

1細胞ゲノム科学

二階堂愛^{1,2,3} 笹川洋平^{1,3} 團野宏樹^{1,3}

¹理化学研究所生命機能科学研究センター バイオインフォマティクス研究開発チーム

²筑波大学 グローバル教育院 ライフイノベーション学位プログラム 生物情報領域

³Knowledge Palette, Inc

目次

目次	1
0. この文章について	2
1. なぜ1細胞トランスクリプトームを計測するのか?	2
2. 1細胞RNA-seq法の原理	3
2.1. RNAシーケンス法とは	3
2.2. 1細胞RNA-seqの実現に必要な反応	5
2.3. PolyA tailing法	5
2.4. Template switching法	6
2.5. RNase H- mediated IVT法	6
2.6. RT-RamDA法	6
3. 1細胞RNA-seq法のハイスループット化	8
3.1. 細胞バーコード法と混合反応	9
3.2. マイクロプレート型1細胞RNA-seq法	10
3.3. 液滴生成流路型1細胞RNA-seq法	10
3.4. マイクロチャンバー型1細胞RNA-seq法	10
3.5. マイクロプレート型と液滴生成流路型の違い	11
3.6. 組み合わせインデックス法	14
4. 1細胞RNA-seq法の性能の向上と評価	16
4.1. 1細胞RNAシーケンス法の性能とは	16
4.2. 精度向上のための分子バーコード法	17
4.3. 分子バーコードを用いた実験精度や手法性能の評価	18
5. 1細胞RNA-seq法の今後	21
5.1. 細胞アトラス構築	21
5.2. 多検体1細胞RNA-seq	21
5.3. 細胞系譜・空間配置	22
5.4. 網羅的な遺伝子への摂動と1細胞RNA-seq	24

5.5. 1細胞マルチオミックス	25
5.6. 多様化する1細胞RNA-seq	26
6. 謝辞	26
7. Reference	26

0. この文章について

この文章は、二階堂愛が執筆した1細胞ゲノム科学、特に1細胞トランスクリプトーム周辺についての動向について書かれたものである。この文章は、許可のない転載や配布などの二次利用は一切認めない。ただし個人利用については妨げない。私や私から許可を得た研究室のメンバーはこの文章をもとに、総説等の出版物などに文章や図を転載する場合がある。その際は、その旨を明記する。出版された原稿はプロの編集を受けて、誤字脱字の修正や語彙統一などの修正が含まれており質が高いため、そちらを読むことを推奨する。この文章や図は笹川洋平氏や團野宏樹氏から提供して頂いた文章を含む。注意を払って執筆しているが、メモの域を出ない部分もありその正しさについては保証しない。この文章は任意のタイミングで更新されており、今後継続して公開され続ける保証はない。本文へのコメントや質問などは歓迎している。

1. なぜ1細胞トランスクリプトームを計測するのか？

多細胞生物は、様々な機能を持つ細胞が協調することで個体が生命活動を行うことができる。細胞は同一個体で同一のゲノム情報を持つにも関わらず、多様な機能を持つ細胞型(celltype)や細胞状態(cell state)が存在する。細胞型や細胞状態は、細胞内で発現する遺伝子によって規定される。遺伝子はRNAへ転写されて、その一部はRNAのまま機能する。残りはタンパク質に翻訳される。これらのRNAやタンパク質が細胞を構成物になったり、遺伝子群を制御・調整することで、細胞の様々な機能が発現する。そのため、1細胞ごとのRNAの種類や量を調べれば、個体や臓器、組織に含まれている様々な細胞の特性を調べられる。一般に検体で発現している全遺伝子のRNA分子(転写産物: transcript)やその量をトランスクリプトーム(transcriptome)と呼ぶ。トランスクリプトームを計測する方法として、現在は、次節で解説するRNAシーケンス法(RNA-sequencing, RNA-seq)が用いられる。

RNAは、核酸であるため、増幅してシーケンスできる。一方で、タンパク質は増幅を行えないため、1細胞から同時にたくさんのタンパク質の種類や量を同定することは困難である。ゲノムDNA配列は、各細胞型で基本的には同じ配列であるため、細胞型の特性はわからない。エピゲノムやゲノムは、分子が2コピーしかないため、検出が困難である。RNAは、十分に発現が高い遺伝子であれば、1,000コピーほどあるため、ある程度の感度があれば検出しやすい。以上の理由から、細胞の特性を理解するにはRNAの量と種類、すなわち1細胞トランスクリプトームを計測すればよい。1細胞トランスクリプトームを計測する手段として、現在もっとも利用されているのが、1細胞RNAシーケンス法(single-cell RNA-seq, scRNA-seq)である。

細胞集団や組織からRNAを抽出して、それらの機能を調べる方法がこれまで利用されてきた。一般にscRNA-seqと対比して、bulk RNA-seqと呼ばれる。当然、組織や細胞集団はそのな

かに複数の既知・未知の細胞型や細胞状態を持つ。疾患や標的の現象を理解するためには、各細胞型や状態ごとに細胞特性を計測しなければ理解できない。というのも、多細胞生物の生命現象は、細胞型や状態ごとに臓器や組織の機能の役割を分担したり、協調したりすることで、成り立っている。臓器や組織の細胞型や細胞状態を計測するためには、それぞれを取り出して、バルクで調べる方法がある。ところが、事前知識なく、どのような細胞型や状態が含まれているかまだわかっておらず、それぞれを取り出す方法もほとんど確立していない。事前知識なく、臓器や組織に含まれるすべての細胞型や状態を捉えるには、1細胞ごとにたくさんの細胞を計測すればよい。得られたデータから細胞型や細胞状態を推定してから、現象の解明を進めればよい。また事前にどの遺伝子が重要か不明であるため、1細胞に含まれたRNAをすべて計測できれば、データから重要な遺伝子を推定できる。

このような背景から、臓器や組織を理解するために、1細胞内の複数種類のRNAの量をたくさんの細胞で調べる方法が必要となるわけである。本章では1細胞RNA-seq法の原理と多様性、今後の開発の展開などについて述べる。

2. 1細胞RNA-seq法の原理

2.1. RNAシーケンス法とは

scRNA-seqを理解するに、まず bulk RNA-seq法について説明する¹。これは、臓器や組織片、培養細胞集団からRNAを抽出して、そこに含まれるRNA量を定量する方法である。通常は、10 ngのTotal RNA、すなわち1万個程度の細胞の細胞からRNAを抽出し、DNAシーケンサーで配列決定できる分子に変換する。

RNA-seqを実現するには、RNAをDNAに変換する必要がある。なぜなら、現在、よく利用されているillumina社のDNAシーケンサーは、DNAしかシーケンスできないためである。さらに、DNAの長さは200-300塩基対程度でなければならない。しかし、RNAは数千塩基を越えるものも多く、そのままではシーケンスできない。細胞から抽出されたTotal RNAは、rRNAやtRNAなどの細胞型の情報を持たないRNA分子が99%程度含まれており、これらの分子をシーケンスしても意味がない。そのため、それらを取り除いてからシーケンスしなければならない。

そこで、RNAを細胞から抽出した後、polyA selectionによりmRNAを精製する。rRNAは、RNAの末端にポリA配列を持たないため、polyAの相補鎖であるpolyT配列で、polyA RNAを捕捉して精製する(図1)。このRNAをシーケンスする方法をRNA-seq、あるいはpolyA RNA-seqと呼ぶ。もうひとつの方法は、rRNA depletionである。これはrRNAに相補鎖を持つDNAでrRNAを捕捉し、取り除く方法である。このRNAを読む方法はTotal RNA-seqと呼ぶ。polyAを持たないRNAでも機能的で細胞型特異的なものがある。また、polyA配列が付加される前のpre-mRNAが存在する。Total RNA-seqはこれらのRNAもシーケンスが可能である。

次にRNAを断片化する。断片化されたRNAをランダムな6文字のプライマー(random hexamer)で逆転写反応を行う。これにより断片化されたRNAはcDNAに変換される。これによって、シーケンスに適切な長さで、かつ、DNAに変換されたことになる。このDNAにシーケンスプライマー部位を付加するために、シーケンスアダプターをライゲーションで付与すれば、シーケンスができるDNAが得られる。これをシーケンスライブラリと呼ぶ。

このライブラリをサンプルあたりランダムに数千万断片のシーケンスを行う。得られた配列断片は、無作為に選ばれてシーケンスされているので、どのRNA由来かは不明である。そこで、これらの配列をRNA配列データベースに計算機中で、マッピングし、そのマッピング数をカウントすることで、どの遺伝子がどのぐらいRNAを転写したかを知ることができる。最終的に得られるのは、サンプル x 遺伝子の次元の行列であり、行列の要素は、RNAの発現量を示す。これを遺伝子発現行列と呼ぶ。この行列から似たサンプルを検索したり、サンプル間で異なる遺伝子を統計的検定によって得たりすることで、サンプルの特性を知ることができる。

このような一般的なRNA-seqは1細胞RNAに応用できない。その理由は、RNA精製を伴う方法は、微量の細胞からRNAを抽出する困難で、精製時にRNAをロスしてしまうためである。また微量RNA由来のcDNAを効率的にライゲーションしなければ、ライブラリを作製できない。

1. PolyA selection



2. RNA断片化



3. 逆転写



4. 末端修復とリン酸化



5. シーケンスアダプター付加

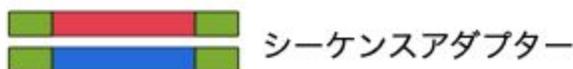


図1. RNA-seqの反応原理

図1. RNA-seqの反応原理

1では、dTが付加されたビーズを利用して、mRNAを回収する。2ではmRNAを断片化、3ではランダムプライマーで逆転写する。4では末端修復とリン酸化によりアダプターのライゲーションの準備をする。5ではシーケンスアダプターをライゲーションしている。

2.2. 1細胞RNA-seqの実現に必要な反応

前節で紹介した一般的なRNA-seqは、1細胞RNAに応用するのは困難である。その理由は、微量な細胞からのRNA精製は、精製時のロスによりRNAを失ってしまう。一方で、RNA精製でrRNAを除かなければ、ほとんどのシーケンスがrRNAに占められてしまいます。また、微量RNA由来のcDNAに、効率的にシーケンスアダプターをライゲーションしなければ、シーケンスライブラリを作製できないが、この効率は高くない。仮にライブラリを作製できたとしても、シーケンス可能な量には数万倍足りないため、DNA増幅が必要になる。つまり、1細胞RNA-seqを実現するには、rRNAの除去をすることなく、RNAをDNAへ変換し、DNAを増幅した後、DNAを断片化、最終的にシーケンスライブラリの作製を行う必要がある。

これを実現するために、1細胞RNAを精製することなく、細胞を細胞融解液に直接採取し、細胞からRNAを取り出す。当然、ゲノムDNAやrRNA、タンパク質など細胞成分が溶け出してくる。細胞融解液には、RNA分解酵素の阻害酵素やDNA消化酵素を入れておき、不必要な成分を酵素的に分解しておく。次に、rRNAを避けて、細胞機能の理解に意味があるRNAのみを捉えるため、poly-A RNAを標的としてオリゴdTプライマーと逆転写酵素を用いて逆転写反応を行う。このとき、オリゴdTプライマーの3'末端に、後の増幅に利用する配列を付加しておく。これをタグと呼び、タグを付加することをタギングと呼ぶ。付加される配列は、1細胞RNA-seqの方式によって異なる。例えば、PCR増幅を行う方法は、PCRプライマーがアニールできる共通配列を付加しておく。一方、in vitro transcription (IVT)によってDNAを増幅する場合は、T7配列を付加しておく。次節ではこれらの詳細を説明する。このように全遺伝子に共通の配列をタギングし、全遺伝子を増幅する技術を一般にはwhole transcript amplification (WTA)と呼ぶ。WTA法はタギングとcDNA増幅法によって、大きく4手法に分類される。

2.3. PolyA tailing法

この手法では、逆転写による第1鎖合成の際に、オリゴdTプライマーにPCR用配列を付加されたプライマーを用いる。逆転写後、RNA-cDNAハイブリッド鎖を二本鎖DNAにし、かつ、PCRで増幅できるようにするには、第2鎖DNAの5'端にPCR用の共通配列を付加する必要がある。そこで、第1鎖の3'端にTerminal Deoxynucleotidyl Transferase (TdT)を利用して、酵素的にポリA配列を付加する(PolyA tailing)([図2A](#))。このポリA配列をターゲットとしたオリゴdTと後のPCR用の共通配列(タグ)からなるタギングプライマーを用いて、第2鎖を合成する。これで両端にPCR増幅用のタグが付いた二本鎖DNAが得られる。

PolyA tailingは非常の効率の高い反応である。また逆転写反応が途中で終了しても、3'末端にAAAが付加されるため、第2鎖合成に進むことができる。逆転写とPolyA tailing反応が独立して実施されるため、後に述べるtemplate switchingよりも反応ステップが1つ増える。また反応温度や時間の影響を大きく受けるため、若干実験がシビアである。逆転写とタギングが別のステップになっているため、それぞれで最適な条件を追求することができるため、効率を向上しやすい。PolyA tailingはQuartz-Seq²やQuartz-Seq³で採用されている。

2.4. Template switching法

第1鎖合成の際にオリゴdTプライマーにPCR用配列を付加する点はPolyA tailing法と同様である。逆転写酵素によっては、3'末端にポリヌクレオチドを付加する性質がある。template switchingでは、第1鎖cDNAにグアニンが3文字付加される性質を利用し、シトシン3文字にPCR用タグが付加されたtemplate switchingオリゴをアニールさせて、第2鎖DNAを合成する(図2B)。これにより、両端にPCR用タグが付いた二本鎖DNAが得られる。template switchingは逆転写中に起きるため、ステップ数がpolyA tailingより1ステップ少ないため、若干簡便である。逆転写がRNAの5'末端まで辿りつかなければ、グアニンが付加されないため、タギングがおこなわれず、その後のステップに進めない。そのため、RNAの全長をとりやすいが、遺伝子の検出漏らしが多くなることを原理的に防げない。Template switchingはSTRT⁴やSmart-seq⁵、Smart-seq²⁶、Drop-seq⁷などで採用されている。

2.5. RNase H- mediated IVT法

この方法は、DNAを鋳型にして相補なRNAを合成するin vitro transcription (IVT)を利用して、1細胞由来RNAを増幅する方法である(図2C)。第1鎖合成の際にオリゴdTプライマーにPCR用配列の配列ではなく、T7配列を付加する。後にこのT7配列からT7 RNA PolymeraseによりRNAを線形増幅するためである。dTプライマーで第1鎖cDNAを合成後、ランダムなRNAプライマーで第2鎖を合成する。ライゲーションでそれらを繋げた後、第2鎖のDNAを鋳型として、T7からIVTによりcRNAを合成する。この方法はCEL-seq⁸やCEL-seq²⁹で採用されている。CEL-seq²は非コード領域にたくさんのリードがマップされるなど原因不明の欠点がある。

2.6. RT-RamDA法

これまでの方法はオリゴdTで第1鎖を逆転写する際にタグを付加する点が共通である。さらに、PolyA tailingやTemplate switching法では、PCR用のタグを逆転写プライマーサイトからみて逆端にも付加する必要がある。この方法には以下のような欠点がある。まず、dTプライマーで逆転写するため、rRNAなどを避けることができるが、ヒストンや長鎖非コードRNAなどの様々なnon-polyA RNAはシーケンスできない。また、3'端から逆転写を行うため、非常に長いRNAは途中で逆転写が止まってしまい、5'側のシーケンスが得られない。逆転写プライマーのプライミングや反対の鎖のタギングが失敗してしまうと、その分子は計測からドロップアウトしてしまう。増幅可能な分子に変換できるかどうかは、両端のタギングの効率のかけ算になるため、dropoutの確率を低くすることが技術的に困難である。

この問題を避けるにはタギングを伴わない核酸増幅法、ポリA配列を頼りにしないランダムプライミングを利用した方法が必要となる。RT-RamDAは、RNAを鋳型として逆転写反応によりcDNAを合成しつつ、1つの標的分子からcDNAが何度も合成させる反応である(図2D)。これはタグなしのランダムプライマーで起きる。まず、準無作為プライマー(not-so-random primer, NSR)を用意する。これは6文字のランダムプライマー配列からrRNAにマッチする配列を除いた配列から核酸合成し、1本のチューブに混合したものである。このNSRを用いて、第1鎖cDNA合成を行う。同時にDNaseIと1本鎖安定化酵素を入れておく。逆転写中にDNaseIはRNA-cDNAハイブリッド鎖のcDNA側だけにランダムなニックを入れる。すると、このニックからRTaseが

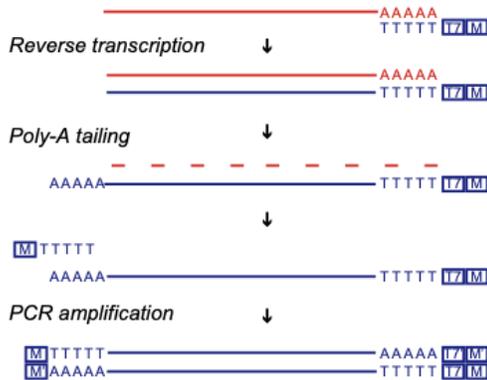
cDNA合成を再び始める。RTase1にはstrand displacement活性を持つものがあり、前に鎖があっても剥しながら合成を続ける。合成中、あるいは合成されたcDNAは一本鎖DNA安定化酵素の働きによりDNase1からの切断を免れる。またこの酵素の働きによってstrand displacementが起きやすくなる。cDNA合成とランダムニックの形成が同時に繰り返し等温反応で起きるため、逆転写しているだけで核酸が数十倍に増幅される。cDNAが増幅されているので、核酸精製など通常の分子生物学的テクニックが使えるようになる。cDNAを精製した後、第2鎖を合成した後、断片化してライゲーションするかタグメンテーションによりライブラリを合成すれば、すぐにシーケンスが可能である。この手法を採用したものとしてはRamDA-seq¹⁰がある。

RamDA-seqは1細胞で完全長のTotal RNA-seqを初めて達成した手法である。

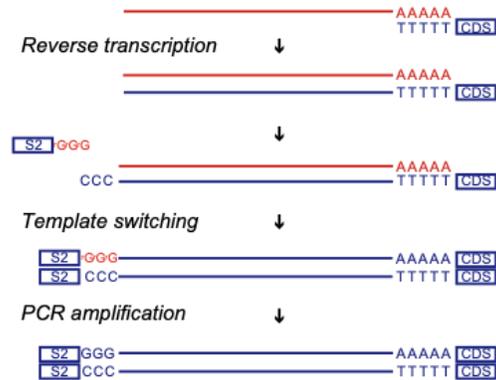
ランダムプライミングなのでRNAの全長からリードが得られるし、dTプライマーに依存しないので、non-polyA RNAも検出できる。1つのRNA分子からたくさんのcDNAを得られるため、遺伝子の捉えもちが非常に少ないのが特徴である。また、断片化されたcDNAが得られるため、ライブラリ作製時のPCRで分子長による増幅バイアスが起きにくい。通常のscRNA-seqはdTプライマーで逆転写、タギング、PCRの3つの段階が成功しなければ、シーケンスライブラリにならず、遺伝子が検出されない。各段階の精度を多少向上させても、遺伝子検出確率は、各段階の成功確率のかけ算になる。たとえば、各段階の成功確率が0.7で3段階だと、 0.7^3 であるため、検出成功率は0.34程度になる。そのため多くの遺伝子を取り漏らすことになる。ところが、RamDA-seqは逆転写の時点でRNAのあらゆる場所からcDNAが合成され、ただちの増幅するため、ほとんどの遺伝子からなんらかのcDNA合成される。すでに増幅されたcDNAをシーケンスライブラリするため、ほとんどの発現遺伝子が検出可能となる。

図2

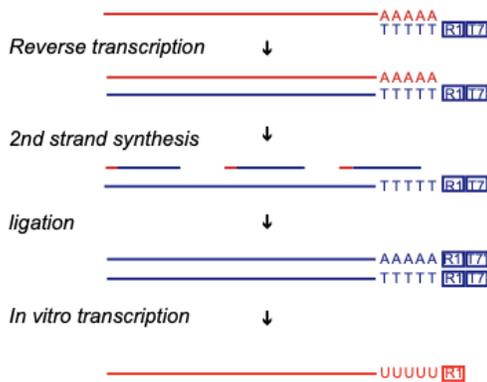
A 1. poly-A tailing



B 2. Template-switching



C 3. In vitro transcription



D 4. RT-RamDA

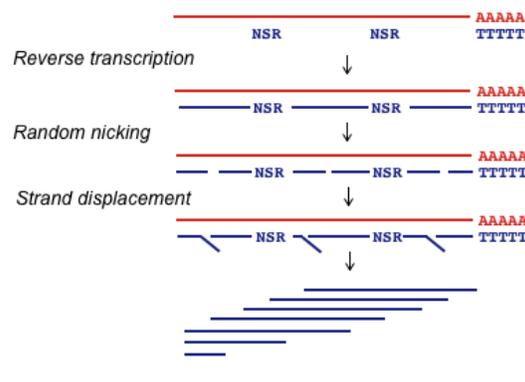


図2. 1細胞RNA-seq法の代表的な基礎反応. 赤線はRNA、青線はDNAを表す。A) TTTTTはポリT、T7はT7配列、MはPCR用プライマー配列を示す。B) CDSとS2はPCR用プライマー配列、CCCはポリC、GGGはポリGを表す。C) R1はシーケンスリード用のプライマー配列、T7はT7配列を示す。D) NSRはnon-so-randomプライマーを表す。

3. 1細胞RNA-seq法のハイスループット化

個体や臓器、組織の機能を理解するためには、たくさんの細胞をシーケンスする必要がある。なぜなら、個体や臓器、組織には複数の細胞型や細胞状態が含まれているからである。通常のbulk RNA-seqでは、一度の実験でせいぜい数十サンプルをシーケンスするだけである。ところが、1細胞RNA-seqでは数百から数万の1細胞をシーケンスしなければ、そのなかに含まれるすべての細胞型や細胞状態を網羅できず、幹細胞や特殊な免疫細胞のような希少な細胞も解析できない。1細胞RNA-seqは、細胞の採取からシーケンスライブラリ作製まで複雑な手順を1細胞ずつ独立に実施しなければならない。そのため、実験の労力やコストが非常にかかる。この問題を解決する方法として、細胞バーコード法と混合反応を利用する方法がある。また、1細胞を高速に取り分ける技術も必要である。本節ではそれらについて解説する。

3.1. 細胞バーコード法と混合反応

1つずつの細胞を実験するのはコストが高い。そこで、1細胞由来の全RNAに目印をつけた後、混合して増幅、シーケンスなどを行い、目印を頼りにデータを1細胞ごとに振り分ける戦略をとる。まず、dTプライマーに細胞バーコード (cell barcode) と呼ばれる配列を付加しておく。大抵は、数十文字列のランダム、あるいは事前に設計されたDNA配列が用いられる。この配列は1細胞ごとに異なる。1細胞ずつ逆転写時を行うと、同じ細胞由来のRNAは同一の細胞バーコードが付加される。この後は、各細胞の逆転写液を混合してしまっても、シーケンス後に細胞バーコードを頼りにデータを分離できる。1細胞と細胞バーコードを1対1で出会わせて逆転写しなければならないが、いくつかの方式が提案されている。次項では代表的な3つの方法について述べる(図3)。

図3

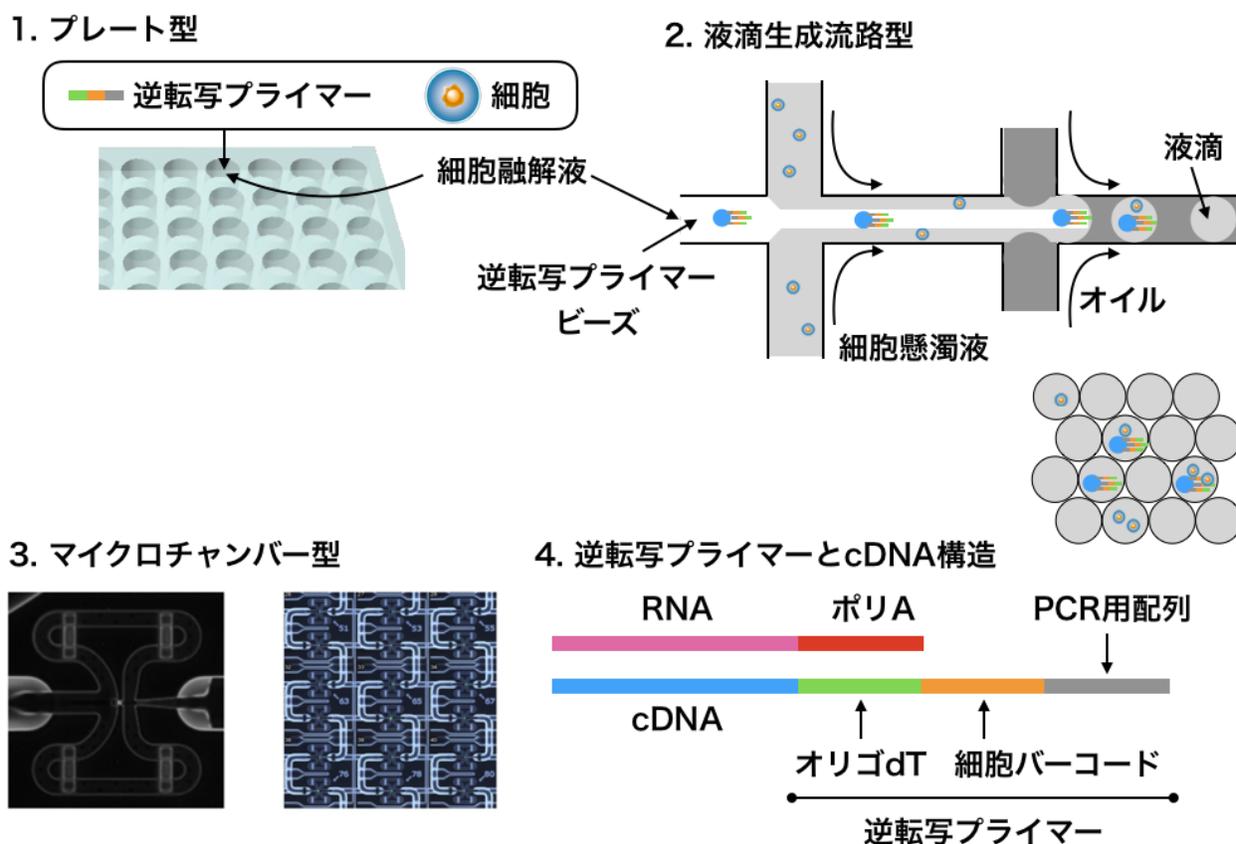


図3. 1細胞RNA-seqの細胞採取法. A)プレート型はセルソーターと96/384ウェルのマルチウェルプレートを利用して細胞を採取する。事前に細胞溶解液と逆転写プライマーを入れておく。B)液滴形成流路の模式図。右下のように液滴は1つのチャンバーに集められる。C)Fluidigm C1型のマイクロチャンバー。左は1つの細胞がトラップされるチャンバーの拡大図。右はチャンバーが連なっている図。D)逆転写プライマーとcDNAの構造の模式図。Template switchingやPolyA tailingで利用される代表的なプライマー構造を示す。

3.2. マイクロプレート型1細胞RNA-seq法

分子生物学実験で良く使われるSBS規格のマイクロウェルプレートを利用する方法である(図3A)。マイクロウェルプレートは96や384ウェルプレートが利用される。このプレートに事前に細胞融解液と逆転写プライマーを分注しておき、セルソーターで1細胞を各ウェルに採取する。事前に分注された逆転写プライマーはウェルごとに異なる細胞バーコードが含まれている。そのため、逆転写後、混合してしまっても細胞バーコードのDNAシーケンスから、データを1細胞ごとに振り分けられる。

このような方法を採用している代表的なscRNA-seq法としてはQuartz-Seq²³やMARS-seq¹¹, CEL-Seq²⁹, SCRB-seq¹²などがある。

3.3. 液滴生成流路型1細胞RNA-seq法

マイクロ流体デバイスを用いて、液滴を高速に生成し、その中に逆転写プライマーと細胞を閉じ込める方法がある(図3B)。まず、細胞懸濁液と、逆転写プライマーが固相化されたビーズが含まれる細胞溶解液、オイルを流路に流す。すると偶然、1細胞とビーズが1対1で含まれる液滴が生成される。液滴のなかは細胞溶解液なので、RNAが溶け出し、ビーズに捕捉される。その後、液滴を融合すれば、各細胞を混合できる。代表的な手法としては、Drop-seq³やinDrop-seq¹³がある。

3.4. マイクロチャンバー型1細胞RNA-seq法

この方式では、特殊な形状のマイクロ流路を利用して細胞を捕捉する。まず流路内に96ヶ所のチャンバーを配置しておき、流路に細胞懸濁液を流すと、1つのサイトに1つの細胞が捕捉される。既に細胞が捕捉されている場所には、物理的にそれ以上の細胞が捕捉されない構造に設計されている。これにより、流路に細胞懸濁液を流すだけで、1細胞の採取が可能となった。

1細胞採取流路は、形状・仕掛け・素材などによって複数種類考案された。しかし、トランスクリプトーム解析では、1細胞採取に加え、1細胞ごとに多段階の分子生物学反応を行う必要がある。例えば、各反応ステップで別の酵素を加えたり、その酵素反応に最適になるように溶液組成を整えるためのバッファーを加える必要がある。これを細胞ごとに行うためには、「細胞数×反応ステップ数」分の溶液添加を制御する仕組みが必要である。スタンフォード大学のQuakeらは、2000年に空圧制御によるマイクロバルブを開発した^{14,15}(図3C)。このバルブは、柔軟性の高いポリジメチルシロキサン(Polydimethylsiloxane: PDMS)を多層に組み合わせて作られており、溶液が流れる流体チャネルのすぐ上または下で交差するように制御チャネルが配置されている。制御チャネルには空気が充填されており、その空圧を機械制御することで、バルブの開閉が可能になる。複数箇所のバルブの開閉を時間的に制御し、溶液を、適切なタイミングに、適切なチャンバーへと移動できる。彼らはこの仕組みを利用し、マイクロチャンバーを用いた、1細胞ごとの多段階反応が可能な自動化システムを構築した。本システムは、彼らが設立したFluidigm社から装置(Fluidigm C1)として販売されている。本システムは操作が非常に簡便であるため、実験コストは高いものの1細胞RNA-seqの世界的な普及に多大な貢献をし

た。しかし、この方式は、細胞のサイズに適したマイクロ流路を選択する必要がある。また、自分の解析対象の細胞集団が、様々な細胞サイズが含まれている場合は、細胞採取にバイアスがかかったり、ダブレットが増えたりする。現在は、先に述べたプレート型や液滴流路型が中心になっており、あまり利用されなくなりつつある。

バルブを利用したマイクロチャンバー以外にも、SBS規格ではない極小のマイクロウェルをPDMSで作製し、1細胞捕捉に利用する方法もある。代表的な方法としては、SeqWell¹⁶、Mirowell-seq¹⁷、iCell8¹⁸、Nx1-seq¹⁹などがある。しかし、どれも検出遺伝子感度が低く、今後の解析が期待される²⁰。

3.5. マイクロプレート型と液滴生成流路型の違い

マイクロプレートとセルソーターを利用した方法は、1細胞を採取後、-80℃で数ヶ月以上凍結保存できるため、一度にサンプルを用意できない時系列サンプルなどの複数検体の実験が行いやすい。セルソーターでは細胞に照射されたレーザービームの散乱光から、様々な細胞特性が計測できる。インデックスソーティング機能を用いれば、細胞特性を計測した細胞そのものを特定のウェルに採取できる。これによりセルソーターからの細胞特性情報とトランスクリプトームデータを1対1対応する。セルソーターからのシグナルを用いることで、細胞特性だけでなく、ダブレットの除去もデータに基づいて行える。もちろん、セルソーターでは、目的細胞のみを取り分けられることは、シーケンスする細胞数を減らすことができるため、大きな利点である。また、サンプルや送液、受け手であるプレートを4℃に冷却しておけるため、細胞を健康に保てる。

通常、384プレートが数枚から十数枚程度、つまり数千細胞程度の実験デザインに向いている。実験には自動分注装置があると良いが、マルチチャンネルピペットでも実施できる。逆に細胞数をある程度少なくして実験も可能なので、研究をスモールスタートできる。

一方で、採取できる細胞数は、96/384ウェルフォーマットに制限されるため、実施する細胞数を増やしていく(スケーラビリティが低い)。また、セルソーターの操作には多少の修練が必要であり、十分に装置の原理を理解しなければ、正しく細胞を採取することが難しい。

SONY SH800, BD Aria II/III, Beckman Mofloなど各社のセルソーターが、96, 384ウェルプレートへのソーティングに対応している。

液滴生成流路方式のDrop-seqと比較して、Quartz-Seq2は検出できる遺伝子数が2.6倍程度あり³、細胞分類だけでなく、細胞機能や細胞状態の違いなど、より詳細な細胞機能の解明に向いている。Quartz-Seq2はヒト細胞アトラスプロジェクトで実施された国際的なベンチマークコンテストで圧倒的な性能を示し、もっとも優れた方法であることが証明されている²¹。特に細胞あたりの検出遺伝子数や細胞型ごとのマーカー遺伝子の検出率は他の手法を圧倒している(図4)。

図4

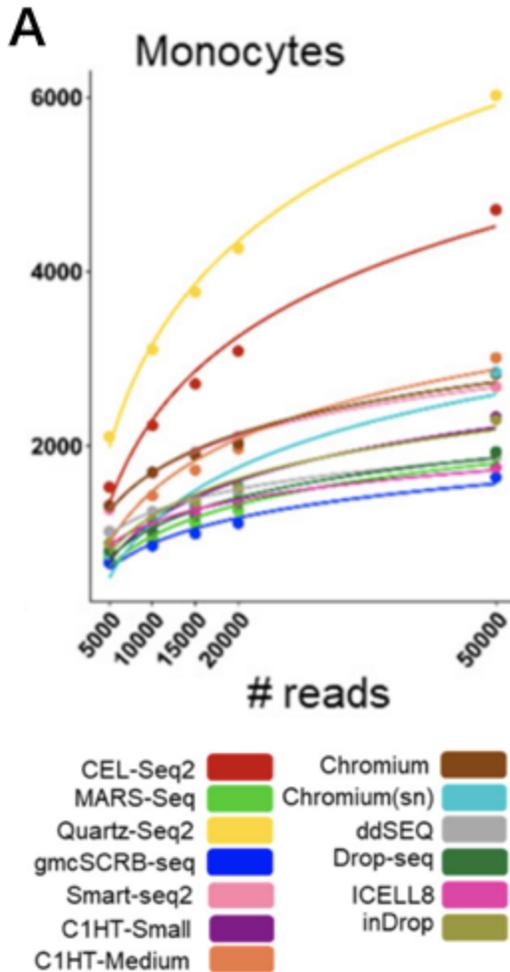


図4. リードあたりの検出遺伝子数。Monocyte1細胞あたり50,000リードのシーケンスを得た場合、Quartz-Seq2では平均6,000遺伝子が検出される。Chromiumでは2,000遺伝子程度。

一方で、液滴流路は流速は早く、数分間で数千から数万細胞を採取できるため、スケーラビリティが高い。通常は流路はPDMSのような安価な材質でできているため、セルソーターと異なり容易にマルチチャンネル化できるため、スケーラビリティを上げやすい。この方法は、専用の流路とビーズ、送液装置が必要であるが、10x genomics社よりChromiumという装置が販売されており、ユーザビリティがよく、比較的簡便に利用できる。ほかにもDolomite Bioや、illuminaとBioRad社の共同開発の装置であるddSeqなどが上市されている。ただし、これらの装置で、特定の細胞のみシーケンスした場合は、セルソーターなどで事前にソーティングが必要になる。装置にもよるが、サンプルや送液などが冷却できず、素早く実験する必要がある。高濃度の細胞懸濁液を用意することで、採取できる細胞数を増やすが、2細胞以上の混入が起きる確率が高くなる。さらに、細胞捕捉率が方法・装置によって5%から80%と様々であり、検体が希少で、細胞数が少ない場合、低い細胞捕捉率は問題となりうる。またセルソーターと異なり液滴に閉じ込められたものが生細胞かどうかを区別できない。そのため死んでしまった細胞から出てくる細胞片が流路に閉じ込められ、そこに含まれているRNAがシーケンスされてしま

う。このため、ミエリンの多い成体脳など、デブリが多く発生する組織での解析が困難である(図5)。装置や試薬を自分で用意できると、コストを下げることが可能である。市販の装置や試薬を購入すると、プレート型とほぼコストはかわらないか、より高額になる。

図5

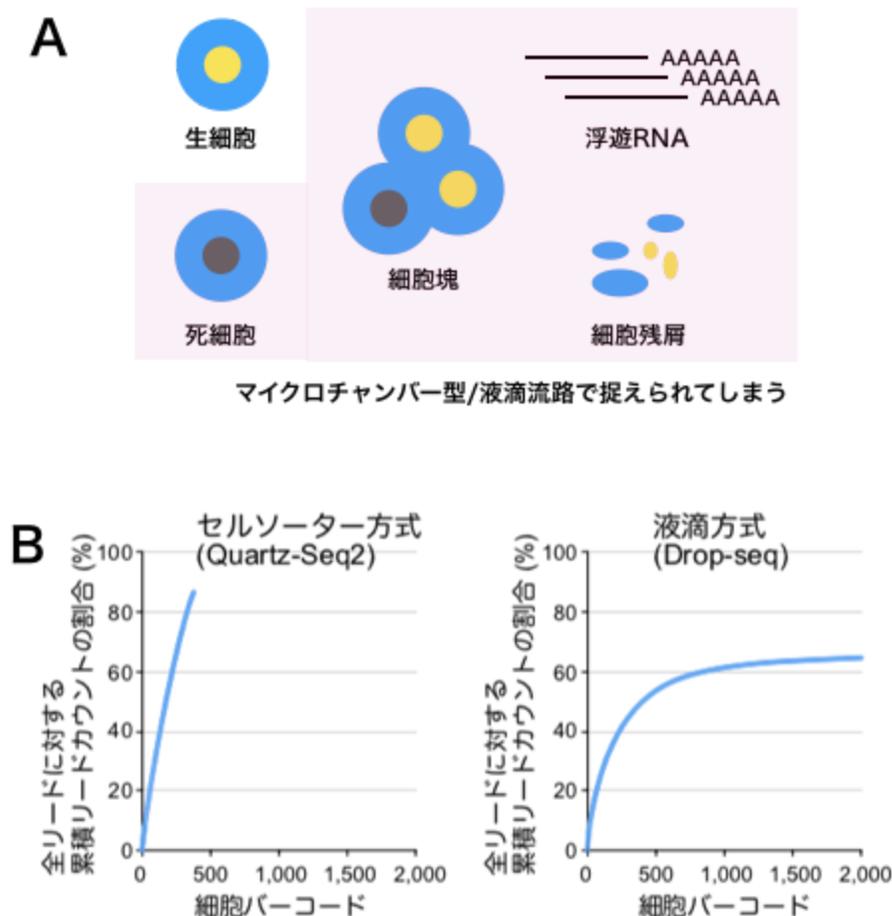


図5. A) 細胞懸濁液に含まれるオブジェクト。生細胞だけでなく、死細胞や細胞塊、細胞残屑、浮遊RNAなどが含まれ、液滴流路方式やマイクロチャンバー方式では、それらもシーケンスされてしまい、ノイズとなる。セルソーター方式では生細胞だけソーティングできるため、これらの混入は少ない。B) 累積リード数の変化。1細胞ずつ得られたリード数を累積したプロットをみると、セルソーター方式では、得られた細胞数に対して直線的にリードが飽和する。液滴方式の場合、リードが生細胞以外のオブジェクトをシーケンスしてしまうため、細胞バーコードが割り当てられず、プロットが早めに飽和してしてしまう。

逆転写プライマーを固相化されたビーズが必要になるが、1つのビーズに同じ細胞バーコードを付加しなければならない。さらにそのようなビーズを数百万通り以上作らなければならず、通常の研究室で自作するのは困難であるが、労力をかければエマルジョンPCR、split-and-pool³、マイクロ流路とPCRを利用した方法¹³などで作製できる。通常は、市販の高額なキットに同梱されているものを利用するのが一般的である。ライブラリによっては、ペアエ

ンドでシーケンスする必要があり、プレートベースの2-4倍のシーケンスコストがかかる場合もあるので注意が必要である。検出できる遺伝子数が少なくいくらシーケンス量を増やしても、プレートベースの方法に匹敵することはない。

これらの方法以外に、マイクロ加工技術で作られたマイクロチャンバーに細胞を採取して、WTA反応を行う方法も提案されている。これらの技術の利点・欠点・入手可能性を理解・考慮し、自身の研究戦略に相応しい技術を選びとることが重要であろう。

3.6. 組み合わせインデックス法

組み合わせインデックス化 (Combinatorial cellular indexing)方式について述べる(図6)。まず、96ウェルプレートに1細胞ではなく複数個の1細胞をそれぞれ採取する。次に、細胞バーコードが付加されたOligo-dTプライマーで細胞内逆転写反応を行う。細胞バーコードは、ウェルごとに異なっている。これを付加しても、ウェル内の数百細胞は同一の細胞バーコードが付加されて、1細胞を区別ができない。そこで、全ウェルの細胞を1つのチューブに混合した後、再び、複数個の1細胞をランダムに96ウェルプレートに取り分ける。次に、2番目の細胞バーコードを細胞内でライゲーションする。これにより、ほとんどの細胞が異なる2つの細胞バーコードを持つ。この混合(pool)と分離(split)を繰り返して細胞バーコードを付加し、1細胞ごとに含まれる細胞バーコードがユニークになる。この方法は、混合・分離をくりかえすことから”split-and-pool”とも呼ばれる。

図6

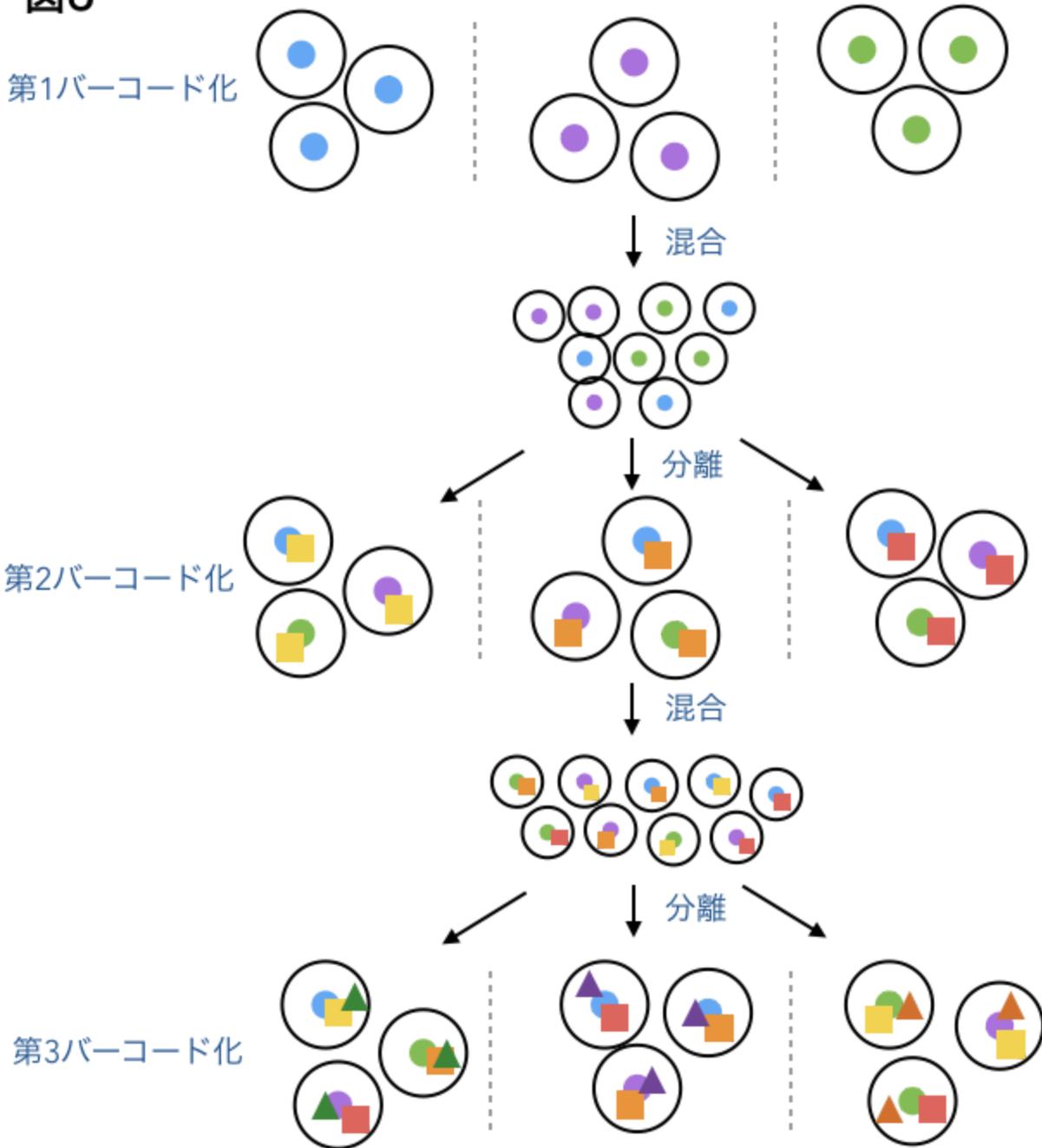


図6. 組み合わせインデックス法の模式図。数千細胞程度が含まれるウェルで、細胞バーコードが付加された逆転写プライマーを利用して逆転写を行う。このとき、細胞を融解せず細胞内で逆転写反応を行う。この細胞を一度プールしてから再び数千細胞程度に分割し、各ウェルに入れる。そこで再び細胞内でバーコード化を行う。これを繰り返すと各細胞は異なるバーコードの組でインデックス化される。

細胞を混合・分離するには、1細胞由来のRNAを液相に分離してはならず、粒子として隔離されている必要がある。そのため、細胞内でバーコード化しなければならない。この方式が、1細胞

胞ゲノム科学に最初に応用されたのは、RNAシーケンス法ではなく、クロマチンが開いた領域を検出するエピゲノミクスのシーケンス方法である²²。近年、1細胞RNAシーケンス法に適応されSPLiT-seqという方法が報告された²³。細胞バーコードを多段階的に簡便に付加できるため、液滴方式より更に、1細胞RNAシーケンス法を高出力化できる潜在能力を秘めている。一方で、cDNAに対して繰り返し標識を行うため、標識効率はその回数のべき乗で悪くなる。そのため、理論的に検出遺伝子感度が低くなりがちである。また細胞により、標識するRNA分子が細胞などから物理的に拡散により失われるなど、方法として改善の余地がある。Dingらのベンチマーク論文²⁴では、SPLiT-seqの検出遺伝子数は、ほかの手法にくらべて悪いことが示されている。

4. 1細胞RNA-seq法の性能の向上と評価

4.1. 1細胞RNAシーケンス法の性能とは

ここでは、scRNA-seqの遺伝子発現量の性能について解説する。scRNA-seqの重要な性能指標として定量性がある。定量性は大きく分けて2つの意味がある。1つはある遺伝子に着目したときに、その遺伝子がそもそも検出できるか、という感度の問題である。前節で解説してきたように、scRNA-seqの反応ステップでロスが起きると、シーケンスライブラリに変換されずに、検出漏れが生じる。RNAの配列、二次構造、結合タンパク質の有無、長さなどによって、逆転写が成功しない遺伝子がある。また逆転写が成功しても、第2鎖cDNA合成のためのタギングが成功しない場合がある。タギングに成功しても、PCR増幅の段階でほかの遺伝子より増幅率が低いと、シーケンスライブラリにならない。シーケンス量が十分でなれば、ライブラリになってもシーケンスされることなく検出されない場合もある。これらの感度は実験回や実験者の影響を受けばつく。これらを抑えるには、各反応効率を高めたり、反応ステップ数を減らすほかない。それらの工夫については前節で解説した。感度は単純に1細胞あたりの検出した遺伝子数や転写産物数を数えたり、その分布をみる。検出された遺伝子数を数える際に、論文によって閾値が違うため注意が必要である。後に解説する分子バーコードを利用した手法では、UMIカウントが0より大きい遺伝子を数えるのが一般的である。しかし、単に検出した遺伝子数を数えるだけでは、本当に細胞のなかにあったRNA分子を数えたのかわからない。つまり偽陽性を検出しただけかもしれない。そこで、bulk RNA-seqで検出できた遺伝子を正解とみためて、scRNA-seqでも検出できた遺伝子数を数える方法がある。この方法を有効であるものの採用する論文は少なく、高性能を謳っている手法でも十分に評価がされてないことになり、これが業界に大きな混乱を招いている。

2つ目は、細胞内のRNA分子数を正確に定量できるか、という精度(accuracy)の問題である。精度は真の値にどれだけ近く計測できるかという正確さ(trueness)と、どれだけ再現良く計測できるかという精密さ(precision)からなる。ここでは、JIS Z 8101やJIS Z 8103の定義を採用した。scRNA-seqでは、RNAの長さや配列のGC%、二次構造のとりやすさなどによりPCR増幅の偏りが生じる。これが正確さを減少させる要因となる。精度を向上させるには、各反応効率を高めたり、反応ステップ数を減らすことが有効であるが、分子バーコードを利用する

ことで補正できる。次の説では分子バーコード法について解説する。精度を計測するにはCV²-plotやポアソン分布へのフィッティングを利用することが多い。

1細胞RNA-seq法の性能として、スループットがある。ここではいくつかの細胞をどのぐらいの時間でライブラリ作製ができるか、という意味である。スループットはコストや労力との兼ね合いもあるため、主観的な評価指標であるが、実際に実験を計画するうえでは重要であろう。

完全長型のscRNA-seq法では、リードの被覆度を利用して、その完全長のカバー率を評価する。あらゆるRNAの長さを100%に正規化し、各領域でどのぐらいのシーケンスリードが存在(被覆)するか、という指標である。完全長型は遺伝子全体にリードが被覆するが、3'端だけを読む高出力型scRNA-seq法は3'端にリードが偏る。完全長型scRNA-seqを実施した際に、RNAの質が悪くRNAが分解しているときや、WTAがうまくいかない場合は、このプロットが乱れる。遺伝子の長さを正規化したプロットをrelative gene body coverage plotと呼ぶ。一方、遺伝子の長さを正規化せずに物理的な塩基数をX軸に取るabsolute gene body coverage plotと呼ぶ。後者のプロットは実際の反応がどの程度うまく起きているか判断するのに適切である¹⁰。

4.2. 精度向上のための分子バーコード法

1細胞RNA-seq法はPCRなどの増幅法を用いるため、分子の長さやGC%などにより増幅率に差が出てしまい、分子数定量の精度が損われる。そこで、RNAやcDNAなどの分子ごとに、ユニークなDNA配列、分子バーコードをつけてシーケンスする。これによって、仮にPCR増幅バイアスなどで、核酸分子の存在比率が変わっても、分子バーコードによって補正できる。これに1細胞RNAシーケンス法の遺伝子発現定量性を向上できる(図7)。

分子バーコード法は、2003年にその理論が提唱されており²⁵、Fodorラボにより、はじめて実験・報告された²⁶。分子バーコードを付与する方式として、2つの方法がある。1つは、逆転写反応の際に、cDNAに分子バーコードを付与する方式で、TaipaleラボとLinnarsonラボの共同で報告されている²⁷。彼らは、初めて分子バーコードをUMI (unique molecular identifier)と呼び、後に同方式を1細胞RNAシーケンスに実装したことから、1細胞RNAシーケンス法では、UMIの方がメジャーな呼び方である²⁸。現在の1細胞RNAシーケンス法の多くで、cDNAの5'末端もしくは、3'末端に、UMI配列を付与する。UMIによりPCRバイアスを除去したリードは、UMIカウントと呼ばれ、シーケンサーにより検出されたmRNA分子とほぼ同意義と考えられている(図7)。

もう1つは、断片化されたゲノムDNAもしくは断片化されたRNAから作られた二本鎖cDNAに対して、DNAライゲーションにより分子バーコードを付ける方式で、FodorラボとXieラボから報告されている^{26,29}。1細胞RNA-seqでは、合成されたcDNAに対して増幅前にライゲーションを行うステップは存在しない。微量なcDNAのライゲーションは効率が悪く、事前にPCR増幅などが必要になる。scRNA-seqでは、PCR増幅のばらつきを減らすための分子バーコードを付加したいので、PCR増幅後のcDNAに分子バーコードをライゲーションする意味はない(図7)。

次にUMIの効果を評価する方法を紹介する(図7)。図7Bでは、Quartz-Seq2でのUMIの補正効果を示している。384 wellの全ウェルに、精製した1細胞レベルのtotal RNA(10 pg)を入れて、遺伝子発現量の検出の技術的ばらつき(technical noise)を評価している。X軸には、各遺伝子の平均発現量、Y軸に各遺伝子の発現量のばらつきをプロットすることからCV²-mean plotと呼ばれる。RNA分子数のばらつきは理論的にはポアソン分布に従う。ポアソン分布から生成される確率変数は、CV²-mean plotでは、対角線の直線に相当する。そのため、技術的なノイズがなくな

れば、実測値が対角線に乗る。Quartz-Seq2³の実測値では、UMI補正前でも技術的なばらつきは少ない。これはもともとPCR増幅バイアスが非常に少ないことを示す。さらに、UMI補正するとポワソン分布の理論値にほぼ重なる。つまり、UMIを利用して技術的なばらつきをほぼ消すことができ、非常に遺伝子発現の定量性が高いことを示す。このようにCV²-mean plotを利用することで、増幅の前後の技術的なばらつきを評価できる。

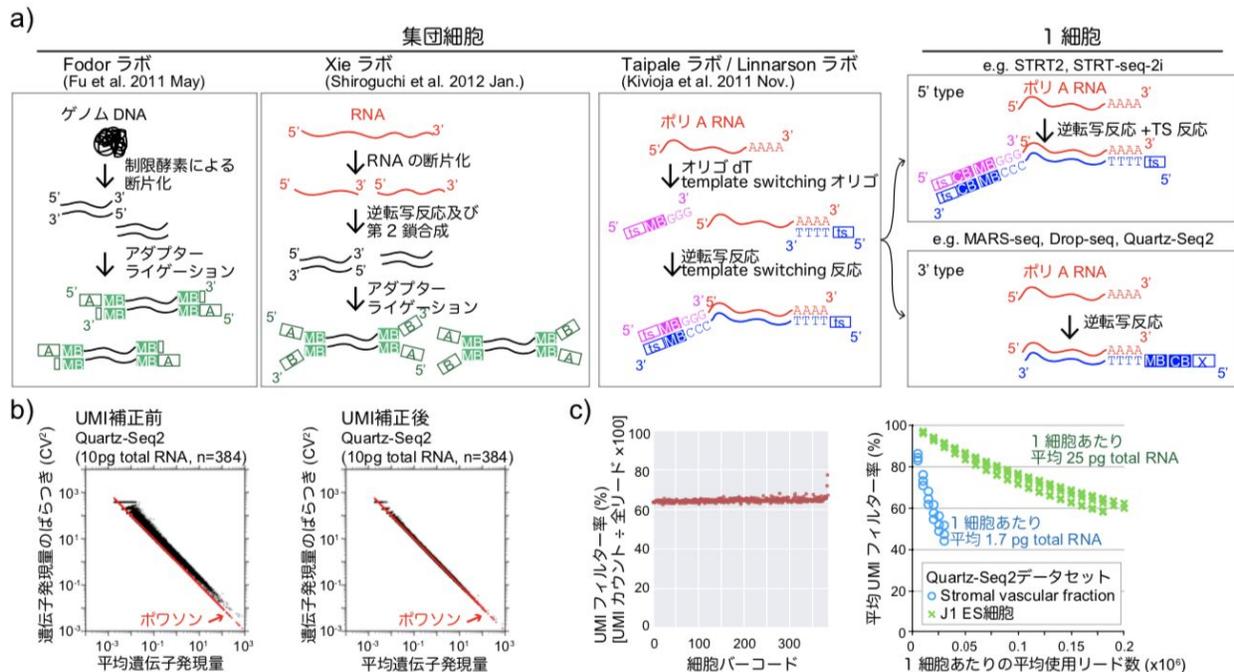


図7. 1細胞RNA-seq法と分子バーコード技術

a) 主要な分子バーコード技術および1細胞RNA-seq法への流れ。b) 分子バーコードによるPCRバイアスの低減。c) UMIフィルター率と1細胞あたりにアサインされるリード数の関係。

4.3. 分子バーコードを用いた実験精度や手法性能の評価

UMI補正前後のリードカウントの比率を見ると、シーケンスリード数が適切であったか判断できる(図8)。この比率は、1細胞がもつ全mRNA分子数やシーケンスリード数に依存する。1細胞あたりの平均total RNA量が、1.7pgのstromal vascular fraction (SVF)と25pgのES細胞の場合、SVFの方が、検出されるUMIカウント数が早く飽和し、結果早く比率が落ちる。UMIカウントが飽和していればリード数が十分であることがわかる。また、高い感度を持つ1細胞RNAシーケンス法ほど、比率が下がりにくく、多くのmRNA分子を捉えられる。そのため、シーケンス法の性能比較にも利用できる。

高出力型1細胞RNAシーケンス法では、1細胞あたりに使用できるリード数が、細胞集団のRNAシーケンスと比較して、1/100、約10万リード以下と少なくせざるをえない(図3a)。それはシーケンスコストを抑えるためである。そのため、限られたシーケンスリードを漏れなく利用し、感度や精度を最大化することが重要である。ところが、これまでに説明したように、非1

細胞由来リードや、PCR重複によるリードなど様々な要因で、シーケンスリードはロスしていく。

そこで、UMI conversion efficiency (UCE)³という評価指標で、どれだけロスなくリードを利用しているかを定量化する(図8B)。この指標は、シーケンスされたすべてのリードのうち、どのぐらいがUMIカウントに変換されるかを示す。ここで、UCEを用いて様々な1細胞RNAシーケンス法の性能評価を行った例を挙げる。まず、あらゆる実験条件を正確にあわせて、UCEを計算した。その結果、Quartz-Seq2が最も高いUCEを示し、リードを効率良く使っていた。UCEが高いと多くの種類のmRNA分子を検出できる。その結果として検出される遺伝子の種類数が多くなる(図8C)。Quartz-Seq2は、高い検出遺伝子数を保ったまま、多数の細胞を解析できるため、希少な細胞の状態や細胞機能を類推可能である³。このようにUCEを利用すれば、1細胞RNAシーケンス法の開発者はその性能比較できる。また、1細胞RNAシーケンス法のユーザは、自身が行う実験の品質評価が行えるため、実験ごとに計算するのがよいだろう。

次に外部コントロールRNAによる1細胞RNAシーケンス法の評価について述べる。まず、ポリA鎖のついた92種類の外部コントロールRNA (ERCC spike-in RNA)をあらかじめ決めた分子だけ用意し、1細胞RNAシーケンスの際に混入させる。このRNAは、細胞の品質に影響されずに増幅されるため、実験のコントロールとして利用できる。外部コントロールRNAの50%検出確率(50% detection probability)という指標を使うことで、実験やその手法の性能評価ができる。まず、92種類の外部コントロールRNAのUMIカウントを検出できたかどうかの2値にする。次に、入れた外部コントロールRNAのcopy数をX軸に、2値化されたUMIカウントをY軸に92点をプロットする。このプロットからRNAの50%検出されたコピー数を近似で導き、RNA分子の検出限界とする。これが低ければ低いほど、少ない分子数を検出できると判断する³⁰。

このように外部コントロールRNAは1細胞RNAシーケンスの性能・品質評価に利用されるが、その利用には十分な注意が必須である。まず、反応原理によって、外部コントロールRNAのほうが細胞由来mRNAより検出しやすい場合もあるし、その逆もある(図8C)。また、1細胞あたりの入力リード数に依存して、外部コントロールRNAの検出力が異なる。そのためリード数を厳密に揃えて比較する必要がある。さらに、外部コントロールRNAの検出率は、1細胞がもつtotal RNA量に非常に強い依存性がある(図8D)。たくさんの量の外部コントロールRNAを入れれば、その指標では見かけデータが良くなる。しかし、外部コントロールRNAばかりがシーケンスされ、細胞由来mRNAの検出感度は下がることになり、本末転倒である。このような基本的な考えかたが不十分で、まったく異なる細胞種や条件で行われた1細胞RNAシーケンス実験データから、外部コントロールRNAのデータだけ比較している論文が多く散見され、この分野に大きな混乱を招いている。本来、知りたいのは1細胞由来mRNAの情報であり、外部コントロールRNAの検出力ではない。また、外部コントロールRNAの分注の際のエラーが存在する。十分に精度の高いscRNA-seqの場合、この分注エラーのほうが大きく、増幅のばらつきではなく外部コントロールRNAの分注ばらつきの評価になってしまう。マイクロウェルプレート型ではなくDroplet型のscRNA-seqでは、外部コントロールRNAを加えるのが原理上非常に困難であり、そもそも利用できない。以上の理由から、外部コントロールRNAよりも、細胞由来mRNAの検出力を見ることが本質的であり、正確な性能評価法であろう。外部コントロールRNAの利用は今後推奨されないだろう。

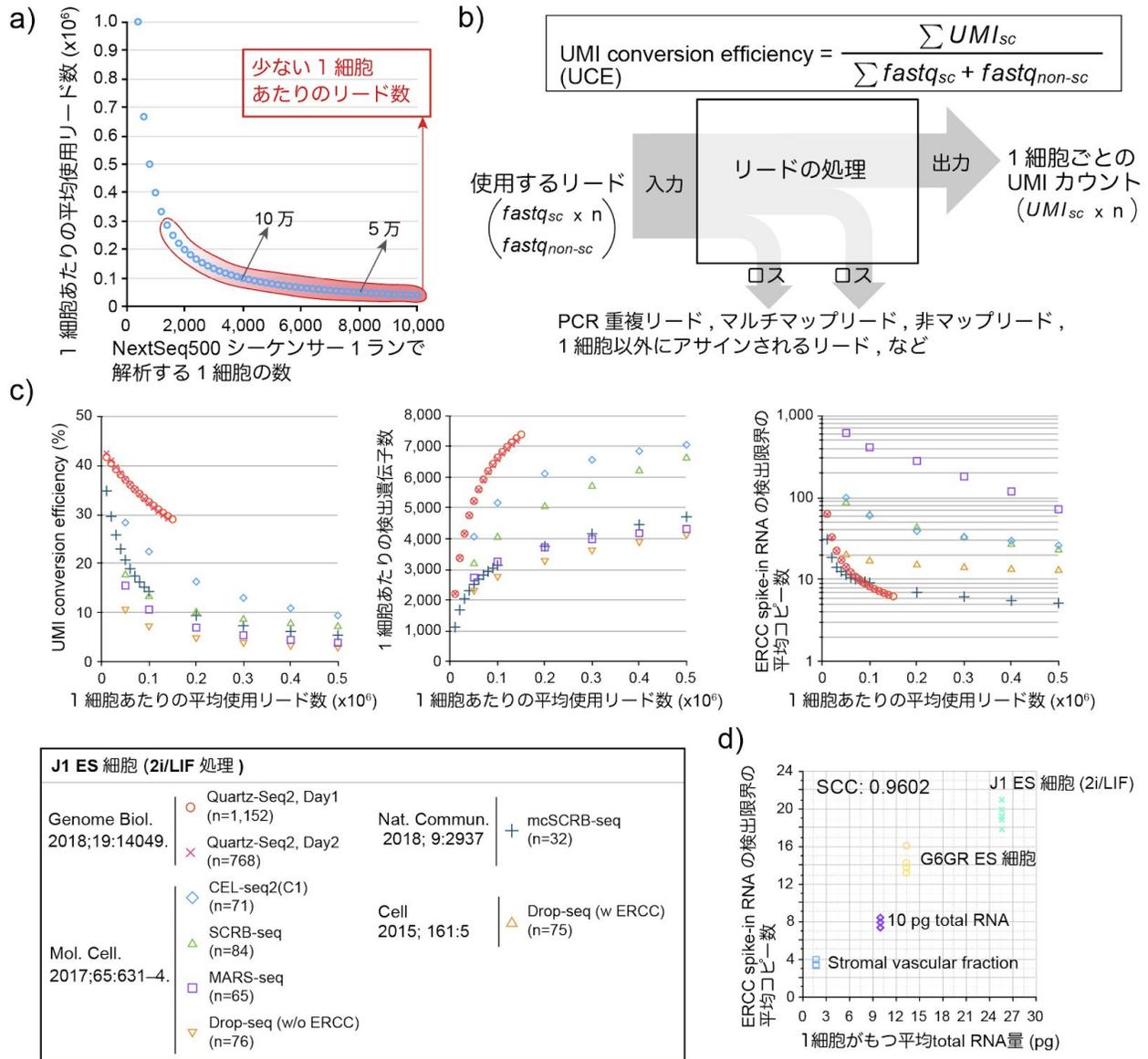


図8. 高出力1細胞RNAシーケンス法の定量性能

a) 1細胞あたりに使用される入力リード数とNextSeq500 1ランで解析する細胞数の関係。 b) UMI conversion efficiencyをもとめる計算式 c) マウスJ1 ES細胞をつかった1細胞RNAシーケンス法の比較。 d) Quartz-Seq2における外部コントロールRNAの検出限界コピー数と解析対象の細胞がもつtotal RNA量の関係。

5. 1細胞RNA-seq法の今後

5.1. 細胞アトラス構築

ハイスループット型scRNA-seq登場により、現在、ヒト全細胞型を収集するHuman Cell Atlas³¹という国際プロジェクトがスタートした。このプロジェクトでは、肝臓、肺、心臓、初期発生、脳、小腸など様々な臓器の1細胞RNA-seqのデータを集積する国際協調プロジェクトである。ヒト全細胞アトラスの完成によって疾患の解明や創薬技術などが進展すると期待されている。このプロジェクトでは細胞の分類ができるが、RNAの3'端しかシーケンスできないので、RNA配列はデータベース化されない大きな問題がある。RamDA-seqはその解決のひとつとなる。高出力型のscRNA-seqは3'端しかシーケンスしないが、WTA後は、全長に近いcDNAが得られている。そこで、このcDNAを長鎖DNAシーケンサーで読むことで、polyA RNAのみであれば、全長に近い配列を得ることもできる。がんなどのサンプルでは変異がどの細胞型で特異的に蓄積するかという問題が設定があり、その場合はこのように全長を読むscRNA-seqが有効になるだろう。

5.2. 多検体1細胞RNA-seq

今後の1細胞RNAシーケンス法の発展として、多検体の1細胞RNAシーケンス法の開発が重要となるだろう。例えば、大量の時系列やコホート、薬剤スクリーニングなどでは、数百から数万規模のサンプルに対するscRNA-seqが必要になってくる。高出力型1細胞RNAシーケンス法は、たくさんの1細胞を解析できるが、元になる細胞集団としては、少数の検体が由来である。数十から数百の検体をすべて1細胞RNAシーケンスすることは困難である。それをある程度解決する技術として、Cell Hashingが提案されている³²。Cell Hashing(図9)では、細胞表面に結合する抗体にポリA配列と抗体ごとにユニークなバーコード配列をもつDNAプライマーを付与しておく。これで細胞を標識した後、検体を混合して1細胞RNAシーケンスを実施する。つまり、検体ごとに細胞を標識すれば、検体の段階で混合反応できるということである。これは検体バーコードと呼ぶことができそうである。この技術は、そもそも、表面タンパク質を標的とした抗体を利用し、表面タンパク質の種類や量とトランスクリプトームを同時に1細胞で計測できるCITE-seqを応用した技術である³³。これにより100程度の検体での1細胞RNAシーケンスが可能になった。抗体の他に、小分子化合物³⁴や細胞膜に埋め込まれる脂質³⁵を使うMULTI-seqも提案されている。split-and-pool法もたくさんの細胞をシーケンスライブラリ化する手法であるが、多検体への応用も可能であろう。

一方で、Cell Hashingやsplit-and-poolでたくさんのライブラリが作製できても、1細胞あたりのシーケンスライブラリDNA調整費やシーケンスコストが、律速になっている。1000細胞、1万条件で1細胞あたり10万リードが必要となると、総額10億円以上になり、現実的ではない。今後、より高出力な1細胞RNAシーケンスを実現するには、反応とシーケンスコストが1/100-1/1000ほどになる必要があるだろう。今後のさらなる技術発展に期待したい。

図9 Cell hashing

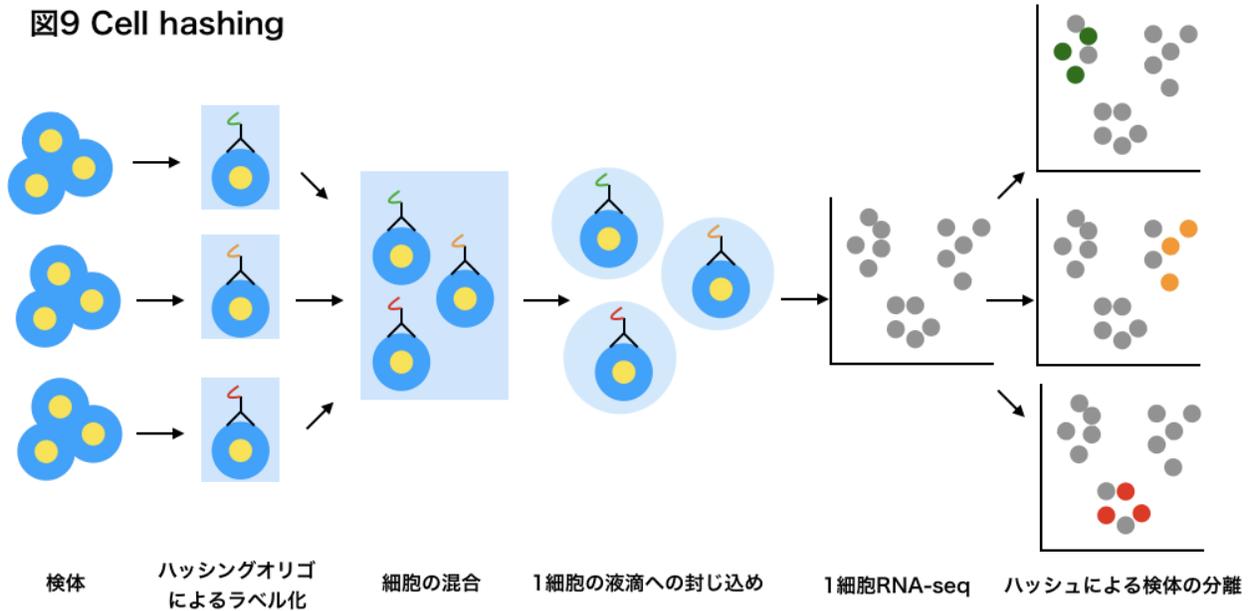


図9. Cell hashing法の模式図。各検体に抗体で標識をする。抗体にはポリA付きのDNAバーコードが付加されているため、細胞を混合して反応しても、バーコードを頼りに分離できる。

5.3. 細胞系譜・空間配置

これまで紹介した1細胞RNA-seq法はすべて1細胞をランダムに分離してからシーケンスを実施する。そのため細胞分化系譜や空間配置の情報は失われてしまう。

近年は、In situ RNA sequencingであるSTARmap³⁶やMultiplex FISH技術であるMERFISH³⁷や、seqFISH、seqFISH+³⁸が発達し、2次元培養されている細胞や組織切片のどこでどの遺伝子が発現しているか計測できるようになってきた(図10)。これによりオルガノイドのどこからどのような遺伝子が発現しているかを調べることができるようになるだろう。

図10 空間発現プロファイル

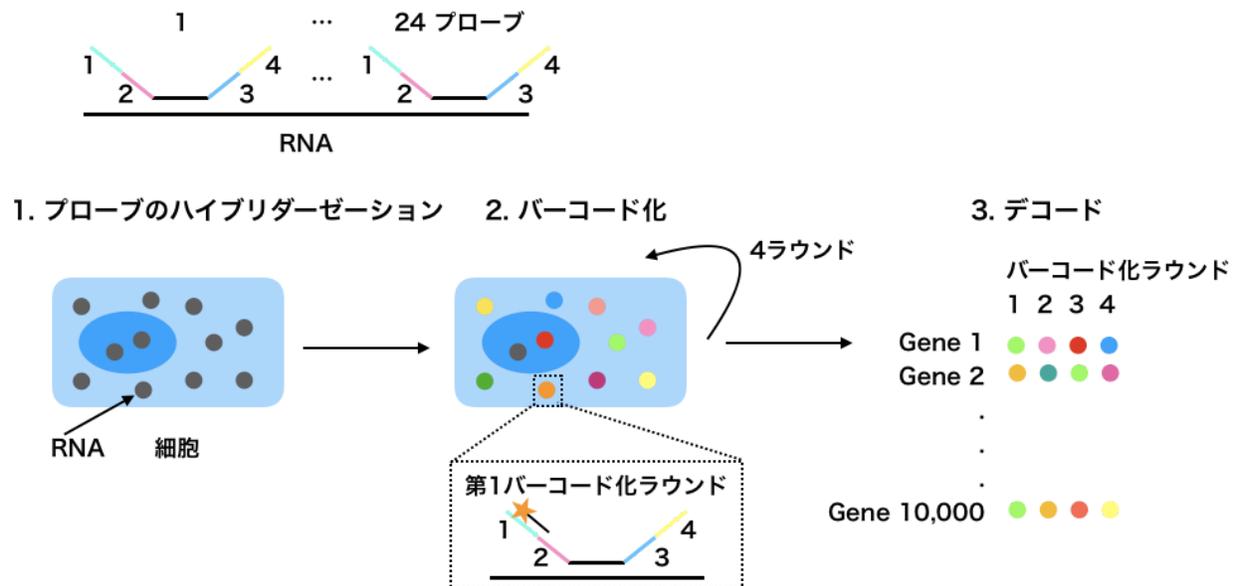


図10. 空間発現プロファイルの計測法seqFISH+法の模式図。遺伝子あたり24種類のプローブが設計されている。各プローブには4箇所の蛍光ラベルされたDNAが結合するサイトがある。このようなプローブを用いてFISHを行い、画像解析することで、1万遺伝子のFISHを実現できる。

また、細胞型間の系譜を計測する方法が提案されている³⁹⁻⁴⁶。原理としては大きくゲノム編集型と組み換えタンパク質型に分けられる。最初に発表されたGESTALT法は、ゲノム上に細胞を区別するバーコード配列を挿入し、その配列から細胞分化の系譜を推定する方法である(図11)。ゲノムにノックインされた蛍光タンパク質の3'UTR領域に、gRNAの標的配列を複数並べたものをバーコードアレイと呼ぶ。この細胞にCas9を発現させると、1細胞ごとにバーコードアレイにランダムな変異が入る。これによって、各細胞が1細胞ごとに異なるバーコードを持つことになる。個体発生や細胞培養後、バーコード領域をシーケンスする。似た配列を持つ細胞は母子・姉妹関係にある可能性が高く、進化系統解析と同じ要領で、細胞系譜を描くことができる。このバーコード配列は、蛍光タンパク質の3'UTRにコードされており、mRNAとして転写している。そこで、1細胞RNA-seqでこれらの細胞をシーケンスすることで、1細胞トランスクリプトームと系譜バーコードを同時にシーケンスできる。つまり細胞型と系譜関係を同時に計測できる。これらの技術は哺乳類細胞やオルガノイドでも原理的には応用可能である。今後はこのような空間配置や系譜情報を得ることでオルガノイドを利用した発生・臓器再生の理解が進むだろう。

図11 細胞系譜解析

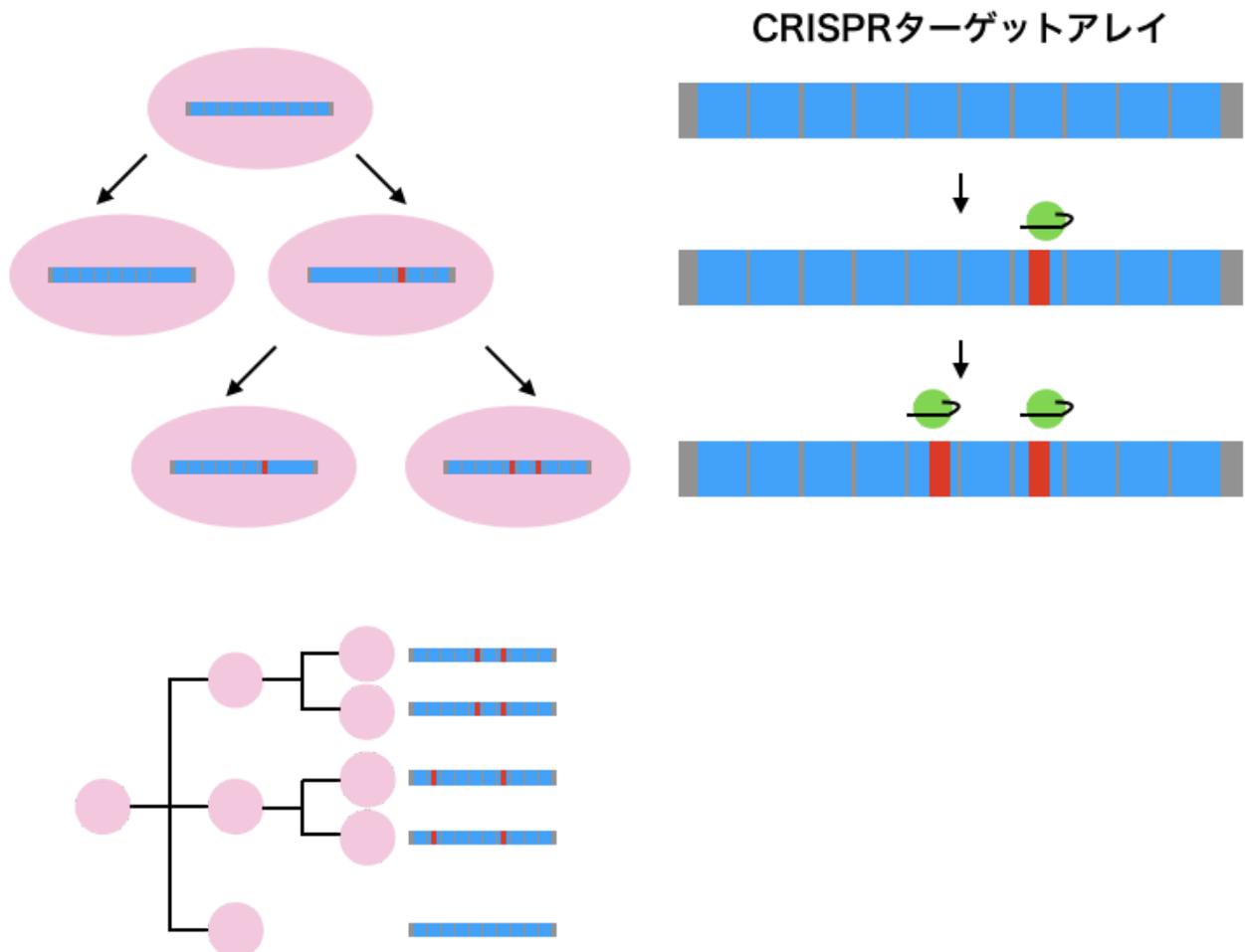


図11. 細胞系譜解析法GESTALTの模式図。ゲノムに、10種類のgRNAのターゲットサイトを導入する。このターゲットアレイをgRNAとCas9がランダムに変異を入れる。同じ変異パターンを持つ細胞は系譜関係にあると言える。

5.4. 網羅的な遺伝子への摂動と1細胞RNA-seq

数千から数万の1細胞RNA-seqが比較的簡単に安価に実施できるようになった。このように生命科学の実験技術のなかで、一度に数千から数万サンプルを計測できる手法はほとんどない。1細胞RNA-seqを大量サンプル計測法として捉えなおす動きがある(図12)。例えば、CRISPR/Cas9による遺伝子ノックアウトライブラリによって、1細胞ごとにランダムに遺伝子をひとつずつ破壊し、1細胞RNA-seqを実施することで、数千から数万の遺伝子破壊とその結果のトランスクリプトームを1度の実験で得られる⁴⁷⁻⁵⁰。

図12 網羅的摂動とscRNA-seq

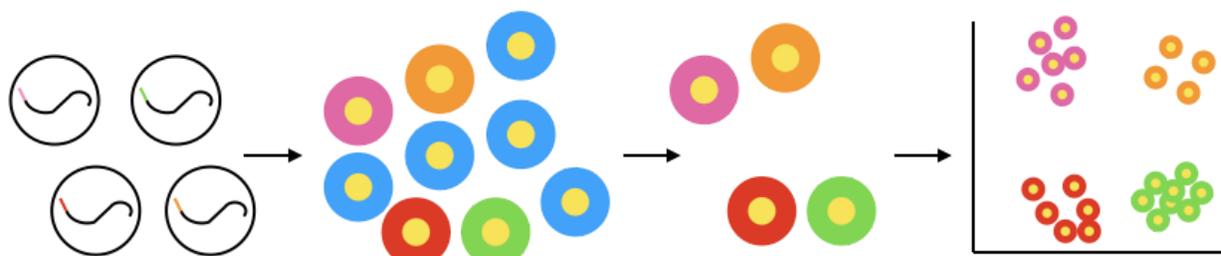


図12. Pertub-seq法の模式図。CRISPRライブラリにポリA付きのDNAバーコードを組み込む。このライブラリを細胞に感染させると、1細胞ごとに異なる遺伝子に変異が導入される。この細胞を1細胞RNA-seqすることで、変異が起きた遺伝子とその結果である遺伝子発現プロファイルの変化を同時に計測できる。

5.5. 1細胞マルチオミックス

ゲノム配列の修飾やクロマチン構造、クロマチンの開閉位置などは、遺伝子発現の制御に関わっており、細胞ごとに特異的なパターンを示すため、1細胞で計測できれば、細胞機能やその成り立ちの理解が深まる。2015年にはトランスポゼースの挿入位置をシーケンスすることで、ゲノムの開構造をシーケンスするsingle-cell ATAC-Seq^{22,51}が登場した。2013-2016年頃には、1細胞のDNAメチル化をシーケンスする技術が開発された⁵²。2013年頃から1細胞ごとにクロマチンの接近位置を特定するHi-Cの実験を行う方法が提案され始めた⁵³⁻⁵⁷。2015年には特定のタンパク質がゲノムのどこに結合するかを1細胞ごとに調べるsingle-cell ChIP-seq⁵⁸が開発された。いずれの1細胞エピゲノムのシーケンス法も、実験精度の問題で、ゲノム全体を真にカバーしているとは言えず、今後の技術開発の発展が期待される。Chromatin integration labelling (ChIL)法はこれを解決する提案のひとつである⁵⁹。特定のタンパク質抗体にトランスポゼースの結合サイトを付加しておき、クロマチン内の結合サイトにトランスポゼースを接近させる。トランスポゼースはタンパク質結合サイトの周辺にT7プロモーターを挿入する。T7からIVTを行い、そのRNAを読めば、タンパク質の結合サイト周辺をシーケンスできるという方法である。これは微量の細胞から実施でき、その効率やゲノムのカバー率も高い。

生命の情報は、DNA-RNA-タンパク質という順番で情報が伝達され、機能が発現するというセントラルドグマがある。これらの情報を1つの細胞が同時に計測することで、各現象の因果関係を解析できる。これを1細胞マルチオミックスと呼ぶ。まず、ゲノム配列とトランスクリプトームを同時に計測するシーケンス法G&T-seq, DR-seq, Simul-seq, SIDR-seq⁶⁰⁻⁶³が登場した。背景としては1細胞RNA-seqと並行して1細胞ゲノムDNAシーケンスの技術も発展してきたためである⁶⁴⁻⁷⁰。ただしゲノム配列のカバー率は高くなくゲノムのコピー数多型の解析にとどまり、1塩基単位での網羅的な変異解析に使える性能ではない。次に、DNAメチル化とトランスクリプトームの同時計測を行う方法scM&T-seq, scMT-seq⁷¹⁻⁷⁴が開発された。クロマチン状態とDNAメチル化を同時に検出する方法scCOOL-seq, scNMT-seq^{75,76}も提案されている。DNAメチル化とHiCを同時に達成するscMethyl-HiC⁷⁷も提案されている。ただし、1細胞レベルのエピゲノムシーケンスの感度が低い、改善の余地が残っている。RNA量とタンパク質の種類、あるい

は量を同時に計測する技術CITE-seq やREAP-seq など^{33,78-80}が登場したが、網羅的なシーケンス技術を利用した方法としては、実用的なものはまだない。

5.6. 多様化する1細胞RNA-seq

RNAは核で転写し様々な修飾を経て、一部はタンパク質に翻訳され、一部はRNAのまま機能する。これまでは長いRNAについてのみ議論してきたが、microRNAを1細胞でシーケンスする手法も提案されている。また核と細胞質をマイクロ流体装置で分離し、それぞれを1細胞RNA-seqする手法SINC-seqも提案されている⁸¹。また細胞を培養中にRNAをパルスラベルし、ラベル化されたRNAのみをシーケンスすることで、新生RNAのみを捉える手法scSLAM-seqやNASC-seqも提案されている^{82,83}。先に紹介したRamDA-seqは、非ポリA RNAやイントロンのある長いpre-mRNAも捉えられる¹⁰。RamDA-seqではRNAが断片化されてcDNAが合成される。1細胞内の異なるisoformを捉えるには、1分子単位でRNAの全長をシーケンスする必要がある。そのためSmart-seq法と同様の原理で合成したcDNAをロングリードシーケンサーでシーケンスする手法RAGE-SeqやSciSO-Seqも提案されている⁸⁴⁻⁸⁶。このようにRNA-seqであっても、多様な手法が提案されており、1細胞レベルのRNA代謝 (RNA processing)の理解に貢献するだろう。

6. 謝辞

理研BDR BiTのメンバーや共同研究者らとの議論の末に得られた知識が含まれており、感謝する。

7. Reference

1. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98 (2011).
2. Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14**, R31 (2013).
3. Sasagawa, Y. *et al.* Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol.* **19**, 29 (2018).
4. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly

- multiplex RNA-seq. *Genome Research* **21**, 1160–1167 (2011).
5. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
 6. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
 7. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
 8. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
 9. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
 10. Hayashi, T. *et al.* Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.* **9**, 619 (2018).
 11. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
 12. Soumillon, M., Cacchiarelli, D., Semrau, S., van Oudenaarden, A. & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* 003236 (2014). doi:10.1101/003236
 13. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
 14. Unger, M. A., Chou, H. P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).
 15. Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**,

580–584 (2002).

16. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
17. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **173**, 1307 (2018).
18. Goldstein, L. D. *et al.* Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* **18**, 519 (2017).
19. Hashimoto, S. *et al.* Comprehensive single-cell transcriptome analysis reveals heterogeneity in endometrioid adenocarcinoma tissues. *Sci. Rep.* **7**, 14225 (2017).
20. Hughes, T. K., Wadsworth, M. H., Gierahn, T. M. & Do, T. Highly Efficient, Massively-Parallel Single-Cell RNA-Seq Reveals Cellular States and Molecular Features of Human Skin Pathology. *bioRxiv* (2019).
21. Mereu, E., Lafzi, A., Moutinho, C. & Ziegenhain, C. Benchmarking Single-Cell RNA Sequencing Protocols for Cell Atlas Projects. *BioRxiv* (2019).
22. Cusanovich, D. A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
23. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
24. Ding, J., Adiconis, X., Simmons, S. K. & Kowalczyk, M. S. Systematic comparative analysis of single cell RNA-sequencing methods. *BioRxiv* (2019).
25. Hug, H. & Schuler, R. Measurement of the number of molecules of a single mRNA species in a complex mRNA preparation. *J. Theor. Biol.* **221**, 615–624 (2003).
26. Fu, G. K., Hu, J., Wang, P.-H. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9026–9031 (2011).

27. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
28. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
29. Shiroguchi, K., Jia, T. Z., Sims, P. A. & Xie, X. S. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1347–1352 (2012).
30. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
31. Regev, A., Teichmann, S. A., Lander, E. S., Amit, I. & Benoist, C. Science forum: the human cell atlas. *Elife* (2017).
32. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
33. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
34. Gehring, J., Park, J. H., Chen, S., Thomson, M. & Pachter, L. Highly multiplexed single-cell RNA-seq for defining cell population and transcriptional spaces. *bioRxiv* (2018).
35. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
36. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, (2018).
37. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
38. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells

- Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342–357 (2016).
39. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
 40. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, (2016).
 41. Frieda, K. L. *et al.* Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
 42. Pei, W. *et al.* Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
 43. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
 44. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
 45. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469 (2018).
 46. Chan, M. M. *et al.* Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
 47. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882.e21 (2016).
 48. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016).
 49. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883–1896.e15 (2016).

50. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
51. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
52. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
53. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
54. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
55. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
56. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
57. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
58. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
59. Harada, A. *et al.* A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296 (2019).
60. Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519 (2015).
61. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285–289 (2015).

62. Reuter, J. A., Spacek, D. V., Pai, R. K. & Snyder, M. P. Simul-seq: combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat. Methods* **13**, 953–958 (2016).
63. Han, K. Y. *et al.* SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. *Genome Res.* **28**, 75–87 (2018).
64. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* **29**, 51–57 (2011).
65. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
66. Gundry, M., Li, W., Maqbool, S. B. & Vijg, J. Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic Acids Res.* **40**, 2032–2040 (2012).
67. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
68. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012).
69. Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402–412 (2012).
70. Chen, C. *et al.* Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* **356**, 189–194 (2017).
71. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).
72. Hu, Y. *et al.* Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* **17**, 88 (2016).

73. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26**, 304–319 (2016).
74. Cheow, L. F. *et al.* Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* **13**, 833–836 (2016).
75. Guo, F. *et al.* Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* **27**, 967–988 (2017).
76. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
77. Li, G. *et al.* Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* (2019). doi:10.1038/s41592-019-0502-z
78. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).
79. Darmanis, S. *et al.* Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Rep.* **14**, 380–389 (2016).
80. Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
81. Abdelmoez, M. N. *et al.* SINC-seq: correlation of transient gene expressions between nucleus and cytoplasm reflects single-cell physiology. *Genome Biol.* **19**, 66 (2018).
82. Hendriks, G.-J. *et al.* NASC-seq monitors RNA synthesis in single cells. *bioRxiv* 498667 (2018). doi:10.1101/498667
83. Erhard, F. *et al.* scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* (2019). doi:10.1038/s41586-019-1369-y
84. Singh, M., Al-Eryani, G., Carswell, S. & Ferguson, J. M. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes.

bioRxiv (2018).

85. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4259
86. Volden, R. *et al.* Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9726–9731 (2018).