## 全長 scRNA-Seq 解析による遺伝子融合、SNP および選択的スプライシングの 検出能向上

#### イントロダクション

臨床および応用分野でのトランスクリプトーム解析に次世代シーケンスを使用するには、多数のシングルセルを正確に並行処理でき、目的とするターゲットから堅牢なライブラリーを調製できる試薬が必要です。このニーズを満たすため、いくつかの自動プラットフォームが誕生していますが、その基盤となるテクノロジー(マイクロフルイディクス、ドロップレットによるカプセル化、マルチサンプル・ナノディスペンシングなど)の多様性は、シングルセルの単離・処理効率に影響を与えます。細胞処理技術に加えて、ライブラリー調製試薬の選択および感度(例:3′末端解析と、全長転写産物 [5′→3′] 解析のいずれを選択するか)によっても、遺伝子発現に関する個別のデータ品質に差が生じます。全長解析を選択すると、転写産物のカバレッジの偏りが小さいため、遺伝子融合、SNP 検出や選択的スプライシングの検討が可能である一方、3′末端解析による発現差異解析(3′DE)データは、遺伝子発現調節の検討に使用できます。

弊社の ICELL8 cx システムは、1 枚のチップ上に 1,200~1,500 個のシングルセルを自動でハイスループット処理することができ、一般的なマイクロフルイディクス法に付随する細胞サイズの制約や、ドロップレット技術ベースのシステムに付随する画像解析上の制約を受けません。ICELL8 cx CellSelect Software の高度な画像解析テクノロジーを用いることで、処理するウェルを柔軟かつ正確に選択可能で、このワークフロー自体を完全自動化することも可能です。いずれのオプションを選択する場合も、解析対象とするターゲット細胞を細胞染色により同定できます。この独自のアプローチにより、シーケンス解析データと個別のウェル(シングルセル、多核細胞、細胞クラスターまたはオルガノイドを含むウェル)との対応を確認することができます。信頼性の高い弊社の SMART-Seq 試薬と ICELL8 cx システムの組み合わせによる全長ライブラリーの構築は、シングルセルのトランスクリプトームについて豊富なデータが得られるハイスループットなソリューションを提供します。遺伝子を高感度に検出できるほか、細胞生物学のさらなる解明に不可欠なスプライスバリアント、遺伝子融合や突然変異に関する情報も得られます。

このテクニカルノートでは、図1で示された SMART-Seq ICELL8 Application Kit のプロトコールを用いて行った全長解析の結果を、X 社のドロップレット技術ベースのシステムによる 3'DE 法の結果と比較します。gene body coverage、遺伝子融合および SNP の検出、選択的スプライシングの同定といった指標を評価します。2つのテクノロジーの直接比較は行っていませんが、弊社の手法を用いた場合、全長配列の情報を取得できるため、gene body coverage が 3'DE 法よりも高く、その結果として、遺伝子融合の検出能、ClinVar データベースに病因 SNP としてアノテーションされている遺伝子位置のリードの深さ、全 SNP の検出能および細胞間でのスプライシングの差の評価能が向上しました。

# Α В Cell dispense On-chip RT-PCR

## SMART-Seq ICELL8 Application Kit のワークフロー

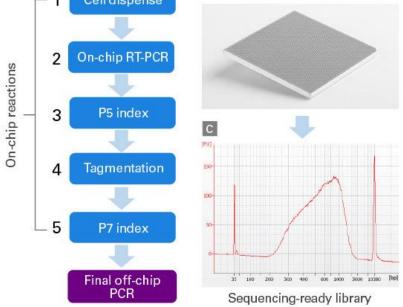


図 1. SMART-Seq ICELL8 Application Kit のワークフロー

パネル A. この所要時間 10 時間のプロトコールは 5 つの分注ステップからなります。ICELL8 ブランクチップの各ウェルに、培養細胞を 1 ウェルあたり平均 1 個となるように分注します (ステップ 1)。ICELL8 cx CellSelect Software を用いて、シングルセルを同 定します。細胞を溶解させた後、cDNA を合成・増幅させます(ステップ 2)。イルミナ社の Nextera Tagment DNA Enzyme を 用いて全長 cDNA のタグメンテーションを行い、格子 (grid) 状に付加したイルミナ社独自のインデックスアダプターとともに 増幅させます(ステップ3~5)。最終的なライブラリーをプールし、さらに増幅させた後、シーケンス解析前に精製します。

パネル B. 5,184 個のウェルを搭載したチップに、細胞、試薬、インデックスおよびタグメンテーション酵素を分注します。 チップには 72 x 72 通りのバーコードが印字されており、各ウェルはバーコードの組み合わせと 1 対 1 で対応しています。

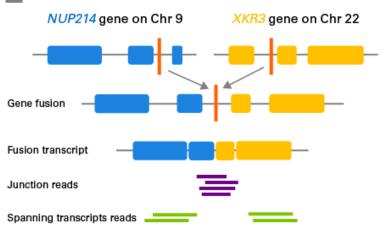
パネル C. シーケンス解析用ライブラリーの代表的な Bioanalyzer トレース

#### 遺伝子融合

2 つの手法による既知の遺伝子融合の検出能を評価するため、Chromium Single Cell 3' Library & Gel Bead Kit を用いた 10x Chromium システム(「X 社の 3'DE 法」)、または SMART-Seg ICELL8 Application Kit のワークフロー(「弊社の Full Length 法」)にて、K562 細胞 を処理しました。X 社の 3′DE 法により構築したライブラリー(シングルエンドリード)は、NovaSeq プラットフォームにて、X 社が推奨 するローディングおよびサイクリング条件に従ってシーケンス解析し、弊社の手法により構築したライブラリー(ペアエンドリード)は、 NextSeq High-Output カートリッジにて、イルミナ社が推奨するローディングおよびサイクリング条件に従ってシーケンス解析しました。

図 2 のパネル A に、NUP214 と XKR3 の融合を示します。 ジャンクションリードを紫色、 転写産物のリードを緑色で表示しています。 図 2のパネルBの表に示すように、X社の手法により構築したライブラリーでは、弊社の手法により構築したライブラリーで検出された5つ の融合遺伝子のうち、1 つが検出できたにすぎませんでした。この結果は、弊社の手法を用いた場合の融合遺伝子のペアエンドリードのカ ウント数によっても裏付けられています。弊社の手法では、既知の BCR--ABL1 融合も検出されました。





## В

Fusion name	Takara Bio Full Length		X社 3′DE		Left breakpoint	Right breakpoint
	Junction*	Spanning*	Junction*	Spanning*		
NUP214-XKR3	11	20	3	0	chr9:131199015:+	chr22:16808083:-
XACTLRCH2	5	0	0	0	chrX:113938193:-	chrX:115163783:-
XACTLRCH2	3	0	0	0	chrX:113938193:-	chrX:115156667:-
C16orf87ORC6	1	9	0	0	chr16:46824386:-	chr16:46696017:+
C16orf87ORC6	1	9	0	0	chr16:46824386:-	chr16:46693093:+
BCRABL1†	1	1	n/a	n/a	chr22:23290413:+	chr9:130854064:+

### 図 2. 融合遺伝子の検出(Takara Bio USA 社比較データ)

合計約 2,300 万リードのうち、1 ヵ所にのみマップされたリードの割合が 85%超であった K562 細胞のデータ。弊社の Full Length 法と X 社の 3'DE (v3) 法を比較しています。

\*ジャンクション全体のリード数("junction")またはジャンクションの両側に存在する転写産物のリード数("spanning")を示します。 †フィラデルフィア染色体陽性の K562 細胞については、弊社の Full Length 法により構築したライブラリーから既知の BCR--ABL1 融合が検出されました。(このサンプルについては、X 社の 3'DE 法による評価は実施していません。) 残りのデータは、BCR--ABL1 融合陰性の K562 細胞株を用いて収集したものです。

#### SNP の検出

2 つの手法による SNP 検出能を評価するため、X 社の 3'DE 法または弊社の Full Length 法を用いて 1,000 個の HEK293 細胞を処理し、それぞれ NovaSeq 又は NextSeq プラットフォームを用いて、ライブラリーのシーケンス解析を行いました。

図3では、カバレッジを生リード数として示しており、X 社の 3'DE 法のデータ (黒線) と弊社の Full Length 法のデータ (青線) の間に、全般的に明確な差が認められています。全転写産物の位置をノーマライズしたところ、X 社の手法によるデータのカバレッジは、3'末端領域に偏っていることが明らかになりました。これに対し、弊社の手法では、遺伝子全体で高いカバレッジが得られました。このため、ClinVarデータベース (約 18,000) で同定されている病因 SNP のうち、遺伝子の 5'末端側に選択的にマップされている SNP を含むトランスクリプトーム領域が理論上、カバーされることになります(灰色のカラム、右軸)。

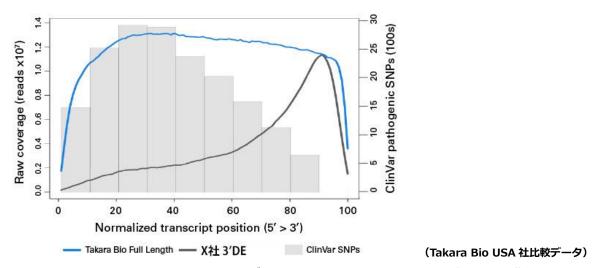


図 3. シーケンス解析のリード数および ClinVar データベースでアノテーションされている病因 SNP の分布

各キットを用いて構築した HEK293 細胞 1,000 個のライブラリーから取得したリードを、キットごとにプールしてマッピングし、病因 SNP が存在しうる遺伝子座(ClinVar データベースに基づく)におけるリードカバレッジを評価したところ、弊社の Full Length 法では、X 社の手法と比較して、SNP におけるカバレッジが高く(図 4)、特に、高いカバレッジ(>30X)では手法間の差がみられました。

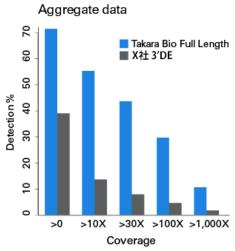


図 4. ClinVar データベースで病因 SNP としてアノテーションされている位置のリードカバレッジ(Takara Bio USA 社比較データ)

SNP 検出能の差は、転写産物の全長にわたるカバレッジ比を表示することにより可視化できます。図 5 に示すように、ノーマライズした転写産物の全長にわたる SNP のカバレッジは、弊社の手法で X 社の手法と比較して約 32 倍高値でした。X 社の手法では、3′末端側にカバ

レッジの偏りがみられ、このことから予想されたとおり、ノーマライズした転写産物の 3'末端から 10%の領域で、カバレッジ比の低下が みられました。ただし、弊社の手法は、3'末端側でも、X 社の手法より大幅に高いカバレッジを示しました。

ClinVar pathogenic SNP coverage ratio

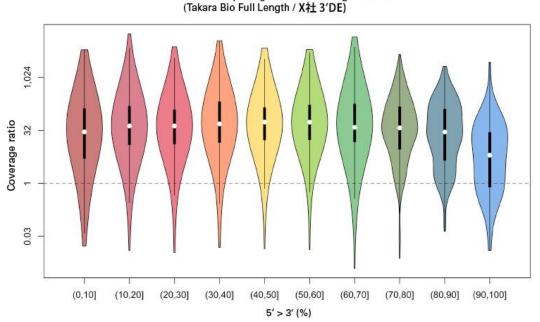
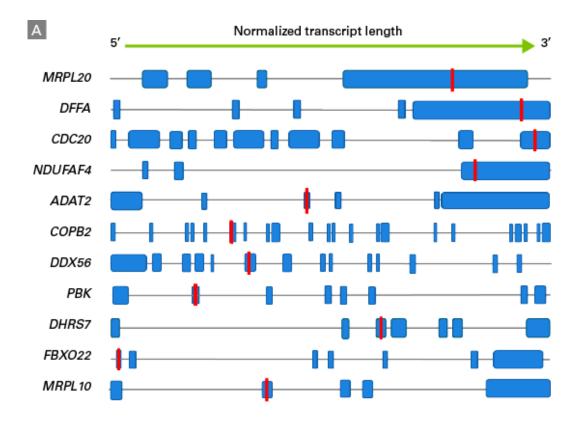


図 5. ClinVar データベースで病因 SNP としてアノテーションされている位置のリードカバレッジ: ノーマライズした転写産物の全長にわたるカバレッジ比(Takara Bio USA 社比較データ)

弊社の手法により 5'未端側で高いカバレッジが得られることの利点は、図6に示す SNP 検出数データに明確に表れています。図6のパネルAには、弊社の Full Length 法を用いて調製した K562 細胞サンプル中の全長転写産物から検出された 11 遺伝子の SNP の位置のスナップショット(ノーマライズされた長さ; エクソンの長さは縮尺どおりに記載されたものでない)を示します。図6、パネルBの表に示すように、X社の手法では、5'未端側の SNP の検出数が、弊社の手法による検出数を下回りました。



## В

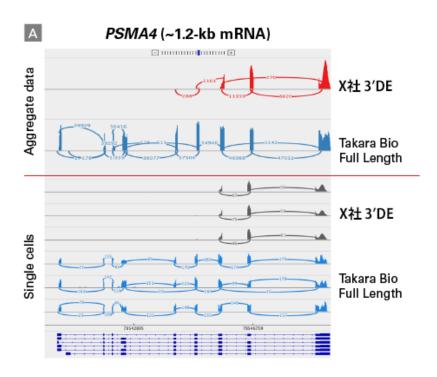
		Takara Bio Full Length		X社 3′DE		
Gene	Wt / variant	Total reads	Variant reads	Total reads	Variant reads	Location
MRPL20	G/C	662	163	127	27	Exon 4 (total 4)
DFFA	C/T	136	45	71	21	Exon 5 (total 5)
CDC20	G/A	161	84	144	86	Exon 11 (total 11)
NDUFAF4	A/C	225	74	98	42	Exon 3 (total 3)
ADAT2	A/C	107	25	17	5	Exon 3 (total 6)
COPB2	C/A	209	144	ND	ND	Exon 6 (total 22)
DDX56	G/A	108	32	ND	ND	Exon 6 (total 14)
PBK	C/G	160	55	ND	ND	Exon 2 (total 8)
DHRS7	T/C	175	173	ND	ND	Exon 3 (total 7)
FBXO22	C/T	169	55	ND	ND	Exon 1 (total 7)
MRLP10	G/T	231	137	ND	ND	Exon 2 (total 5)

#### 図 6. 全 SNP の検出数 (Takara Bio USA 社比較データ)

弊社の Full Length 法と X 社の 3'DE(v3)法を比較しました。パネル A. 11 個の転写産物に含まれるエクソン全体での SNP のおおよその位置。転写産物の長さはノーマライズしています。パネル B. 弊社の手法では、表に示す 11 遺伝子の SNP を検出できましたが、10x 法では、5'未端側の SNP のほとんどを検出不能(ND)でした。

#### 選択的スプライシング

プロテアソームサブユニットa4 (PSMA4; 約 1.2-kb の mRNA) および分枝鎖アミノ酸トランスアミナーゼ 1 (BCAT1; 約 9.6-kb の mRNA) 遺伝子座を例にとり、上記の HEK293 細胞のライブラリーに含まれるスプライシング変異を、Sashimi plot (IGV、version 2.4.10) の作成によって解析しました。図7、パネルAの集合データ(各ライブラリーから得られた全リード)が示すように、弊社の Full Length 法のデータからは、PSMA4 のスプライシングバリアントが複数存在することが十分に裏付けられました。しかし、X 社の手法のデータからは、mRNA が短い場合であっても遺伝子の 3'末端が検出できるにすぎず、スプライシングに関する情報が十分に得られませんでした。各ライブラリーに含まれる3つのシングルセルの解析でも、遺伝子の5'末端からのスプライシング情報の検出能に関して、2つのライブラリー間で差がみられました。図7、パネルBに、弊社の手法を用いて5つのシングルセルから発見されたBCAT1のスプライシングバリアントを示します。



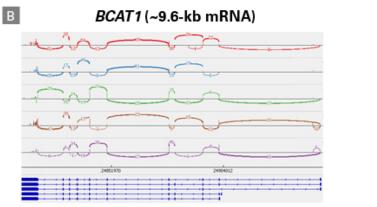


図 7. 選択的スプライシングの検出(Takara Bio USA 社比較データ)

パネル A. PSMA4 で検出されたスプライスバリアント (集合データおよびシングルセルレベル)。

パネル B. BCAT1 で検出されたスプライスバリアント (シングルセルレベル)

## 結論

ICELL8 cx プラットフォームは独自のナノウェルシステムであり、ドロップレット技術ベースのシステムと異なり、複雑な化学反応の進行に必要なサンプル・試薬間の様々な相互作用を可能にします。ICELL8 cx システムは、最大 8 種類のサンプルから細胞を分注後、最大 3 種類のスペクトルチャネルを搭載する顕微鏡を用いて、各ウェルの画像を取得することでシングルセルが入っているウェルを同定し、解析に使用します。取得した画像を用いて、細胞の生死判定およびレア細胞の検出が可能です。以降の解析に使用するウェルを自動または手動で選択し、試薬(SMART-Seq ICELL8 Reagent Kit の試薬など)が分注されます。ICELL8 cx システムの主な利点はオープンプラットフォームであることで、このため、シングルセル情報を迅速に照合できるハイスループット解析を開発することができます。

本テクニカルノートでは、自動化された全長 RNA-Seq 解析に供するライブラリーの構築に高感度な SMART-Seq 法を用いることにより、gene body coverage が高まり、融合、SNP およびスプライスバリアントの検出能向上が実現することを報告しました。これらのアプリケーションは、バイオマーカー同定や新規治療法の開発の進歩に役立つことが期待されます(Zhao 2019)。RNA 解析のその他のアプリケーションとしては、mRNA の 3'末端に焦点を絞った解析用試薬(ICELL8 3' DE Kit)や、T 細胞受容体のクロノタイプと 5'末端のトランスクリプトームのプロファイリングを組み合わせたハイスループット解析試薬(ICELL8 Human TCR Profiling Kit)などがあります。エピジェ

ネティック解析用には、新規の CUT&Tag 技術が開発されており(Kaya-Okur *et al.* 2019)、DNA 解析用には、ハイスループット CVP 解析法が現在開発中です。ICELL8 cx は柔軟なシステムであり、頑健な試薬セットや新たなバイオインフォマティクス解析ツールを組み合わせることで、高感度と高再現性を必要とする様々なシングルセル解析を可能にします。

#### 参考文献

Zhao, S. Alternative splicing, RNA-seq and drug discovery. Drug Discov. Today (2019).

https://doi.org/10.1016/j.drudis.2019.03.030

Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. bioRxiv 568915 (2019).

#### 製品情報

製品コード	製品名	容量
640190	ICELL8® cx Single-Cell System	走一
640222 640223 640224	SMART-Seq® ICELL8® cx Application Kit	For 1 Chip For 3 Chip For 5 Chip

#### © 2019 Takara Bio Inc. All Rights Reserved.

-イルミナ社とは、Illumina, Inc.(本社:米国カリフォルニア州サンディエゴ)の日本法人であるイルミナ株式会社(本社:東京都港区)をいいます。

本紙で紹介した製品はすべて研究用として販売しております。ヒト、動物への医療、臨床診断には使用しないようご注意ください。また、食品、化粧品、家庭用品等として使用しないでください。タカラバイオの承認を得ずに製品の再販または譲渡、およびこれらのための改変、商用製品の製造に使用することは禁止されています。

ライセンス情報については弊社ウェブサイトにてご確認ください。本紙に記載された社名および製品名は、特に記載がなくても各社の商標または登録商標です。

#### タカラバイオ株式会社

首都圏支店 TEL: 03-3271-8553 FAX: 03-3271-7282 関西支店 TEL: 077-565-6969 FAX: 077-565-6995

テクニカルサポートライン、受託窓口 TEL: 077-565-6999 FAX: 077-565-6995

ウェブサイト <a href="https://www.takara-bio.co.jp">https://www.takara-bio.co.jp</a>

# Clontech TakaRa cellartis