

# Omni-C™ Technology

制限酵素を使用しない全く新しいHi-Cで  
完全なゲノム構造の解析を



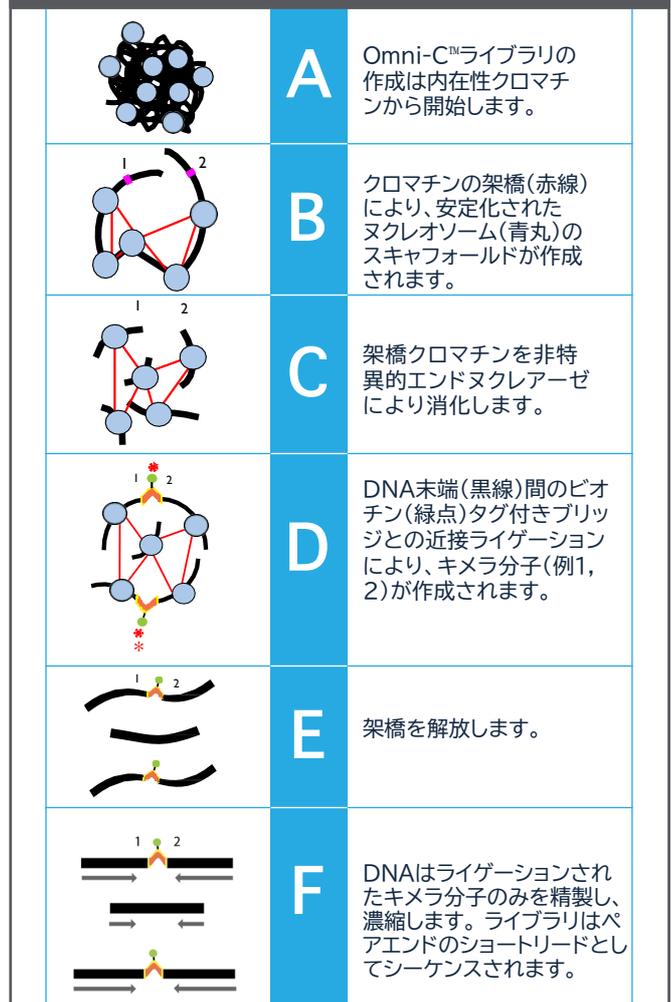
## 製品ハイライト:

- クロマチン構造の総合的解析
- 長距離情報を優れたカバレッジで捕捉
- SNP検出やハプロタイプフェージングなどのダウンストリーム解析

## はじめに

クロマチンキャプチャ法の発見と採用により、ゲノムトポロジーの研究が大幅に加速され、ゲノムアセンブリの品質が向上しました。Dovetail™ Hi-Cをはじめとするゲノムコンフォメーションマッピング技術は、シーケンシングにより、ゲノムの3次元(3D)構造をこれまでにない形で把握することを可能にしました。これらの手法により、遺伝子制御やエピジェネティクス研究に重要なトポジカル・アソシエーション・ドメイン(TAD)やクロマチン・ルーピングなどの3次元構造の研究が可能になりました。さらに、この3次元ゲノム構造は、染色体スケールでのコンティグのスキファールドディング情報を提供するため、ゲノムアセンブリに活用することができます。このように、Hi-Cはこれらの分野での研究を可能にしましたが、クロマチンの消化に**制限酵素(RE)**を使用するため、従来の方法には重大な欠点があります。それは、マッピング可能なヒトゲノムの約20%が、REサイトの密度が低いためにHi-Cを利用できず、Hi-Cデータの解析がこれらのREサイトの捕捉頻度に依存することです。ここでは、**配列に依存しないエンドヌクレアーゼ**をベースとしたDovetail Genomicsの最新近接ライゲーションプロトコルであるOmni-C™テクノロジーを紹介します(図1)。配列に依存しないエンドヌクレアーゼを採用することで、Omni-C™テクノロジーはREサイト密度に起因するバイアスを低減し、より幅広い領域のゲノムをカバーすることにより、各シーケンシングの効率を向上させ、近接ライゲーションアッセイのゲノムカバレッジを大きく向上させます。このように、Omni-C™ライブラリは、従来法と比べ、幅広く均一なゲノムカバレッジを提供し、SNPから、大規模なコンフォメーション変異まで、様々なゲノム構造の発見の可能性を高めます。

図1 - エンドヌクレアーゼベースの近接ライゲーション。



Omni-C™ワークフローは、内在性クロマチンの固定化(架橋)から始まります。架橋後、配列に依存しないエンドヌクレアーゼを用いてin situでクロマチン消化を行います。細胞を溶解して、消化したクロマチンを放出し、近接ライゲーション、クロスリンクの解放、ライブラリのアセンブリを行います。その結果として、イルミナ対応のOmni-C™ライブラリが得られます。

## 製品概要

Dovetail™ Omni-C™キットは、1キットに8反応分が含まれています。イルミナNGSプラットフォームを使用して、配列に依存しないエンドヌクレアーゼベースのHi-Cライブラリを作成し、ゲノム全体を均一にカバーします。Omni-C™アッセイは2日間のワークフローです(次ページ図2)。

- 1日目 - サンプル調製と近接ライゲーション
- 2日目 - シークエンスライブラリの作成

最高の反応効率を確保するために、Omni-C™プロトコルでは、最終的なライブラリの品質を予測する品質管理(QC)チェックを提供しており、標準的な分子生物学的手法により、プロセス中の3つのポイントでプロトコルの成否を簡単に評価することができます(QCステップの例については、図3Aを参照)。各QCチェックポイントは、プロトコルの重要な段階に設定されており、下流段階に進む前の明確な指標となるように設計されています。

このプロトコルは、サンプルに依存するクロマチン断片化のばらつきに対しても安定的になるように設計されており、様々な消化プロファイルにおけるライブラリの品質と、ゲノムカバレッジの安定性が実証されています(図3B, 図3C)。

図2 - Omni-C™ Kitの2日間ワークフロー。

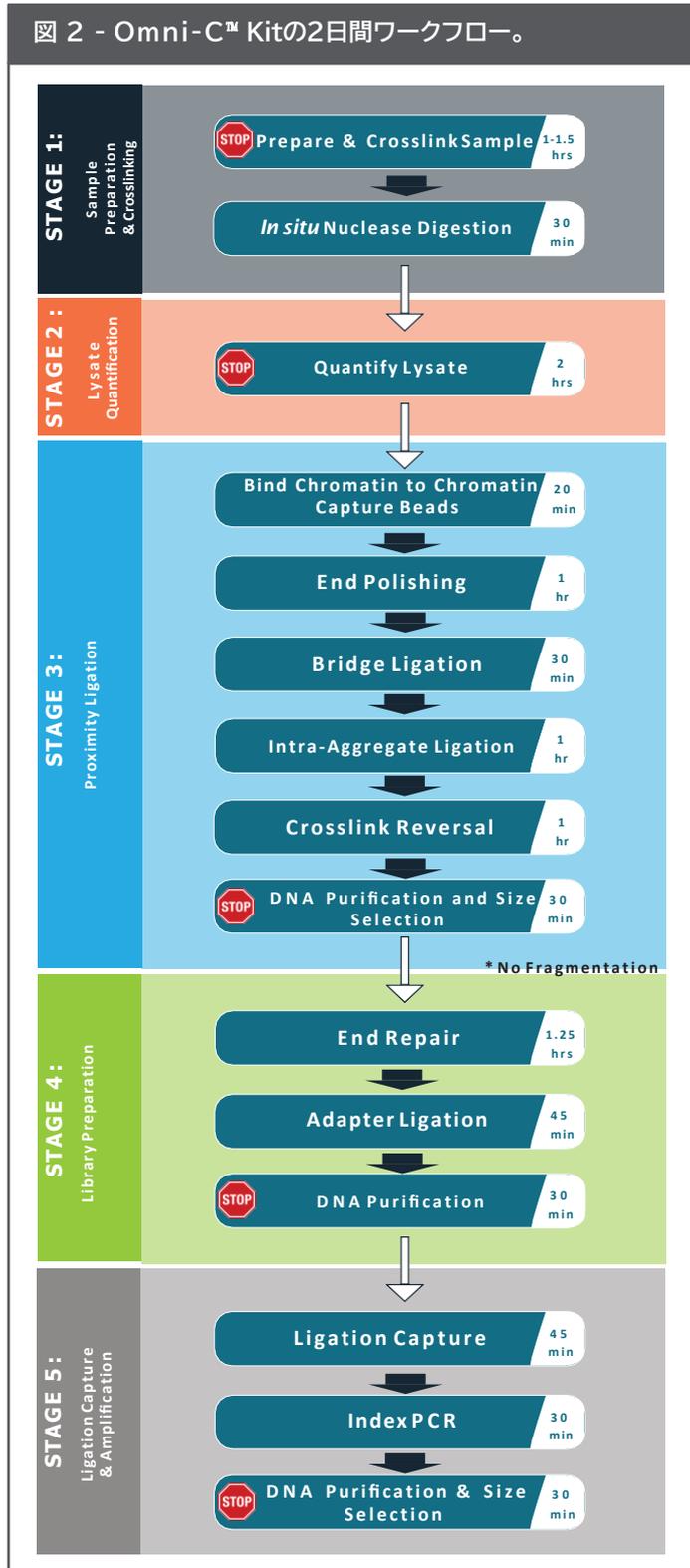
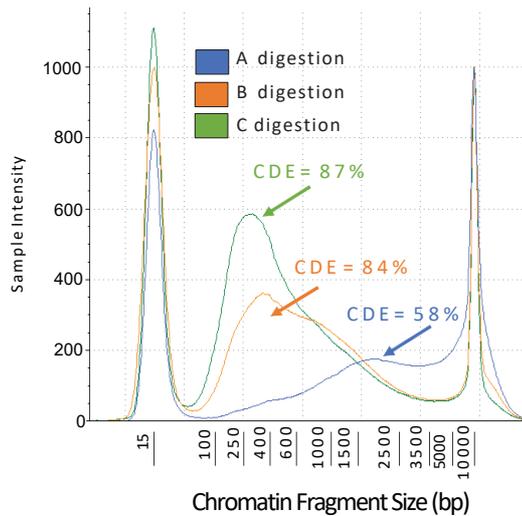
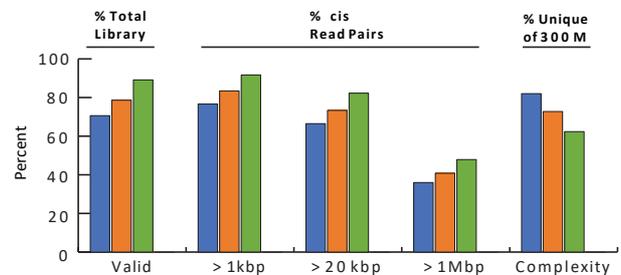


図3 - クロマチン消化におけるアッセイの堅牢性。

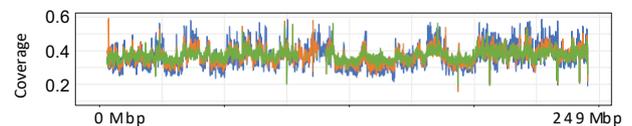
**A Tape Station Traces CDEQC**



**B Library Characteristics**



**C Coverage (chr1)**



ステージ2では、クロマチン消化効率(CDE)品質管理ステップにより、近接ライゲーション(図3A)の前にクロマチン断片サイズの分布を評価することができます。Omni-C™アッセイは、ライブラリ全体の品質(図3B)またはゲノムカバレッジ(図3C)を損なうことなく、幅広いCDEプロファイルを利用することができます。この例のOmni-C™ライブラリは、マウス組織から作成されたもので、組織に対する酵素の比率を変えて消化され、2,000M~4,000Mリードペア(2x150)でシーケンシング、Dovetail社のQCパイプラインで処理したものです。

Dovetail Omni-Cキットは、哺乳類細胞、組織、血液、昆虫、植物、海洋動物など、幅広いサンプルタイプで検証されています。Dovetail Omni-Cキットには、ライブラリの品質を評価するためのオープンソースのインフォマティクスQCツールが付属しています。またOmni-Cのデータは、自由に利用できるオープンソースのNGSおよびHi-C解析ツールと完全な互換性があります。

## データハイライト

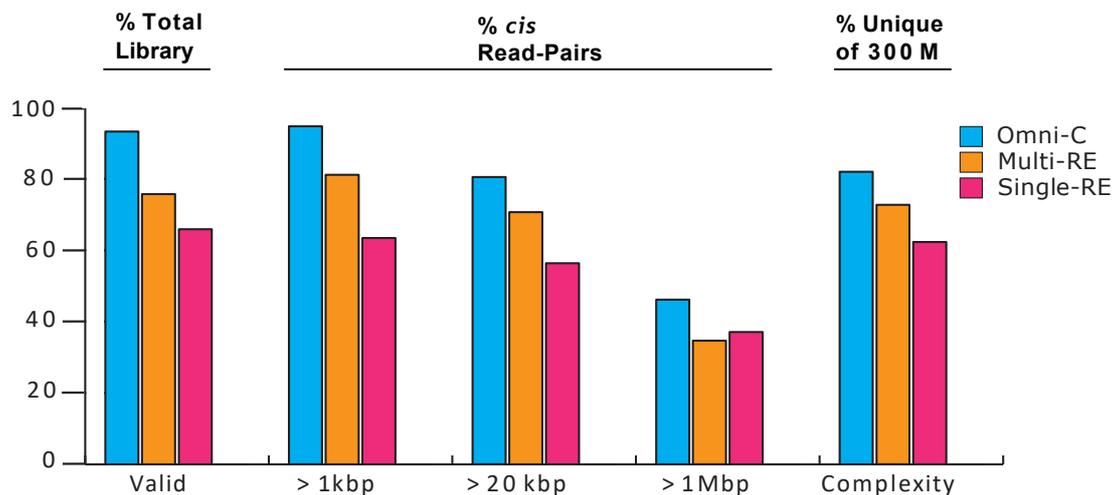
### 長距離情報の捕捉と優れたカバレッジ

2日間のOmni-C™ワークフローでは、有効なリードペアと長距離のシス情報に富む複雑度の高いHi-Cライブラリが作成できます(図3B, 4A)。

図 4 - Omni-C™ライブラリは長距離相互作用情報に富み、ショットガンのようなゲノムカバレッジを提供します。

近接ライゲーションに対するさまざまなアプローチを、有効リード (trans+cis>1kbp)、長距離cis相互作用情報 (>1kbp、>20kbp、>1Mbp)、300Mリードペアあたりのユニークリードの割合で複雑さを比較しています。ヒストグラムプロットは300Mリードペアにおけるショットガンライブラリへの様々なHi-Cアプローチのカバレッジ(塩基あたり)を比較しています。REサイト、GATC(DpnII, MboI, Sau3AI)とGANTC(HinfI)でのカバレッジは、REサイトの上流と下流の両方の絶対値の平均をそれぞれプロットしています。

### A Library Characteristics



### B Coverage Analyses

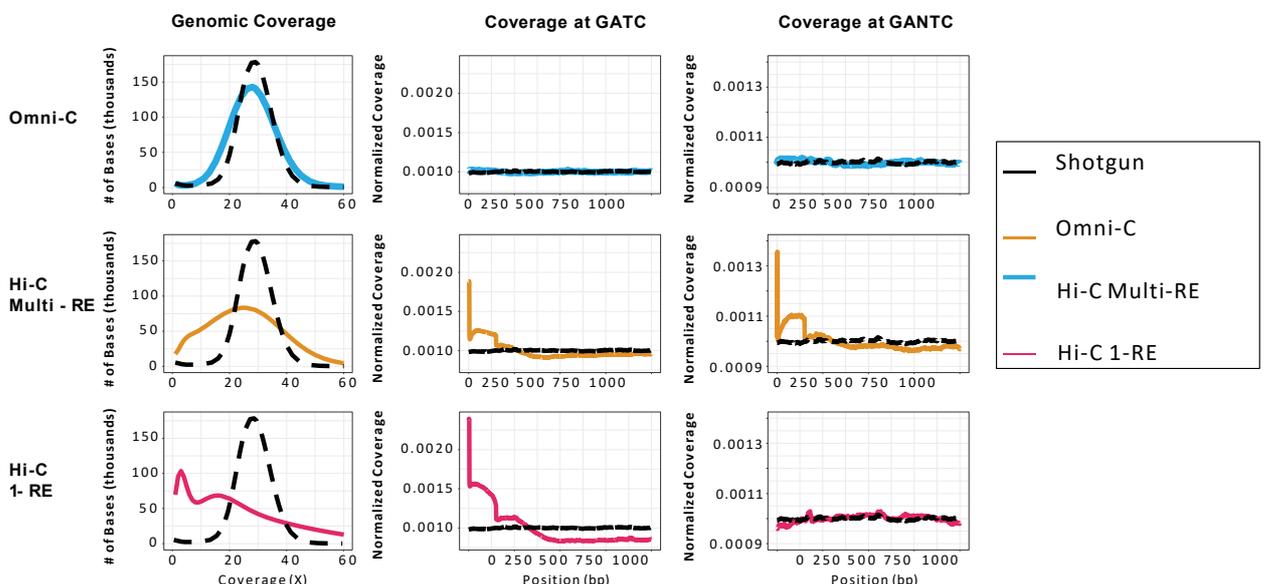


図 5 - Omni-C™アッセイは、幅広いサンプルに対応しています。

下記データのOmni-C™ライブラリは、ヒトとマウスの細胞、組織、血液から作成しました。ライブラリは2,000M~4,000M リードペア(2x150 bp)シーケンスし、Omni-C™ QCパイプラインを使用して評価しました。

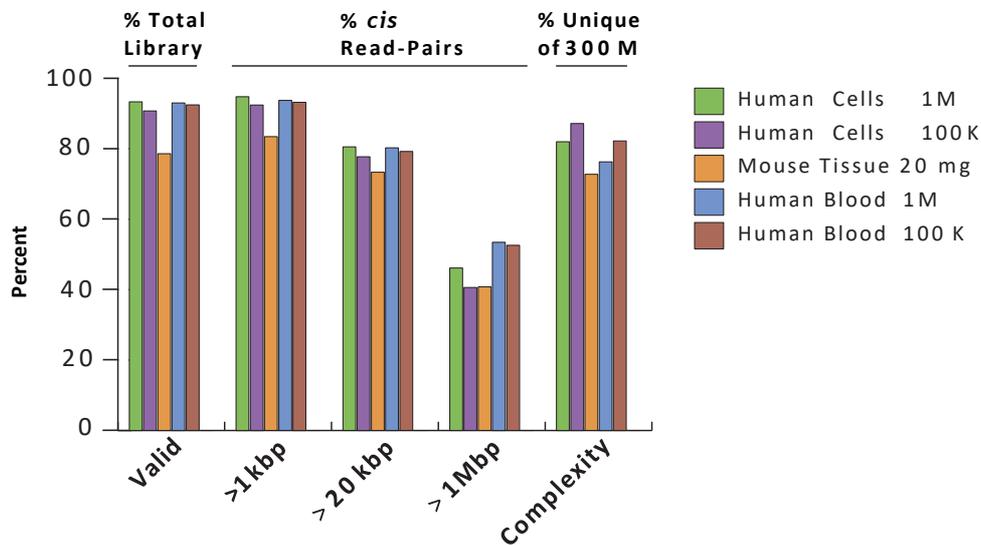
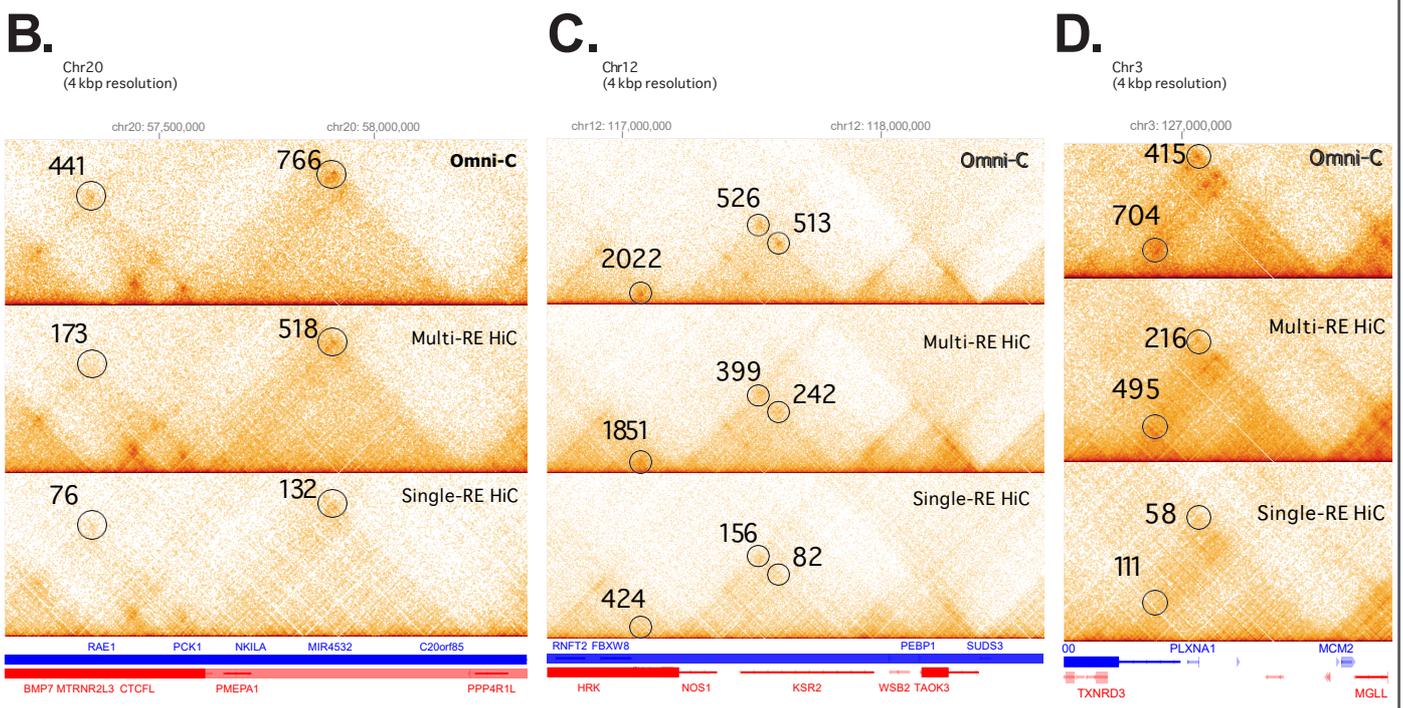
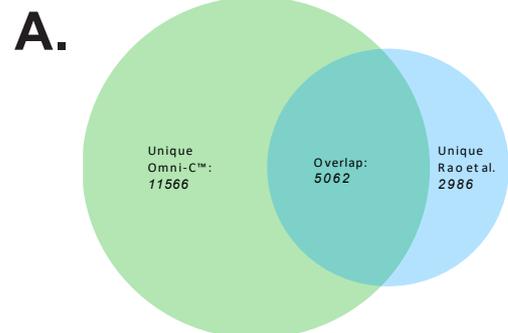


図 6 - Omni-C™アッセイは、独自の方法でクロマチンコンフォメーションの分解能を向上させ、長距離相互作用情報のリードサポートを向上させます。

A) 細胞株GM12878から作成したOmni-C™ライブラリを1.77Bリードペアシーケンスし、HiCCUPsを用いてループを検出しました。結果として得られたループを、Raoらが2014年に報告したループ(4.9Bリードペア)と比較したものです。

B-D) Omni-C™、マルチRE Hi-C、シングルRE Hi-Cライブラリのコンタクトマトリクスを、各ライブラリタイプの800Mリードから4kbの分解能で表示しています。ループは丸で示され、各コンタクトの生リードサポート数が示されています。注目すべきは、Omni-C™ライブラリにはゲノム全体にわたって優れたリードサポートが含まれていることです。



配列に依存しないエンドヌクレアーゼを使用することで、ゲノム全体にわたって均一なカバレッジが得られます(図4B)。Omni-C™データは、ショットガンライブラリのような、塩基あたりのカバレッジが狭いヒストグラムが得られますが、REベースのHi-Cアプローチでは、ゲノム全体のカバレッジが均一でないため、より広いヒストグラムとなります。さらに、REベースのHi-Cアプローチでは、Omni-C™ライブラリでは観測されない、REサイトでのオーバーエンリッチメントによる強いリードバイアスが見られます(図4)。このように、Omni-C™データは、REサイトの近接性に依存しない方法でシングルヌクレオチド情報を捕捉することができます。Omni-C™データに特有のカバレッジの向上により、ゲノムをより網羅的に観測することができ、SNPの検出や、ハプロタイプのフェージングなどのダウンストリーム解析が可能になります。

## 高い柔軟性

Dovetail™ Omni-C™キットは、高い柔軟性を考慮して設計されており、細胞、血液、組織など、幅広いサンプルに対応しています(図5)。標準的なワークフローでは、10万~100万個の細胞、または5~20mgの組織を必要としますが、低インプットワークフローでは、細胞数を1,000個まで減らすことができます(Application Note: Genome conformation profiling from as few as 1000 Cells)。

Omni-C™ライブラリでは、低インプットでも複雑性の高いライブラリが得られます(図5; Application Note)。またOmni-C™データは、ハイブリッドキャプチャなどのターゲットエンリッチメントにも対応しているため、シーケンシングのコストを軽減し、目的の部位周辺の分解能を向上させることができます(Application Note: Targeted Hi-C: Uncovering Local Genome Architecture in Regions of Interest)。

## トポロジー

このようにOmni-C™データでは、特有の均一なカバレッジと長距離相互作用情報の増加により、コンフォメーション構造の検出が強化されています。Rao et al., 2014に比べてリードペア数が1/3であるにもかかわらず、HiCCUPSによるGM12878のOmni-C™データからのループ検出では、5,062個のオーバーラップを持つ16,628個のクロマチンループが検出されました(図6A)。さらに、Omni-C™ライブラリから生成されたコンタクトマトリクスは、他のHi-Cアプローチと比較して、コンタクトあたりのリードサポートをより多く提供するため、より高解像度でのループ構造検出を可能にします(図6 B-D)。

図7 - Omni-C™ライブラリはより完全なコンタクトマトリクスを作成します。

コンタクトマトリクスの空白バンドは、リードカバレッジが低い領域におけるコンタクトマトリクスのバランス中に発生します。低カバレッジ領域では、分母にゼロを使用してコンタクトが正規化されるため、このような空白のバケトルが発生します。この図ではOmni-CデータとシングルRE Hi-Cを比較した3つのマトリクスを表示しています。例1: 5番染色体 がん感受性遺伝子座。肺がんでしばしば過剰発現しており、TERT遺伝子を含んでいます。例2: TRAF2遺伝子を包含する9番染色体の~5Mbp領域。TRAF2はアポトーシスシグナル伝達に重要な役割を果たしています。例3: 乳がんでしばしば過剰発現している疑いのあるがん遺伝子FASNを含む17番染色体。下部には、REサイトのない3kbp領域を示す青バンドがあり、黒矢印と赤矢印で遺伝子トラックを示しています。

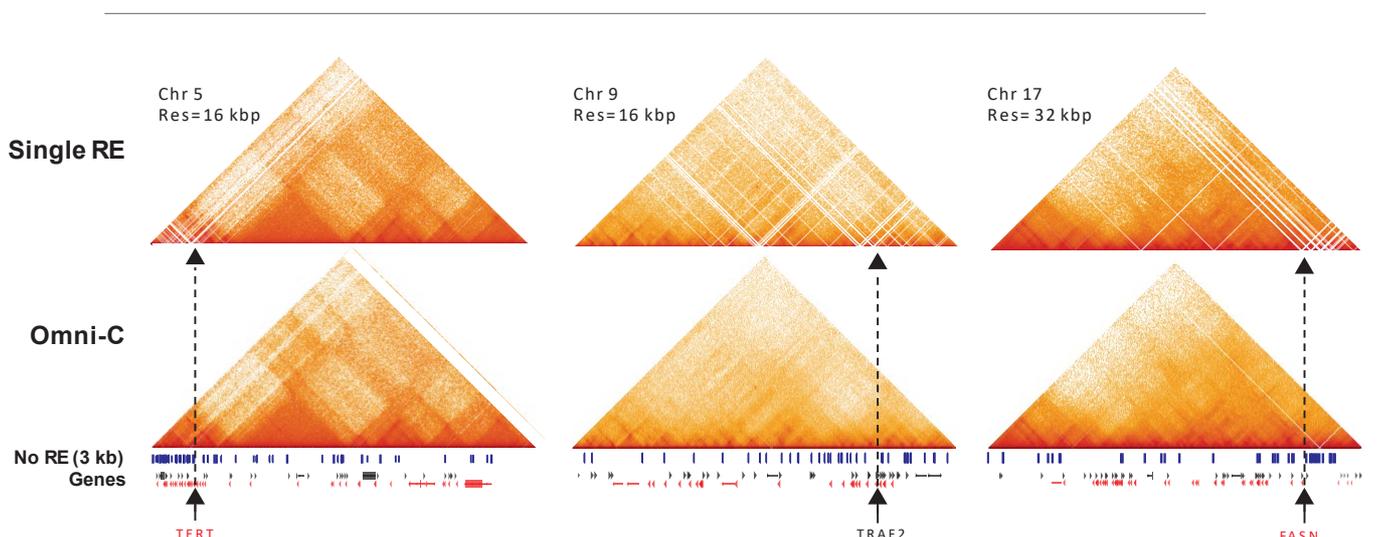
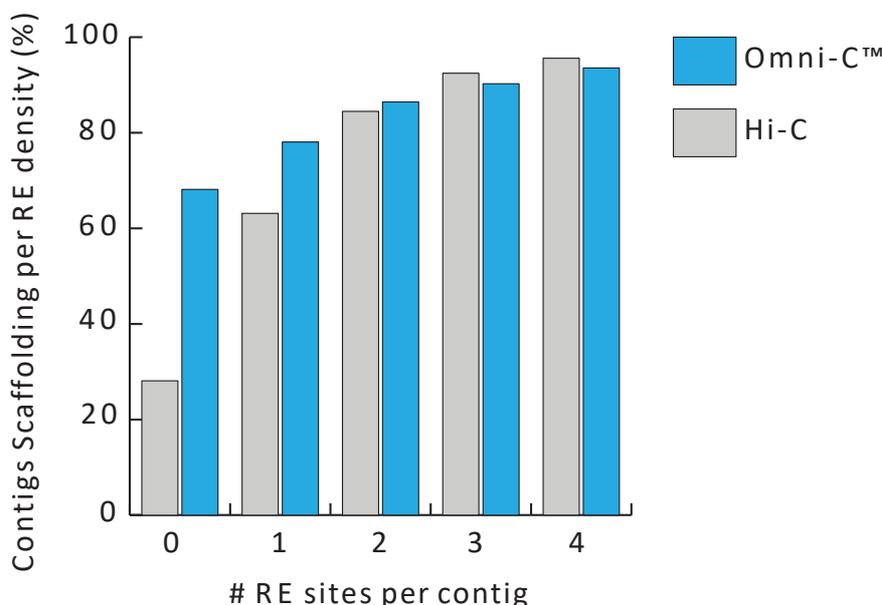


図 8 - Omni-C™は、REサイトの密度が低いコンティグのスキャフォールドを効率的に行うことができます。

ヒトゲノムHG38をランダムなサイズのコンティグに切り出し、GM12878からOmni-C™とREベースのHi-Cの両方を用いてライブラリを作成しました。コンティグをHiRise™アセンブリパイプラインを用いてスキャフォールディングし、コンティグあたりのREサイトの数によってグループ分けしました。スキャフォールディング効率、各REサイトグループ内のコンティグの総数でスキャフォールディングされたコンティグの数を正規化することによって決定しました。



コンタクトマトリクスのバランシング中に、カバレッジの低い領域があると、分母の値にゼロが割り当てられ、結果として空白のベクトルを持つコンタクトマップが作成されてしまいます。図7では、Omni-C™データが、REベースのHi-Cでは捉えられなかったトポロジカルな構造を明らかにしている3つの領域を示しています。REベースのHi-Cでは制限酵素サイト密度の低い領域では、トポロジカルな構造が見落とされています。これらを総合すると、Omni-C™ライブラリに特有のカバレッジの向上により、クロマチン相互作用のより完全な観測が可能であることがわかります。

Omni-C™データとREベースのHi-Cデータをそれぞれ使用して、ヒトゲノムをスキャフォールディングすると、ゲノム中でREサイトの頻度が増加している領域では、REベースのHi-CデータとOmni-C™データをそれぞれ用いたスキャフォールドの品質は類似していますが、REサイトが少ない領域では、REベースのHi-CではOmni-C™データには含まれるコンティグを見逃していることがわかります(図8)。

## まとめ

ここでは、クロマチンの断片化に配列に依存しないエンドヌクレアーゼを使用したDovetail™ Omni-C™ Kitのデータを紹介します。その堅牢なプロトコルは、品質管理ステップを提供し、ユーザーがシーケンス前にライブラリの品質を予測することを可能にします。Omni-C™ワークフローは、あらゆる分子生物学研究室で実行でき、低インプットサンプルを含む多種多様なサンプルに対応できます。Omni-C™データは、ゲノム全体を均一にカバーし、ゲノムトポロジーの包括的な検出を可能にします。

## スキャフォールディング

近接ライゲーションデータの主な用途は、ゲノムアセンブリのためのコンティグのスキャフォールディングです。Hi-Cデータにより正確にスキャフォールディングできるかどうかは、アセンブルされた各コンティグ内のREサイト密度に依存します。



# Dovetail

## GENOMICS

The Genome **Unrestricted.**



**Digital Biology**<sup>®</sup>

輸入/販売元 トミーデジタルバイオロジー株式会社  
phone: 03-6240-0843  
email: [info\\_ap@digital-biology.co.jp](mailto:info_ap@digital-biology.co.jp)  
<http://www.digital-biology.co.jp>