

サンプルのマルチプレックスおよび下流の解析におけるインデックスのミスアサインメントの影響

インデックスのミスアサインメントの原因と、インデックスホッピングの影響を軽減するベストプラクティス

はじめに

次世代シーケンス（NGS）技術の改良により、シーケンススピードが大幅に向上し、データ出力が飛躍的に増加したことで、現在のシーケンスプラットフォームにおいて大規模なサンプルの解析が可能になりました。10年前、Genome Analyzerは1回のランあたりのシーケンスデータの出力が最大1Gbでしたが、今日では、同様のコアテクノロジーに基づいたNovaSeq™システムにより、2日間で最大2Tbのシーケンスデータを生成できます。10年前と比較すると、データ出力量が2,000倍以上に増加しています¹。

この飛躍的に向上したデータ出力量を有効に利用するためには、マルチプレックス法がカギとなります。マルチプレックス法は、ライブラリー調製時に各DNA断片にインデックスと呼ばれるユニークな配列を付加することで行います。これによって、一回のシーケンスランで同時に多数のライブラリーをプールし、シーケンスできるようになります。マルチプレックス法によって得られたデータは、最終のデータ解析前に、プールしたそれぞれのライブラリーを、インデックスの配列情報によって、コンピューター上で振り分けるデマルチプレックスと呼ばれる複雑なプロセスを要します（図1）。

マルチプレックス法におけるライブラリー間のインデックスのミスアサインメントは、マルチプレックス法が開発された当初からシーケンスデータに影響すると知られていた問題です²。本書は、インデックスホッピングが起こりうるメカニズム、インデックスホッピングの測定方法、ならびにシーケンスデータの品質に関するインデックスホッピングの影響を低減させるためのベストプラクティスについて記述しています。

インデックスミスアサインメントの発生メカニズム

インデックスの組み換え、「インデックスホッピング」

Exclusion Amplification（ExAmp）ケミストリーおよび整列化フローセル技術の開発は、データ出力量の増加、コスト削減、ラン時間の短縮など、NGS技術に重要な進歩をもたらしました。この開発によって、1,000ドルゲノムを含む幅広いアプリケーションに対応できるようになりました³。しかし、整列化フローセルで用いるクラスター形成は、

従来のブリッジ増幅を用いたクラスター形成よりも高い割合でインデックスのミスアサインメントを起こすことが確認されました⁴。インデックスホッピングはインデックスのミスアサイ

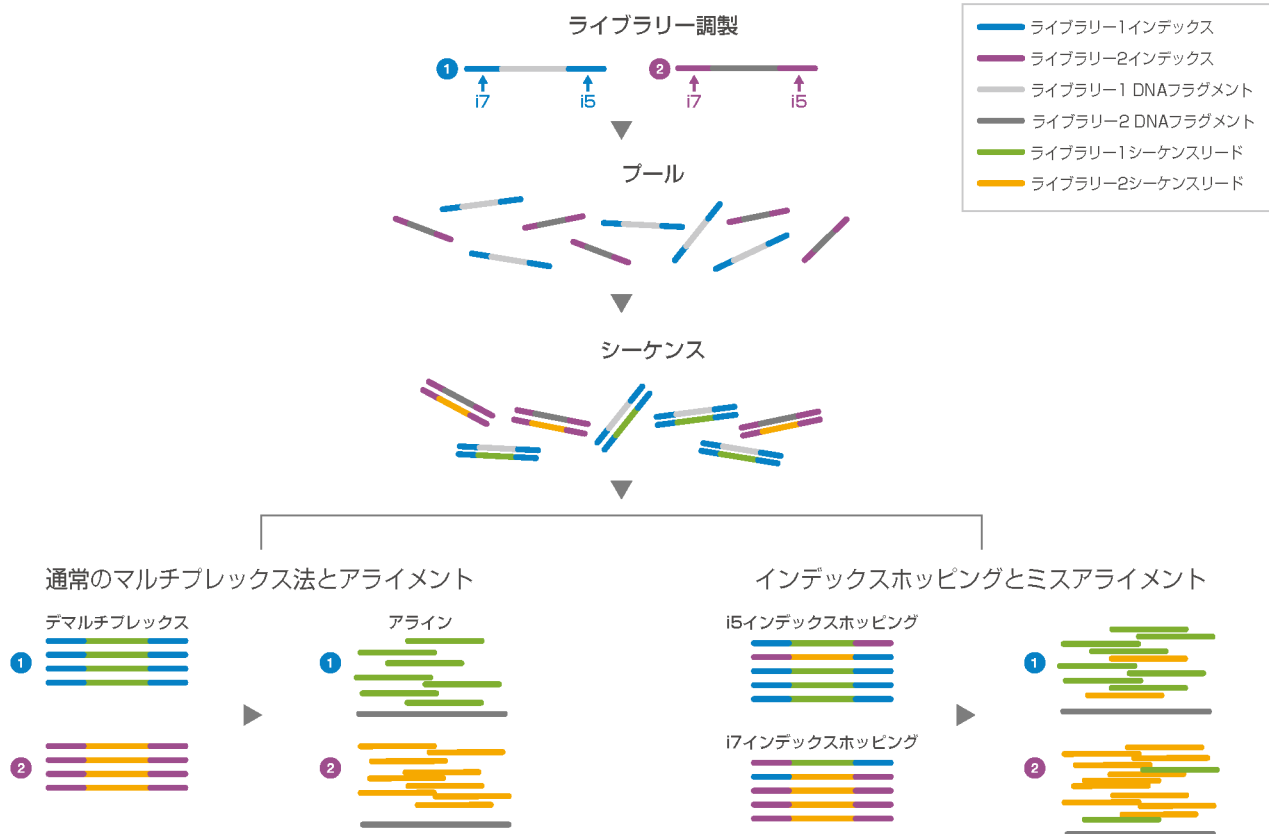


図1：マルチプレックス法とインデックスホッピングの概要—マルチプレックス法では、ライブラリー調製時に各DNA断片にユニークなインデックス配列を付加することにより、1回のシーケンスで同時に複数のライブラリーをランすることが可能となります。シーケンスリードはデマルチプレックスを行うことにより、それぞれのサンプル毎に振り分けられ、適切なアライメント結果が得られます。インデックスホッピングは、シーケンスリードの不正確なアサインメントを引き起こし、リードのミスアライメントまたは下流の解析における不正確なデータの解釈につながる可能性があります。

メントの原因となります。これによりシーケンスリードが本来のインデックスではない同一プール中の別のインデックスが付加されたライブラリーに誤ってアサインされる可能性があり、ミスアライメントおよび不正確な解析結果を引き起こすことにつながります（図1）。インデックスホッピングは整列化フローセル中のインデックスのミスアサインメントの増加を引き起こす主な原因となります。

遊離アダプターまたはプライマーの混入

アダプターを核酸断片に結合した後、遊離した非結合アダプターを除去するためにライブラリーを精製します。ライブラリー精製は、ビーズを用いた方法またはゲル精製法によって行い、遊離したアダプターやプライマーを除去することができます。遊離したアダプターやプライマーの除去が十分でないと、調製したライブラリーに混入することになり、インデックスホッピングおよびインデックスのミスアサインメントを引き起こす可能性があります。この可能性を検証するために、アダプターを除去したライブラリープール中に、DNAインプット量に対してモル濃度で0~35%の異なる濃度のアダプターを混合しました。インデックスホッピングの割合は、混合したアダプターの増加量に一致して直線的に増加しました。（図2）この結果より、調製したライブラリーを、シーケンスランを行う前に確実に精製することの重要性が示されています。

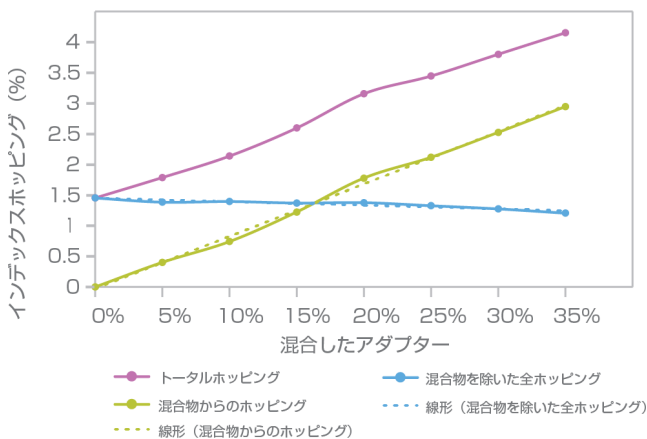


図2：遊離アダプターによるインデックスホッピング—インデックスホッピングの割合をアダプター混合量に対してプロットしています。全インデックスホッピング（赤線）と混合した遊離アダプター量（黄線）との正の相関関係が示されています。

インデックスホッピング発生頻度の測定

ライブラリーをプールした実験から、インデックスホッピングの割合を定量化することができます。ユニークなペアであるi5およびi7インデックスアダプターを用いて、dual indexライブラリーを作成し、それぞれのライブラリーをプールし、シーケンスを行った後、デマルチプレックスを行いました。全ての想定されるアダプターの組み合わせのうち、無効な（サンプルに使用されていない）組み合わせのインデックスホッピングの割合を%で示しました（図3）。例えば、0.17%という値は600対の正確なインデックスペアあたり約1つのインデックスホッピングが発生することを意味しています。

		i7インデックス							
		701	702	703	704	705	706	707	708
i5インデックス	501	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	502	0.00%	26.20%	0.00%	0.11%	0.14%	0.14%	0.00%	0.00%
	503	0.00%	0.17%	0.00%	0.10%	23.41%	0.12%	0.00%	0.00%
	504	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	505	0.00%	0.15%	0.00%	22.91%	0.12%	0.16%	0.00%	0.00%
	506	0.00%	0.14%	0.00%	0.12%	0.12%	23.37%	0.00%	0.00%
	507	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	508	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

図3：ユニークなインデックスを用いたコンタミネーション—全アダプターの組み合わせのインデックスホッピングの割合（%）。有効（緑）および無効（赤）な組み合わせは、それぞれ緑と赤でハイライトされています。インデックスホッピングの発生率は、インデックスの組み合わせによって、偏りを生じることはありません。

インデックスホッピングの影響

ライブラリー調製方法はインデックスホッピング率に影響することが示されています。一般的に、TruSeq[®] DNA PCR-Free Library Prep Kitなどのライゲーションのみを行いライブラリー調製を行う方法では、TruSeq Nano DNA Library Prep KitのようなPCR増幅のステップを含むライブラリー調製法よりも、インデックスホッピングの割合が高いライブラリーを形成します（図4）。従来のブリッジ増幅による不均一化フローセル上にクラスター形成したライブラリーは、インデックスホッピング率（1%）が、ExAmpのクラスター形成による整列化フローセル上のライブラリーをランしたインデックスホッピング率（2%）と比較して低いことが認められます。例えば、TruSeq PCR-Freeライブラリーのシーケンスでは、整列化フローセルよりも不均一化フローセル上で低いインデックスホッピング率が示されています（図4）。

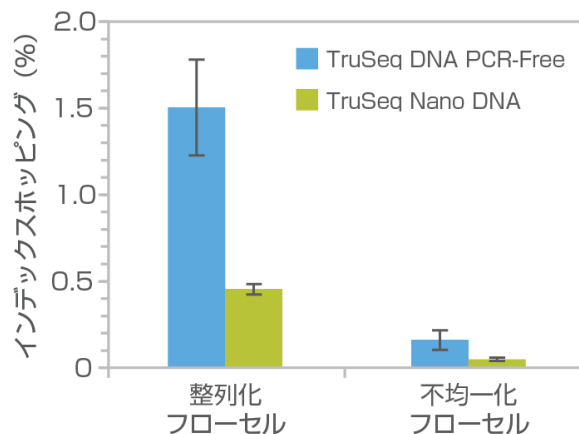


図4：インデックスホッピングの発生率の差—インデックスホッピングの割合は、ライブラリー調製方法に関わらず、不均一化フローセルよりも整列化フローセルが高いことを示しています。PCR増幅のステップを含むライブラリー調製方法（例、TruSeq Nano）は、ライゲーションのみ方法（例、TruSeq DNA PCR-Free）と比較して低いインデックスホッピング率を示します。

RNAシーケンス実験でのインデックスホッピングの影響

非常に高い発現マーカーが存在するサンプルのRNAシーケンス（RNA-Seq）に関してインデックスホッピングの一般的な影響の程度を示すために、stranded mRNAライブラリーを異なる2種類のヒト組織のトータルRNAサンプルから調製しました。こ

ここでは、組織特異的のマーカ―の発現が非常に豊富な組織（肝臓）と、特異的なトランスクリプトに偏らない分散型の発現プロファイルを示す組織（脳）を選択しました。

ライブラリーはTruSeq Stranded mRNA Library Prep Kitを用い、プロトコールに従ってライブラリーを調製しました。サンプルは、ユニークなインデックスセットを付加し、インデックスホッピングを別々に測定しました。HiSeq 4000システムを用いて、肝臓と脳のサンプルをミックスしたレーン、または肝臓サンプルのみ、脳サンプルのみの組織別にプールのレーンで、6プレックスのランを実施しました。

シーケンスデータでデマルチプレックスを行った後、BaseSpace Sequence HubのRNA Express Appと標準的な解析パイプラインを用いて解析を行いました。インデックスホッピング率は解析したレーンにおいて0.3~0.5%と測定されました。FPKM（遺伝子発現強度の単位）の遺伝子発現プロットでは、組織サンプルを混合したレーンの脳サンプルにおいて、アルブミン（肝臓中120,000~950,000カウント）のような肝臓の強発現マーカ―遺伝子が検出されました。これは脳サンプルのみをシーケンスしたレーンには見られていないため、インデックスホッピングによって生じた現象であると考えられます（図5上）。組織サンプルを混合したレーンの脳サンプルにおいて認められたこれらの肝臓マーカ―は、肝臓サンプル中で認められたレベルの~0.13%であることがわかりました。肝臓組織とともにシーケンスした脳サンプルの反復実験を比較解析したFPKM遺伝子発現プロットでは、サンプル間での特異的な発現の差は見られず、両者が同等のバックグラウンドノイズを示すことを表しています（図5下）。これらの結果よりインデックスホッピングの影響を最小にするためには、同類のサンプルを一緒にプールのことが最適であり、それによって優位に高発現する転写産物へのインデックスホッピングの解析における影響を軽減することができます。

インデックスホッピングを減少させるためのベストプラクティス

インデックスホッピングの影響を低減させるために、シーケンサーシステムによる特別な推奨方法、ライブラリー調製ワークフローおよびアプリケーションが特定されています。インデックスホッピングの影響を減少させるための一般的なガイドラインおよび推奨方法を示します（表1）。

推奨する条件以外で調製したライブラリーの保存（表1）は、インデックスホッピング率を増加させることが示されています。それぞれのライブラリーは-20℃で保存し、4℃での保存は避けてください。プールの後、できるだけ早くライブラリーをシーケンスするまたは-20℃で保存することで、インデックスホッピングが低減します。

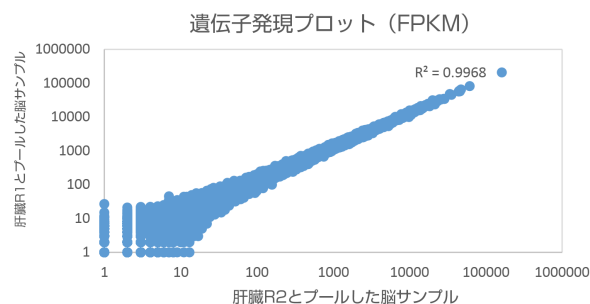
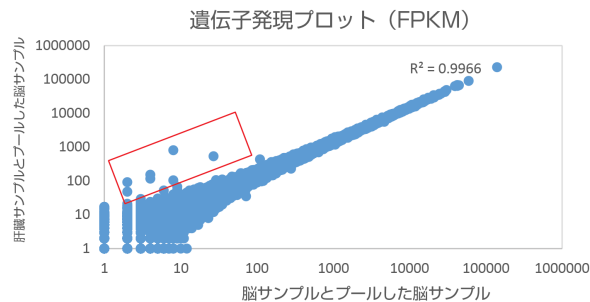


図5：RNA-Seq解析におけるインデックスホッピングの影響—肝臓および脳からのトータルRNAライブラリーを6プレックスでHiSeq 4000システムを用いて、シーケンスしました。組織サンプルを混合、もしくは別々にプールのレーンでシーケンスを行い、比較解析を行ったFPKM発現プロットを示します。組織サンプルをプールのレーンの脳サンプルで、非常に高く発現する肝臓マーカ―遺伝子の検出（赤色の囲み内）は、インデックスホッピングの発生を示しています。下段のプロットで、組織サンプルを混合したレーンの反復実験の比較解析の発現プロファイルではほとんど影響がないことが示されています。

表1：インデックスホッピングを低下させるベストプラクティス

現象の発生を軽減する方法/推奨方法	利点/結果
ユニークなインデックスを使用し dual indexライブラリーを調製 ^a	インデックスホッピングが生じたリードをUndetermined readをして分類
1レーン当たり30カバレッジでヒトゲノムをシーケンス ^b	サンプルのプールとインデックスホッピングを回避
アダプターの除去（精製、スピニングなど） ^c	インデックスホッピング率の低減
推奨温度-20℃で調製ライブラリーを保存 ^c	インデックスホッピング率の低減
同様のRNA-Seqサンプルを混合	高発現遺伝子および低発現遺伝子の混入を低下

a. HiSeq Xシリーズのシーケンスシステムではサポートされていません。
b. HiSeq Xシリーズのシーケンスシステムのみ可能です。
c. [TruSeq Sample Preparation Best Practice](#)および[Troubleshooting Guide](#)をご覧ください。

デュアルインデックスシーケンスのためのサンプルプールガイドライン

TruSeq High-Throughput (HT) Library Prep Kitには、キットによってDNAアダプタープレート（DAP）またはRNAアダプタープレート（RAP）のいずれかが含まれています。アダプタープレートは、96のユニークなインデックスアダプターのコンビネーションを含む96ウェルプレートで、最大96のユニークなインデックス化ライブラリーをマニュアルまたは自動で調製するためにデザインされています。イルミナは、アダプタープレートの利用を最大にし、インデックスホッピングを低減または同定するために用いる、12種類の8プレックスコンビネーション（表2）、または16種類の6プレックスコンビネーション（表3）に

表2：8ブックスコンビネーションのためのプールガイドライン

1		2		3		4		5		6	
アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位
D501-D705	A5	D502-D706	B6	D503-D701	C1	D505-D702	E2	D506-D704	F4	D507-D703	G3
D502-D704	B4	D501-D702	A2	D505-D703	E3	D503-D706	C6	D507-D705	G5	D506-D701	F1
D503-D703	C3	D505-D705	E5	D506-D706	F6	D507-D701	G1	D504-D702	D2	D508-D704	H4
D505-D701	E1	D503-D704	C4	D507-D702	G2	D506-D703	F3	D508-D706	H6	D504-D705	D5
D506-D710	F10	D507-D712	G12	D504-D707	D7	D508-D708	H8	D501-D709	A9	D502-D711	B11
D507-D709	G9	D506-D708	F8	D508-D711	H11	D504-D712	D12	D502-D710	B10	D501-D707	A7
D504-D711	D11	D508-D710	H10	D501-D712	A12	D502-D707	B7	D503-D708	C8	D505-D709	E9
D508-D707	H7	D504-D709	D9	D502-D708	B8	D501-D711	A11	D505-D712	E12	D503-D710	C10
7		8		9		10		11		12	
アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位	アダプターペア	座位
D501-D710	A10	D502-D712	B12	D503-D707	C7	D505-D708	E8	D506-D709	F9	D507-D711	G11
D502-D709	B9	D501-D708	A8	D505-D711	E11	D503-D712	C12	D507-D710	G10	D506-D707	F7
D503-D711	C11	D505-D710	E10	D506-D712	F12	D507-D707	G7	D504-D708	D8	D508-D709	H9
D505-D707	E7	D503-D709	C9	D507-D708	G8	D506-D711	F11	D508-D712	H12	D504-D710	D10
D506-D705	F5	D507-D706	G6	D504-D701	D1	D508-D702	H2	D501-D704	A4	D502-D703	B3
D507-D704	G4	D506-D702	F2	D508-D703	H3	D504-D706	D6	D502-D705	B5	D501-D701	A1
D504-D703	D3	D508-D705	H5	D501-D706	A6	D502-D701	B1	D503-D702	C2	D505-D704	E4
D508-D701	H1	D504-D704	D4	D502-D702	B2	D501-D703	A3	D505-D706	E6	D503-D705	C5

表3：6ブックスコンビネーションのためのプールガイドライン

1		2		3		4		5		6		7		8	
アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル
D501-D705	A5	D501-D710	A10	D502-D704	B4	D502-D709	B9	D503-D703	C3	D503-D711	C11	D505-D701	E1	D505-D707	E7
D502-D706	B6	D502-D712	B12	D501-D702	A2	D501-D708	A8	D505-D705	E5	D505-D710	E10	D503-D704	C4	D503-D709	C9
D503-D701	C1	D503-D707	C7	D505-D703	E3	D505-D711	E11	D506-D706	F6	D506-D712	F12	D507-D702	G2	D507-D708	G8
D505-D702	E2	D505-D708	E8	D503-D706	C6	D503-D712	C12	D507-D701	G1	D507-D707	G7	D506-D703	F3	D506-D711	F11
D506-D704	F4	D506-D709	F9	D507-D705	G5	D507-D710	G10	D504-D702	D2	D504-D708	D8	D508-D706	H6	D508-D712	H12
D507-D703	G3	D507-D711	G11	D506-D701	F1	D506-D707	F7	D508-D704	H4	D508-D709	H9	D504-D705	D5	D504-D710	D10
9		10		11		12		13		14		15		16	
アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル	アダプターペア	ウェル
D506-D710	F10	D506-D705	F5	D507-D709	G9	D507-D704	G4	D504-D711	D11	D504-D703	D3	D508-D707	H7	D508-D701	H1
D507-D712	G12	D507-D706	G6	D506-D708	F8	D506-D702	F2	D508-D710	H10	D508-D705	H5	D504-D709	D9	D504-D704	D4
D504-D707	D7	D504-D701	D1	D508-D711	H11	D508-D703	H3	D501-D712	A12	D501-D706	A6	D502-D708	B8	D502-D702	B2
D508-D708	H8	D508-D702	H2	D504-D712	D12	D504-D706	D6	D502-D707	B7	D502-D701	B1	D501-D711	A11	D501-D703	A3
D501-D709	A9	D501-D704	A4	D502-D710	B10	D502-D705	B5	D503-D708	C8	D503-D702	C2	D505-D712	E12	D505-D706	E6
D502-D711	B11	D502-D703	B3	D501-D707	A7	D501-D701	A1	D505-D709	E9	D505-D704	E4	D503-D710	C10	D503-D705	C5

対する最適なサンプルプールガイドラインを定めています。これらのユニークなインデックスの組み合わせによって、二次解析時にミスアサインされたリードを除去することが可能となります。ミスアサインリードは、「unaligned reads」としてフラグされ、アライメントから除かれます。ユニークなdual indexコンビネーション（表2、3）を用いることは、不正確なインデックスを伴うリードがバリエーションコールまたは遺伝子発現カウントの

サインメントに確実に影響しないためのベストプラクティスです。

まとめ

マルチプレックス法は、NGS技術の大きな進展と重要性を代表するものであり、これによってサンプル解析量の飛躍的な増加が実現します。しかし、マルチプレックス法を用いた場合、ライブラリー調製方法または使用するシーケンサーシステムに関わらず、インデックスホッピングの可能性が存在します。インデックスホッピングは、デマルチプレックス時に間違ったインデックスにシーケンスリードのアサインメントを引き起こす可能性があります。それによってミスアライメントおよびデータの品質にネガティブな影響を与える可能性があります。インデックスホッピングの検証の結果、ほとんどのアプリケーションの解析結果に、大きな影響を与えないことが示されています。インデックスホッピングの恒久的な解決方法は開発中ですが、本書はインデックスホッピングを最小にするためのガイドラインとベストプラクティスを提供するものです。

参考文献

1. Illumina.An Introduction to Next-Generation Sequencing Technology.2016.Accessed April 2017.
2. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform.*Nucleic Acids Res.*2012;2513-2524.
3. Illumina.HiSeq X Series of Sequencing Systems.2016.Accessed April 2017.
4. Illumina.Illumina Sequencing Technology.2010.Accessed April 2017.

イllumina株式会社

〒108-0014 東京都港区芝5-36-7 三田ベルジュビル22 階
Tel (03) 4578-2800 Fax (03) 4578-2810
jp.illumina.com

 www.facebook.com/illumina

代理店

本製品の使用目的は研究に限定されます。 販売条件：jp.illumina.com/tc

© 2017 Illumina, Inc. All rights reserved.

Illumina, BaseSpace, BeadArray, BeadXpress, cBot, CSeq, DASL, Design Studio, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Innuity, iSelect, MiSeq, Nextera, NextSeq, NovaSeq, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は、Illumina, Inc. の商標または登録商標です。

その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様および希望販売価格を変更する場合があります。

Pub.No. 770-2017-004-A-10MAY2017-JPN

本製品の使用目的は研究に限定されます。診断での使用はできません。

