

NGS超入門

MiSeqシステムのご紹介

イルミナ株式会社
サービス・サポート部
仲 健太

© 2014 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPPro, CytoChip, DesignStudio, Epicentre, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, ForenSeq, MiSeq, MiSeqDX, MiSeqFGx, NeoPrep, Nextera, NextBio, NextSeq, Powered by Illumina, SeqMonitor, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

illumina®

イルミナ iSchool – 開講

illumina®

Customer Login | Myillumina | 日本語のサポート | お問い合わせ

研究分野 ▾ 研究手法 ▾ システム ▾ 製品とサービス ▾ インフォマティクス ▾ サイエンス ▾ カンパニー ▾ サポート ▾

SEARCH

変革を求めるすべての人に
私は 遺伝的な答えをもたらしたい

詳細はこちらから

がんゲノム

リプロダクティブヘル
ス

法医学ゲノム

アグリゲノム

疾患ゲノム

微生物ゲノム

ゲノム創薬



イルミナ iSchool

イルミナ iSchool



イルミナキャンペーン



Miniで、Bigに当てよう！

イルミナ iSchool – 講習会

▶ ベーシック講習会（初級）

- 「キャピラリーDNAシーケンサーやマイクロアレイは利用したことはあるけど、次世代シーケンサーを利用したことがない...」「次世代シーケンサーを試してみたいが、何ができるのか分からない...」「次世代シーケンスを始めたいが、教えてくれる人が周りにいなくて困っている...」など、次世代シーケンサーの初学者の方が、気兼ねなく参加できる入門編の講習会。

▶ ハンズオン講習会（中級・上級）

- イルミナの次世代シーケンサーがサポートする幅広いアプリケーションの内、人気のアプリケーションについて、デモデータを用いて操作や解析内容をご理解いただく実践形式の無料講習会。遺伝疾患の変異解析、遺伝子発現解析、微生物ゲノム解析などを取り上げ、研究を前進させるお手伝いをいたします。

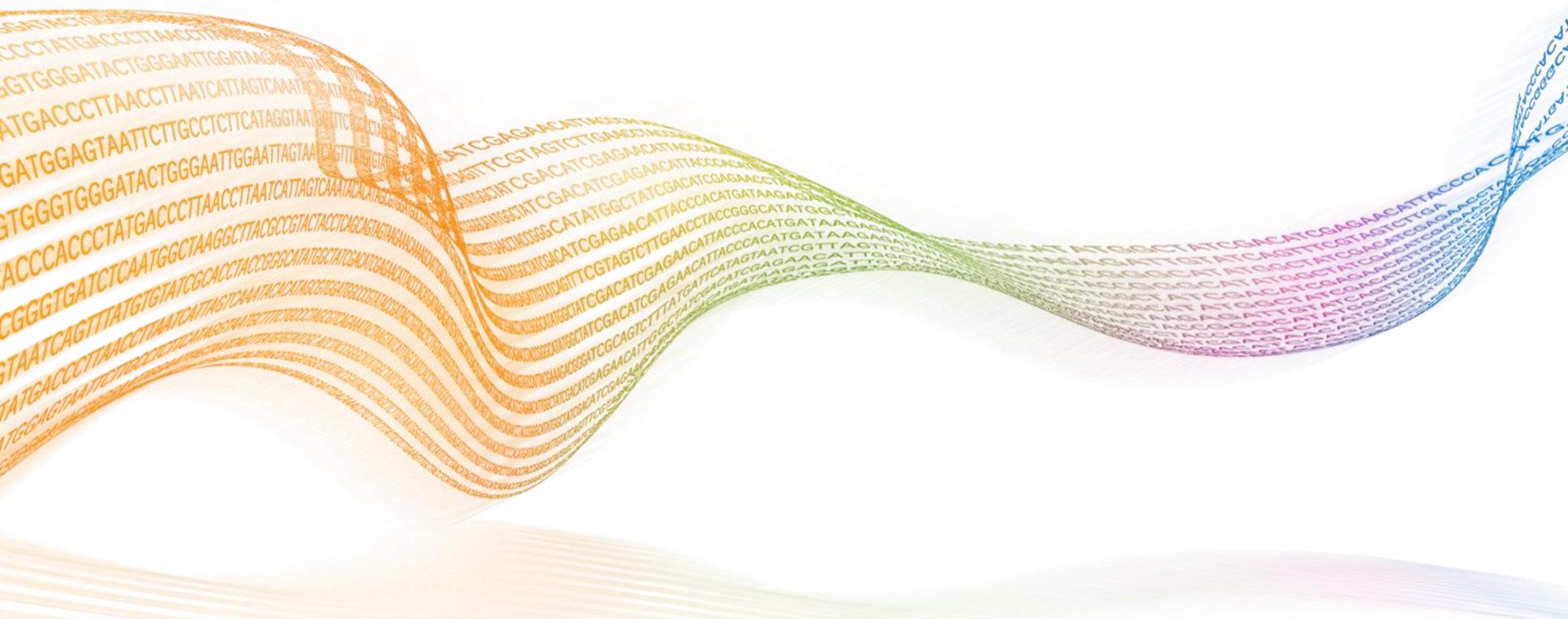
イルミナ iSchool – ウェビナー

- これからイルミナの次世代シーケンサーを利用される方向けに製品の使い方や、製品を使い始めた方向けの使用上のポイント、経験が豊富な方向けのトラブルシューティングのためのノウハウなど、経験に応じた初級~上級までの内容で、実験に役立つ情報をサービス・サポート部のスペシャリストがお伝えいたします。プロフェッショナルは、ユーザー様が実際にどのように次世代シーケンサーを研究に活用しているかを発表していただいています。

イルミナ iSchool – ラボトレーニング

- システム導入前にワークフローを体験したい、新しいラボスタッフにトレーニングを受けてほしいといったご要望にお応えするため、オンデマンドで東京オフィスのラボにてライブラリー調製を含むハンズオントレーニングを実施しています。

第1部 知っておきたいNGSの基礎



第1部 知っておきたいNGSの基礎

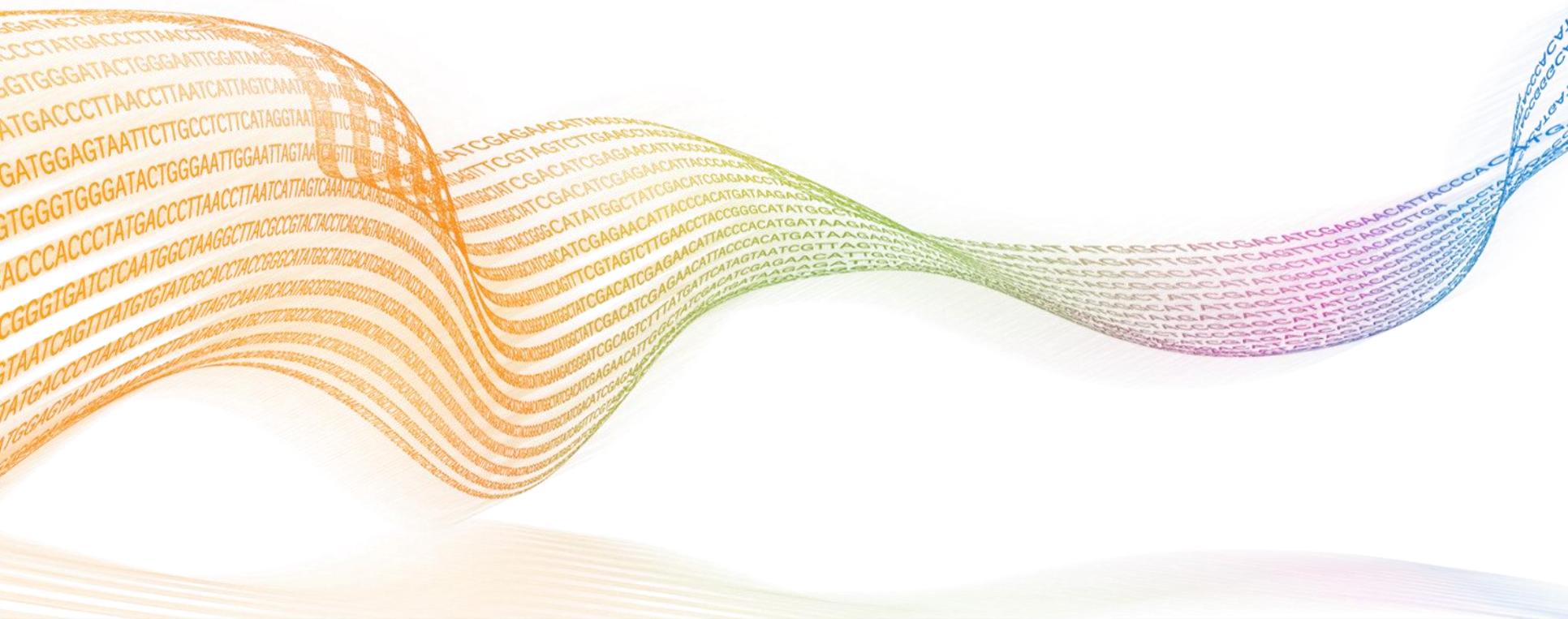
- ▶ 第1章 NGSの原理とワークフロー
- ▶ 第2章 NGSを扱う上で必要な分子生物学の基礎と用語

第1部 知っておきたいNGSの基礎

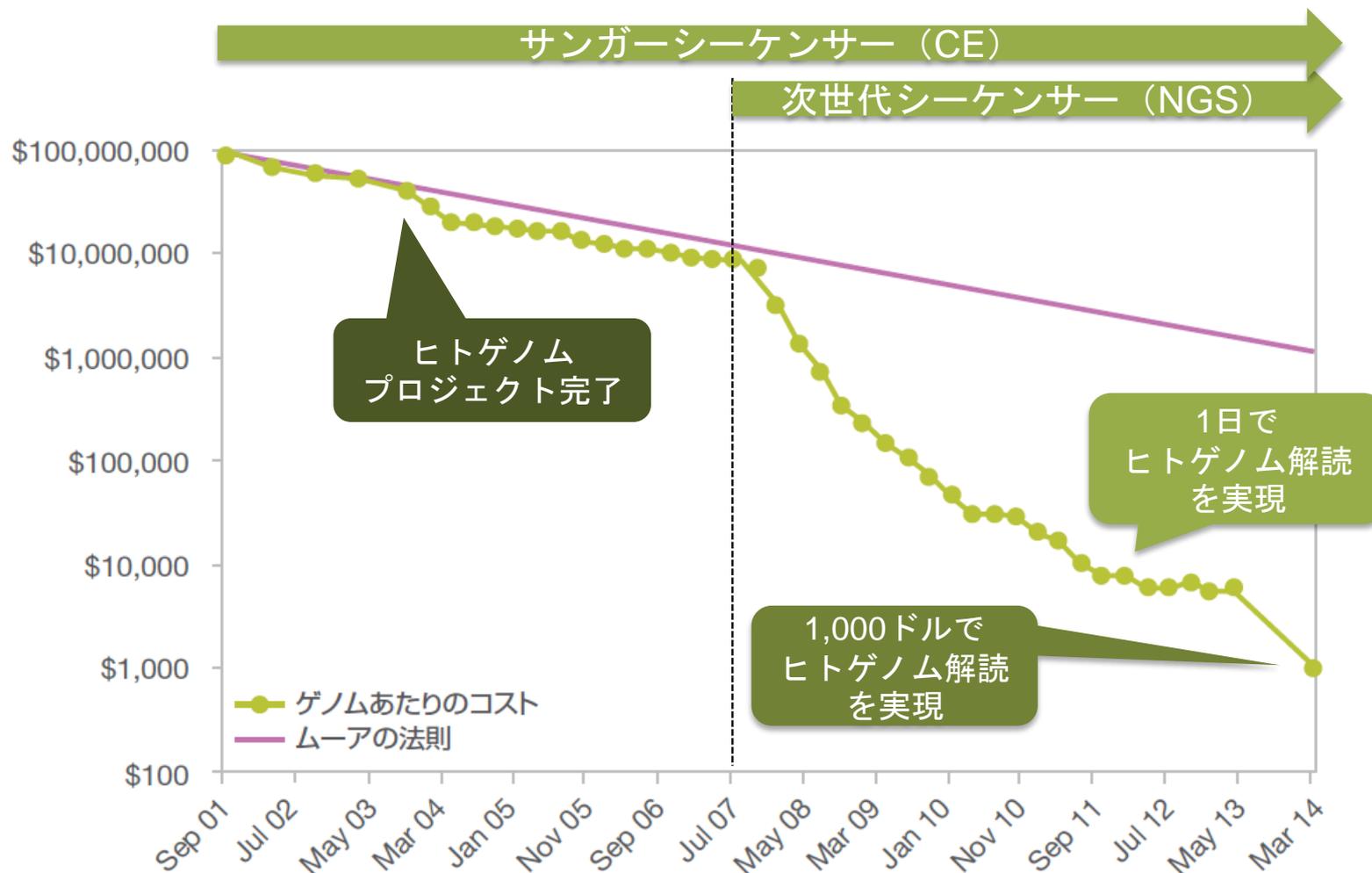
▶ 第1章 NGSの原理とワークフロー

1. 次世代シーケンサーとは？
2. 次世代シーケンサーの解読のしくみ

1. 次世代シーケンサーとは？



次世代シーケンサー（NGS）の誕生



サンガーシーケンサーの課題を克服した「次世代の」シーケンサー

ヒトゲノム解読コストの推移

2003

ヒトゲノム
プロジェクト



\$30億

2006

個人ゲノム解読



\$2,000万

2007

次世代シーケンサー
によるゲノム解読



\$200万

2008

30X カバレッジの
ゲノム解読



\$20万

ヒトゲノム解読コストの推移

2010

\$10,000

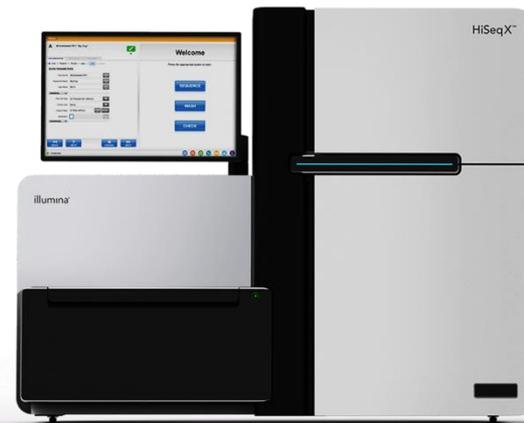
ゲノム



2014

\$1,000

ゲノム

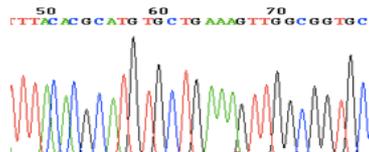
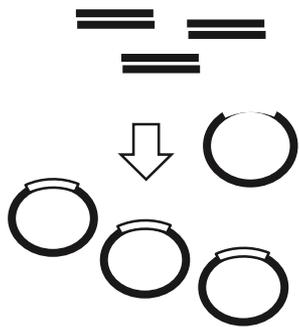


ヒトゲノム解読コスト
\$2,000万 → \$1,000
1/20,000

1,000ドルでヒトゲノム解読を実現

サンガーシーケンサーと次世代シーケンサーの違い

サンガーシーケンサー



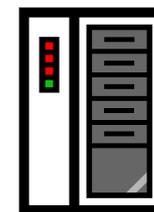
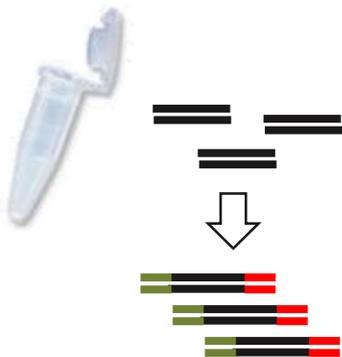
サンプル調製

DNA増幅

シーケンス

データ解析

次世代シーケンサー



ライブラリー調製

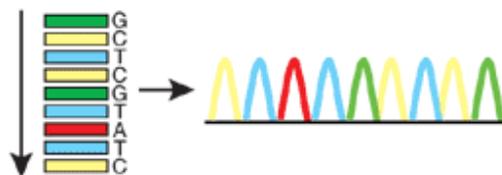
クラスター形成

シーケンス

データ解析

サンガーシーケンサーと次世代シーケンサーの違い

サンガーシーケンサー (1リード/キャピラリー)



リード長

X

リード数

||

データ量

1,000 bp x1

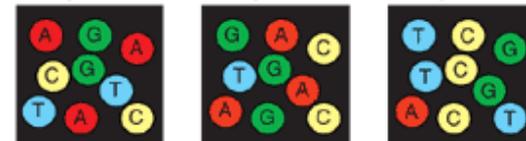
X

96

||

96 Kb

次世代シーケンサー (1リード/クラスター)



What is base 1? What is base 2? What is base 3?

150 bp x2

X

6,000,000,000

||

1.8 Tb

1回で多数のリードを解読

アプリケーションに応じた試薬使い分けが有効

試薬タイプ	試薬量	最大対応リード長	リード数	主な用途
V2 – 50 cycle	50bp	2x25	15M	<ul style="list-style-type: none"> ショートリードのシーケンス (Small RNAの発現プロファイリング) Library QC
V2 – 300 cycle	300bp	2x150	15M	<ul style="list-style-type: none"> Nextera XTキットを使った各種アプリケーション PCRアンプリコンの解析 TruSeq Custom Ampliconとの併用
V2 – 500 cycle	500bp	2x250	15M	<ul style="list-style-type: none"> ロングリードのアンプリコンシーケンス 小型ゲノムのDe novoアセンブリー、など

試薬タイプ	試薬量	最大対応リード長	リード数	主な用途
V3 - 150 cycle	150bp	2x75	25M	<ul style="list-style-type: none"> RNA-seq (トランスクリプトーム) 少数検体のExome、など
V3 – 600 cycle	600bp	2x300	25M	<ul style="list-style-type: none"> ロングリードのアンプリコンシーケンス 微生物のde novoアセンブリー HLA解析、など

1. 次世代シーケンサーとは？

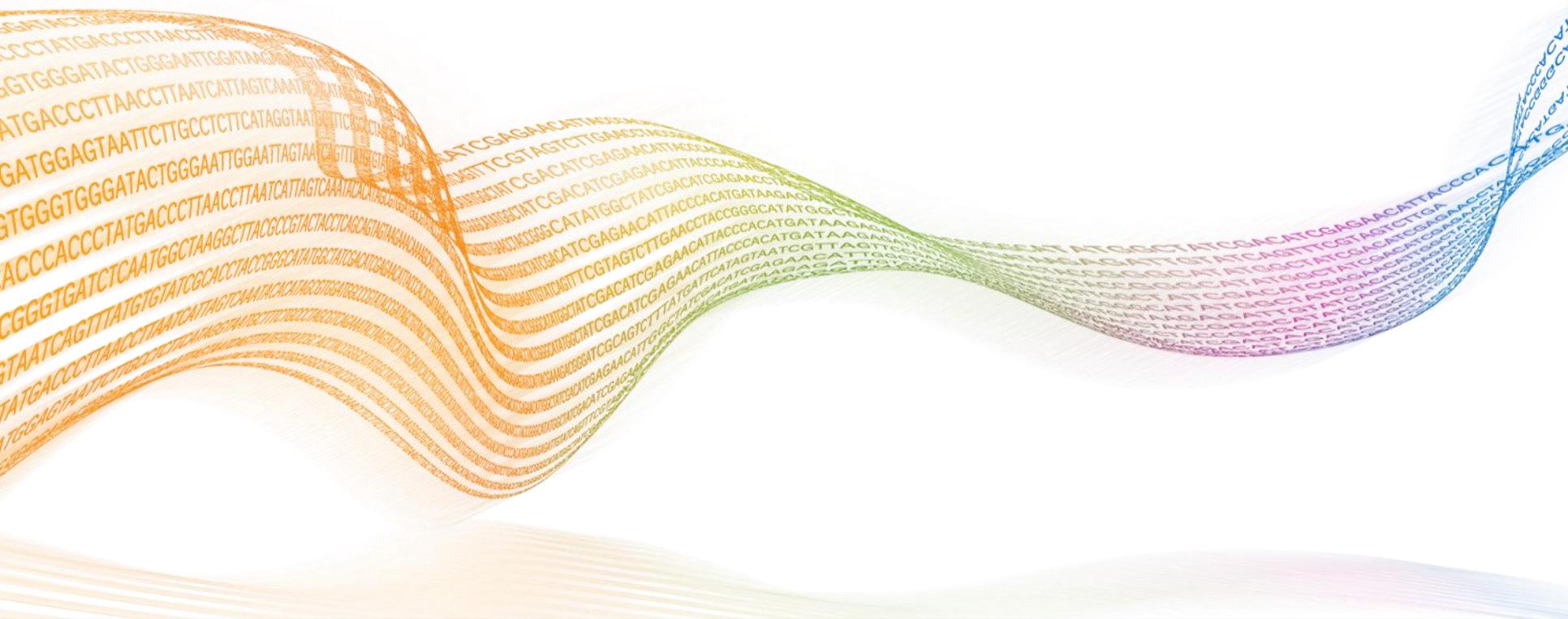
- サンガーシーケンサーの課題を克服した「次世代の」シーケンサー
- 1回で多数のリードを解読
- 1000ドルでヒトゲノム解読を実現

第1部 知っておきたいNGSの基礎

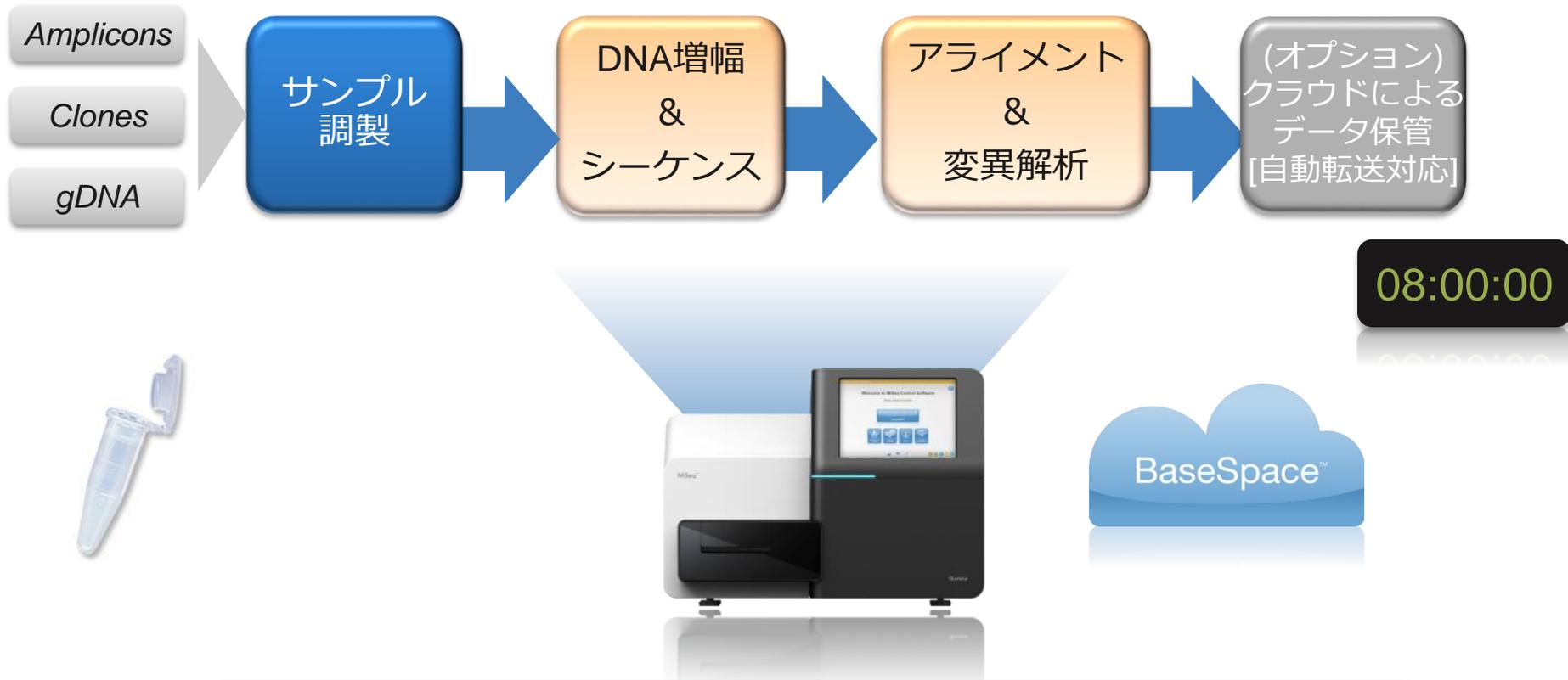
▶ 第1章 NGSの原理とワークフロー

1. 次世代シーケンサーとは？
2. 次世代シーケンサーの解読のしくみ

2. 次世代シーケンサーの解読のしくみ



自動化された快適な使い方を提案するMiSeqのワークフロー



* Nextera サンプル調製キットを使用時で、36bp \times 1のランの場合

MiSeq Sequencing Workflow

1

Library Preparation



2

Cluster Generation



3

Sequencing

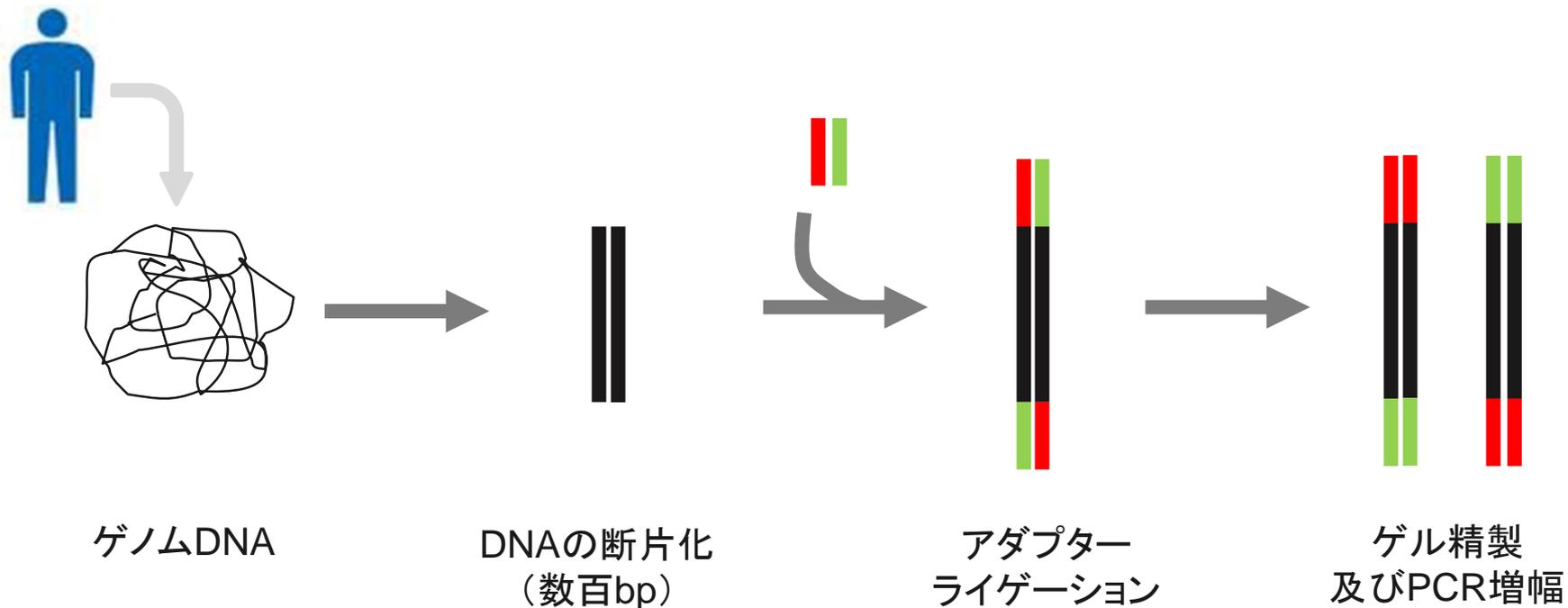


4

Data Analysis



ライブラリー調製の手順

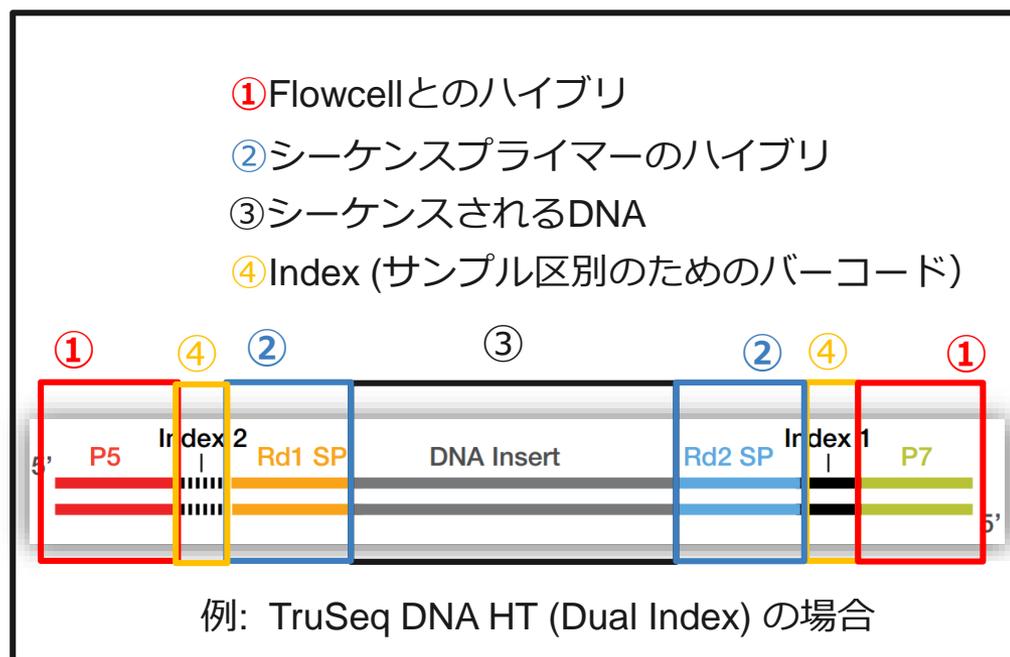


ライブラリー調製とは、次世代シーケンサーで解読できる長さにDNAを断片化し、シーケンスに必要な配列（アダプター）を付加したものを得ること

Library Preparation



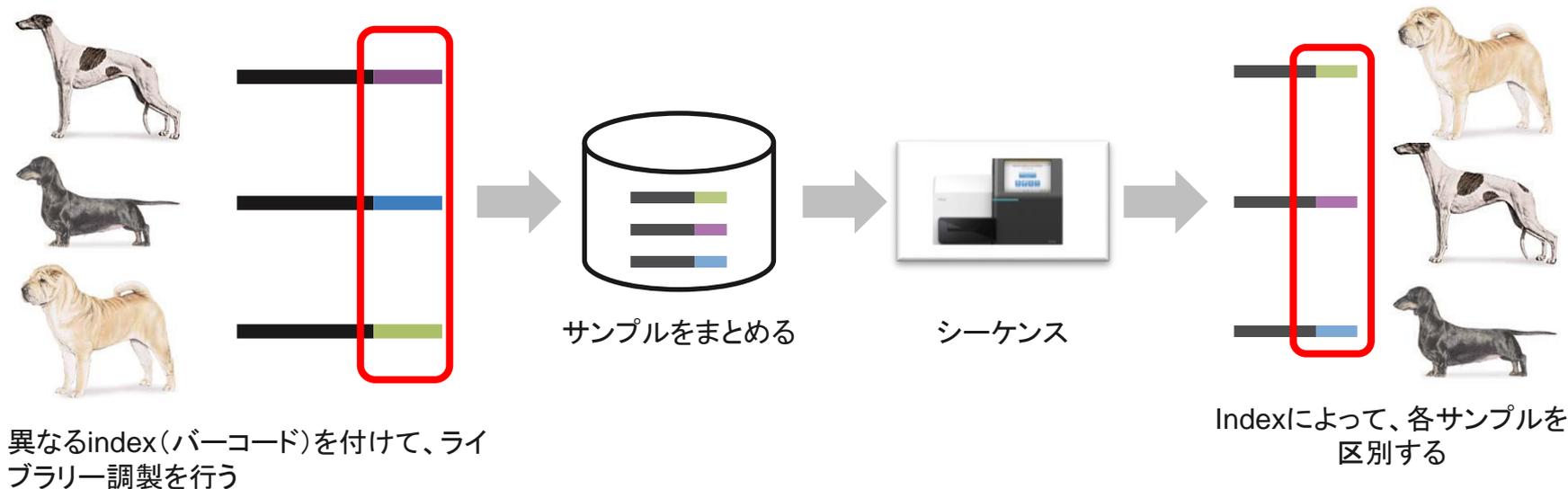
- ▶ ライブラリ調製の目的は、アダプタが両端に付いた核酸断片を得ること
- ▶ Illuminaから販売中の調製キット:
 - TruSeq sample preparation kit ([DNA]、[RNA]、[Small RNA]、[ChIP-Seq]、など)
 - TruSeq Custom Amplicon (TSCA) kit
 - Nextera kit
 - TruSight kit
- ▶ 推奨定量方法 (*キットごとに最適な方法は異なる)
 - qPCR
 - Qubit
 - Picogreen
- ▶ 推奨定性方法 (*キットごとに最適な方法は異なる)
 - Bioanalyzer
 - ゲル電気泳動



正確なライブラリ品質の確認と定量が、最適なクラスター密度と高いシーケンスパフォーマンスを得るためには必要不可欠である。

マルチプレックス法

- ▶ 異なる配列のインデックスを利用することで、1回のランで多サンプルを同時にシーケンスできる



MiSeq Sequencing Workflow

1

Library Preparation



2

Cluster Generation



3

Sequencing



4

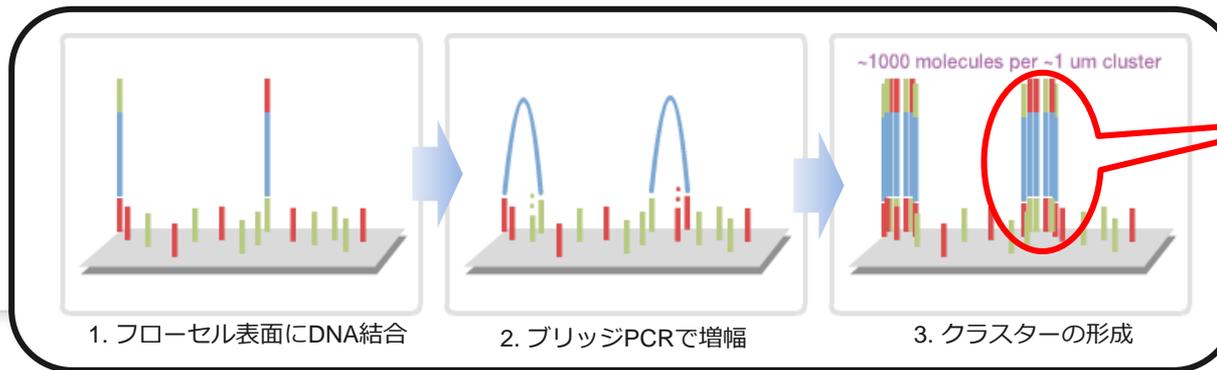
Data Analysis



Cluster Generation

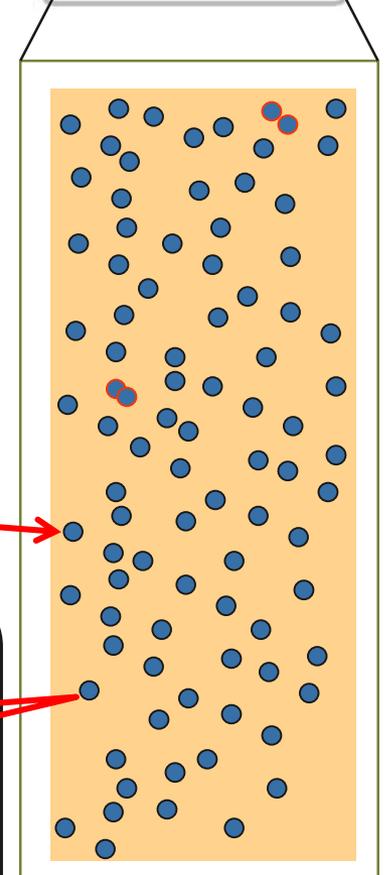
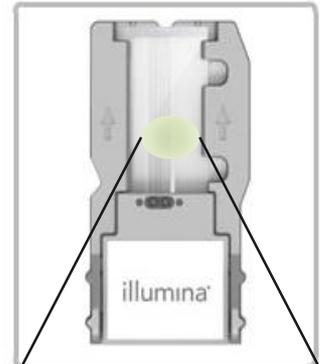


- ▶ クラスタ形成は、シーケンスの際に DNA を検出可能な充分量に増やす目的で行う
- ▶ クラスタは、シーケンスされるフローセルの表面上の DNA が増幅された領域である
- ▶ 各クラスタは増幅された異なる DNA を含む
- ▶ クラスタ形成では 5 つの主要なプロセスを経る
 1. Hybridize sample to flowcell
 2. Amplify sample
 3. Linearize fragments
 4. Block fragments
 5. Hybridize sequencing primer



クラスタ

Flow Cell



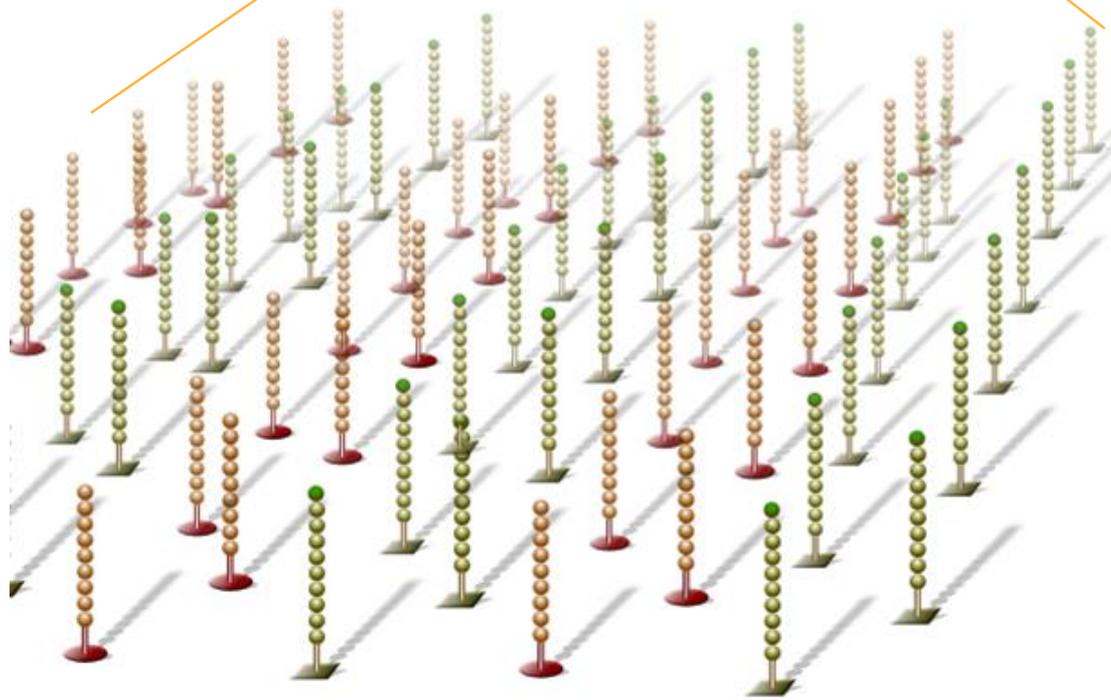
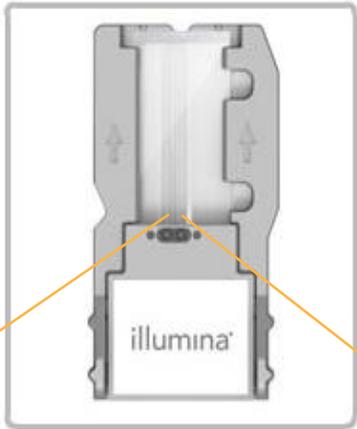
Flow Cellとは?

クラスター形成が行われる場所

試薬の流れる流路を持った、
ガラス基板

流路の内側には、ランダムに配置
されたオリゴの”芝”が並んでいる

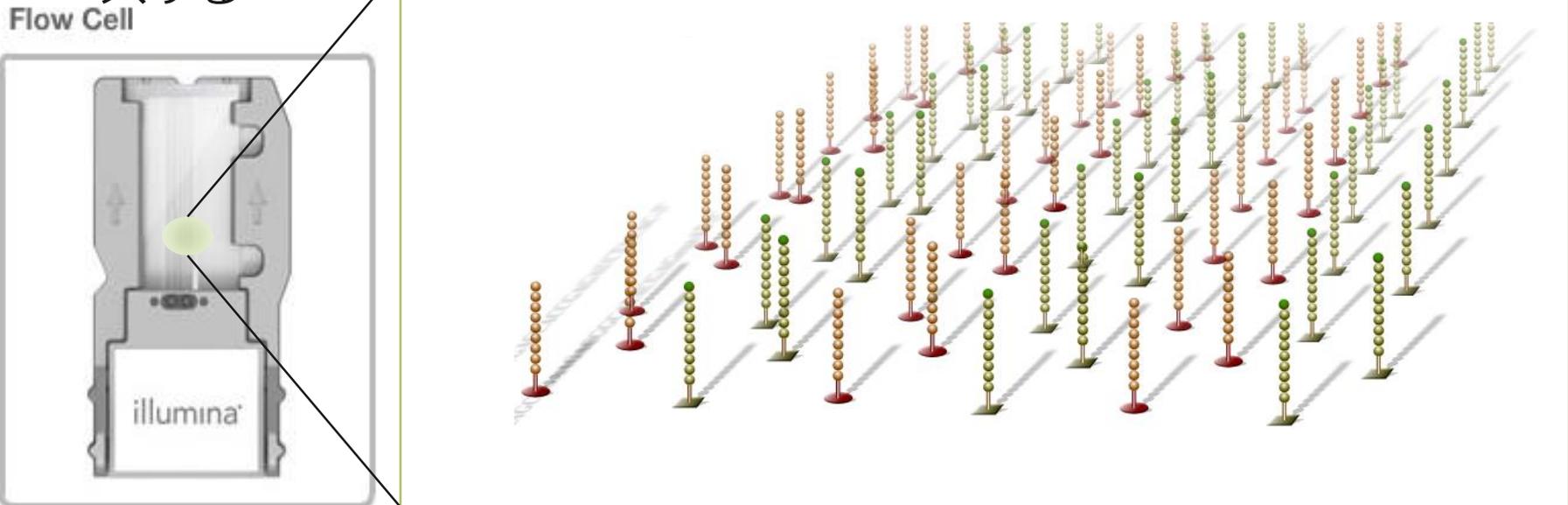
Flow Cell





Flowcell の模式図

- ▶ フローセルにはライブラリ断片のアダプタに相補的なプライマーの‘芝’がある
- ▶ ライブラリ断片はまず始めにフローセル表面のプライマーにハイブリダイズする



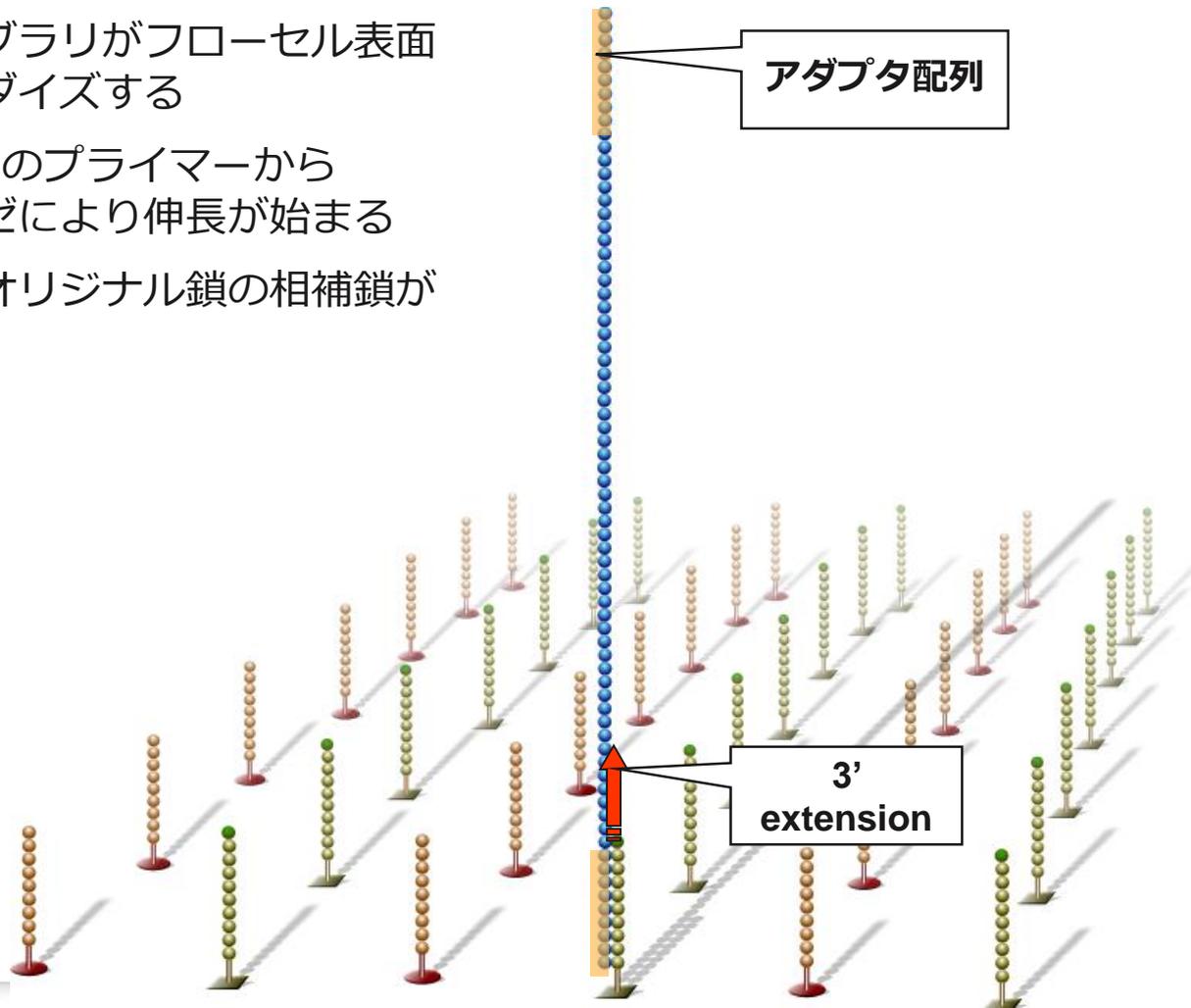
Cluster Generation

1. Hybridize sample to flowcell

Hybridize Fragment & Extend



- ▶ まず、ライブラリがフローセル表面にハイブリダイズする
- ▶ 次に、**‘芝’** のプライマーからポリメラーゼにより伸長が始まる
- ▶ これによりオリジナル鎖の相補鎖が合成される



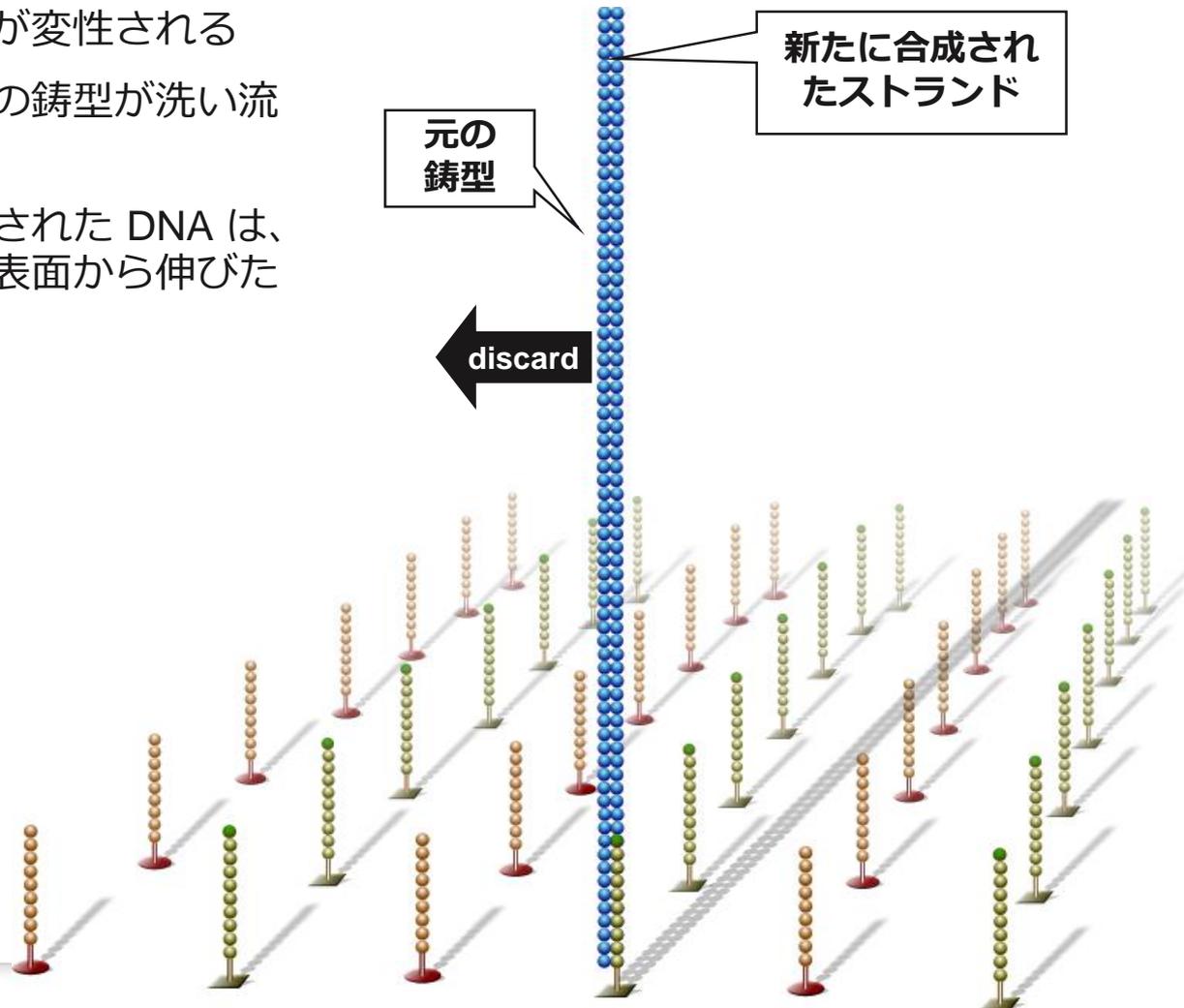
Cluster Generation

1. Hybridize sample to flowcell

Denature Double-stranded DNA



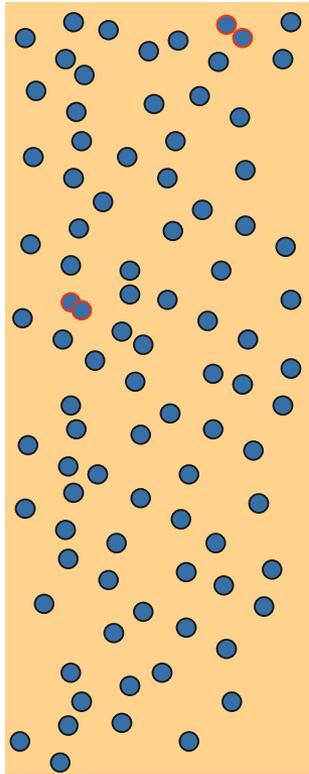
- ▶ 二本鎖分子が変性される
- ▶ オリジナルの鋳型が洗い流される
- ▶ 新たに合成された DNA は、フローセル表面から伸びた状態となる



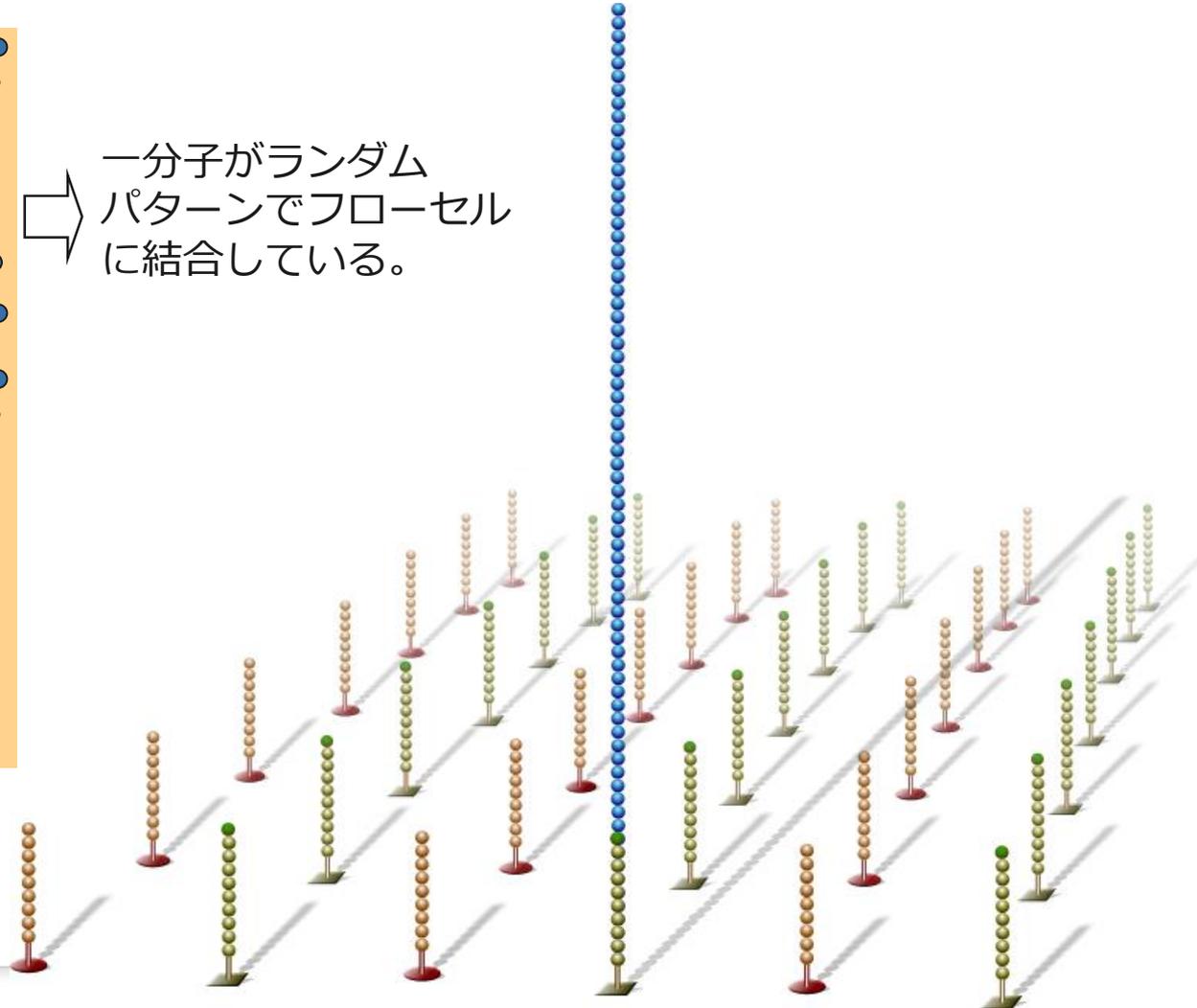
Cluster Generation

1. Hybridize sample to flowcell

Denature Double-stranded DNA



一分子がランダム
パターンでフローセル
に結合している。



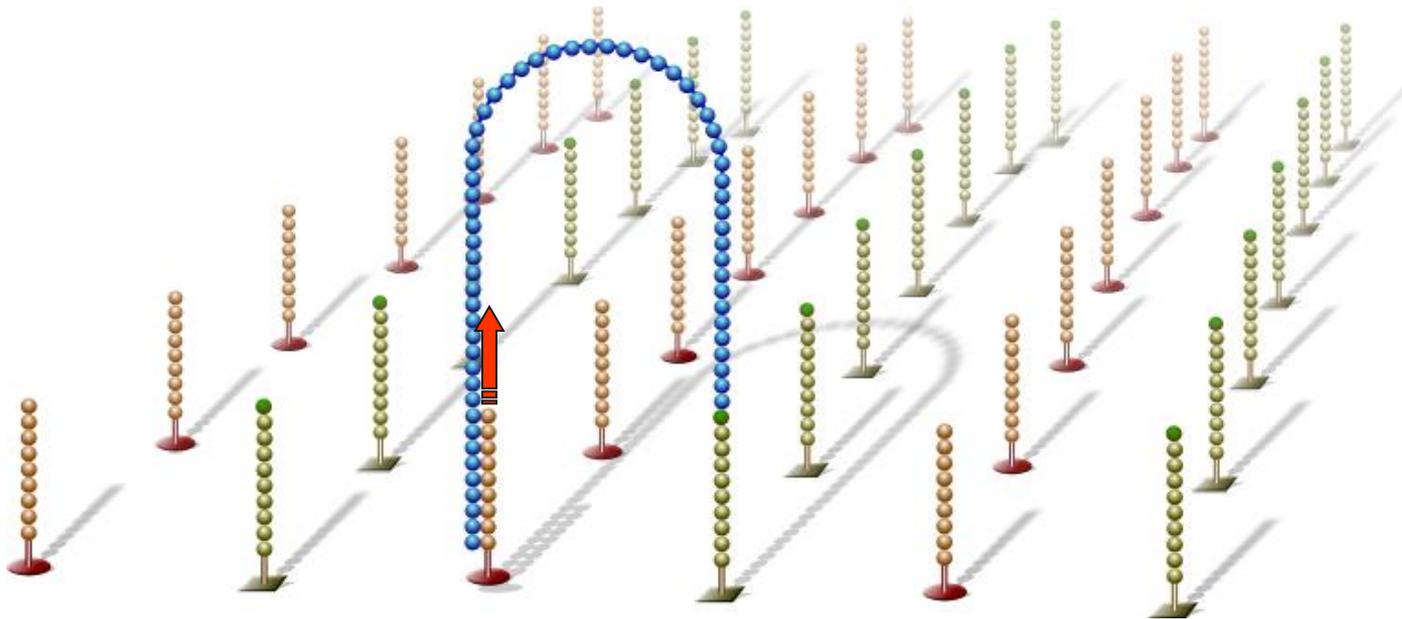
Cluster Generation

2. Amplify sample

Bridge Amplification



- ▶ 一本鎖が曲がって 近傍のプライマーとハイブリダイズしてブリッジを形成する
- ▶ ハイブリダイズしたプライマーはポリメラーゼによって伸長する



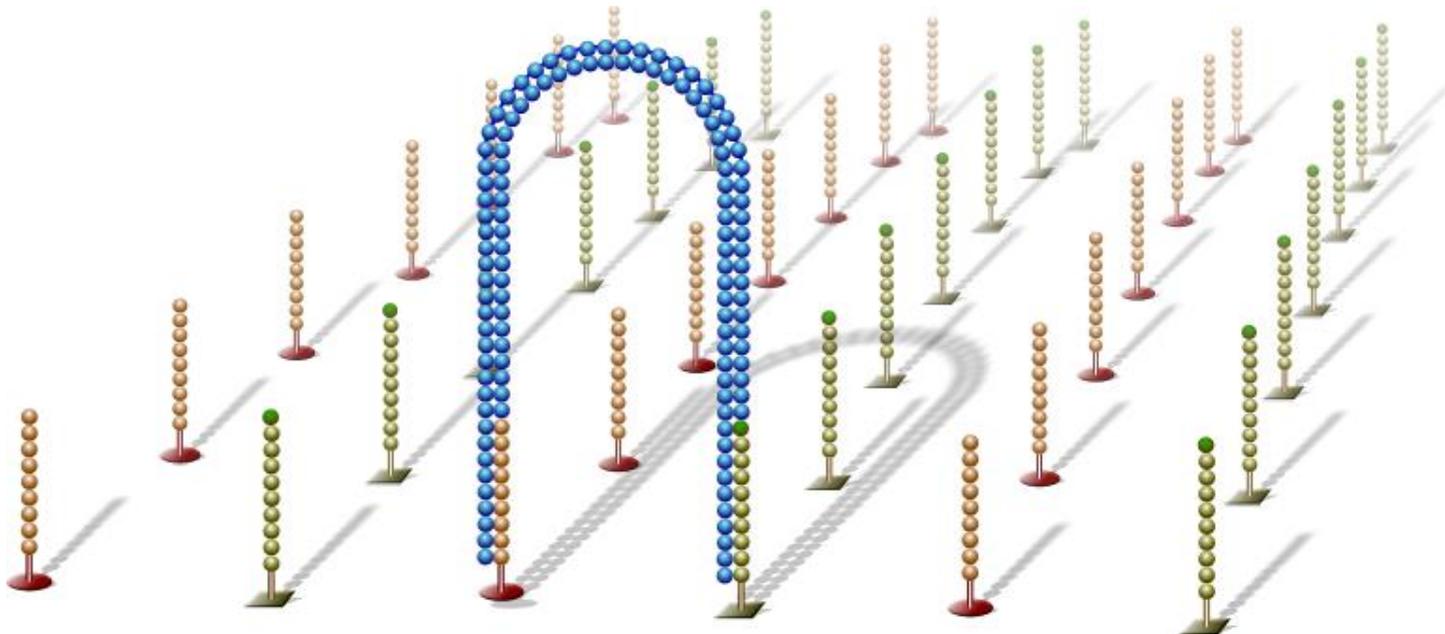
Cluster Generation

2. Amplify sample

Bridge Amplification



- ▶ 二本鎖のブリッジが形成される



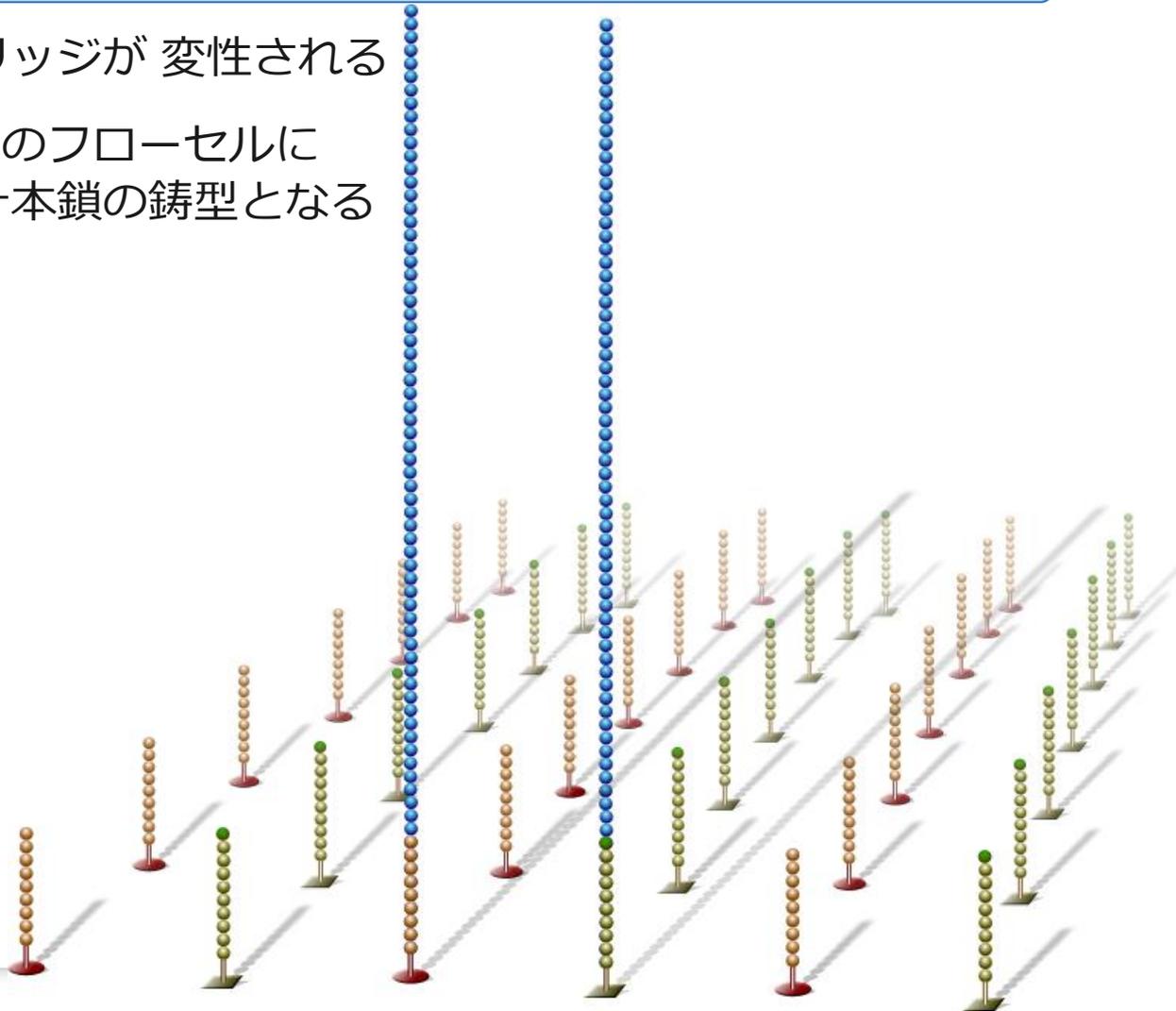
Cluster Generation

2. Amplify sample

Bridge Amplification



- ▶ 二本鎖ブリッジが変性される
- ▶ 結果: 一对のフローセルに結合した一本鎖の鋳型となる



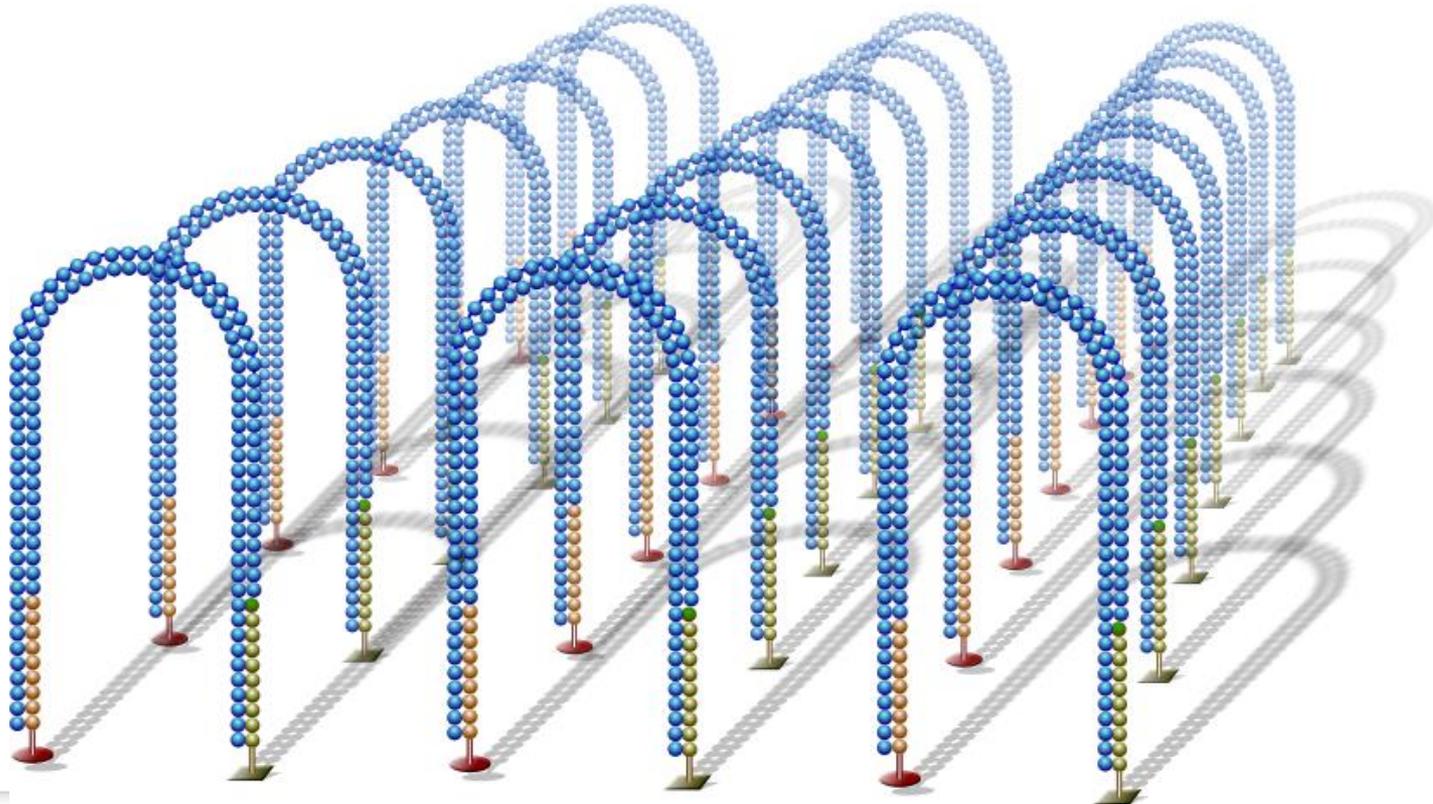
Cluster Generation

2. Amplify sample

Bridge Amplification



- ▶ 多数のブリッジが形成されるまでサイクルが繰り返される。



Cluster Generation

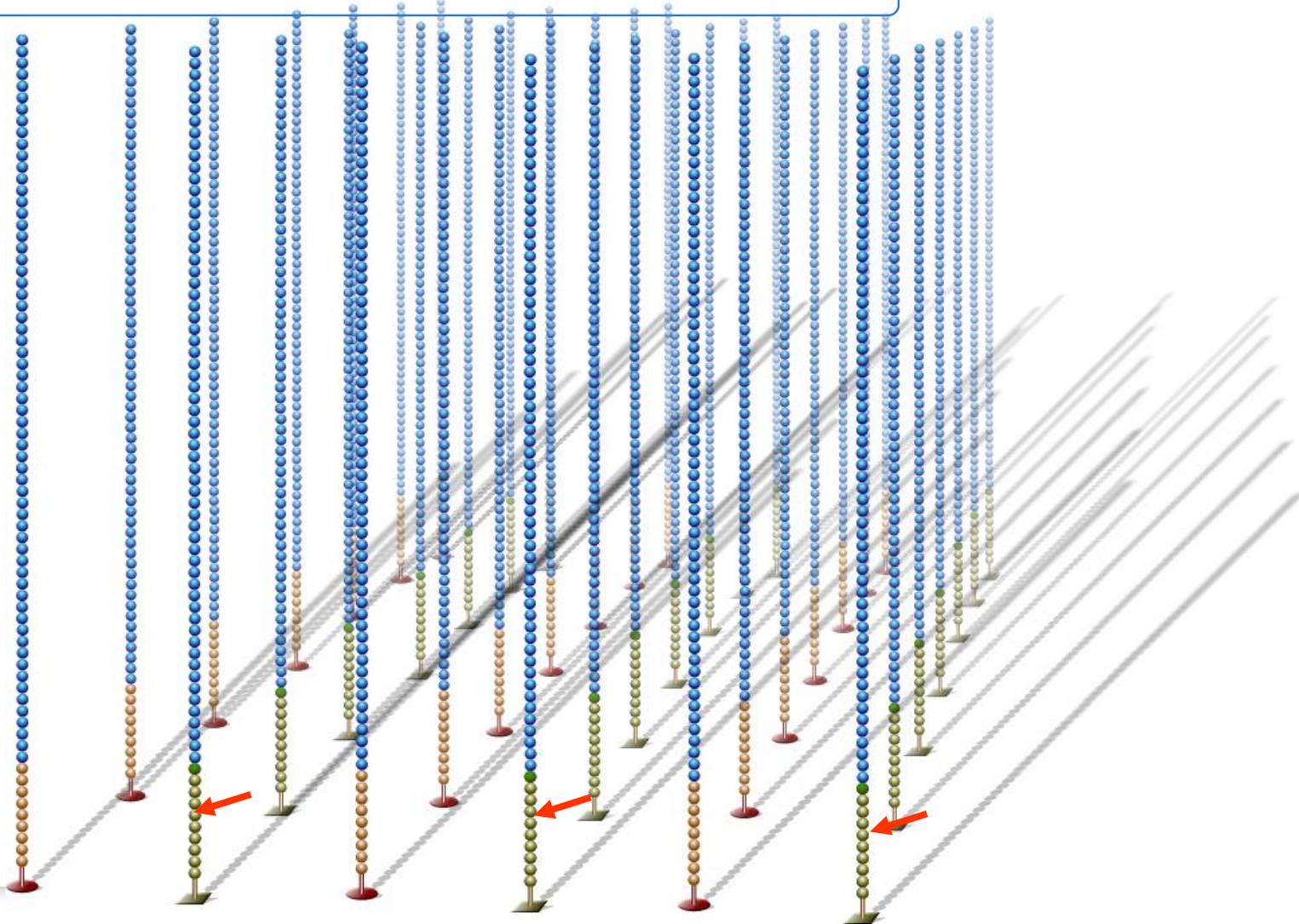
3. Linearize fragments

Linearization



▶ dsDNA のブリッジが変性される

▶ 逆鎖が開裂され、洗い流される



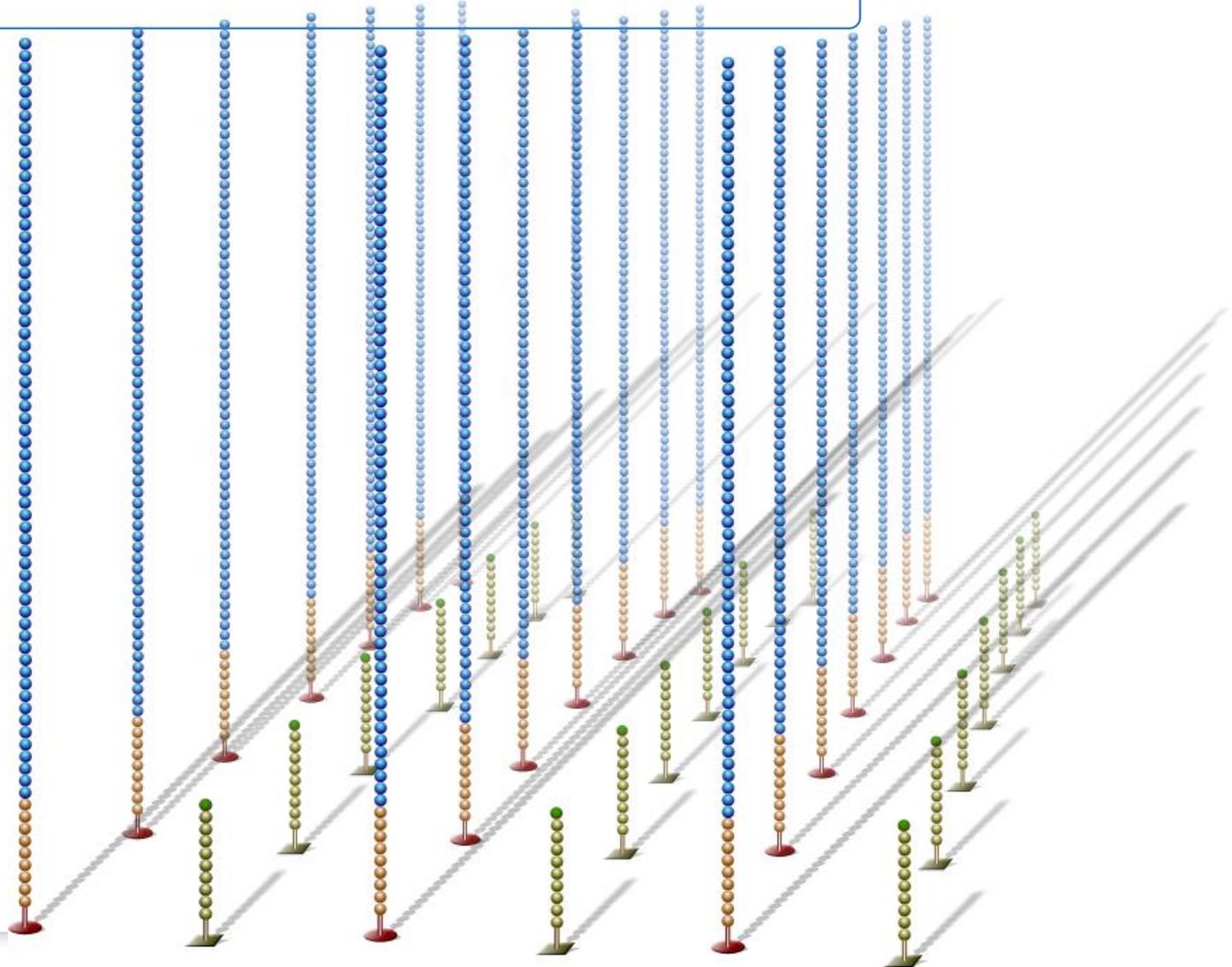
Cluster Generation

3. Linearize fragments

Linearization



- ▶ …順鎖のみが
クスタとして
残される



Cluster Generation

4. Hybridize sequencing primer

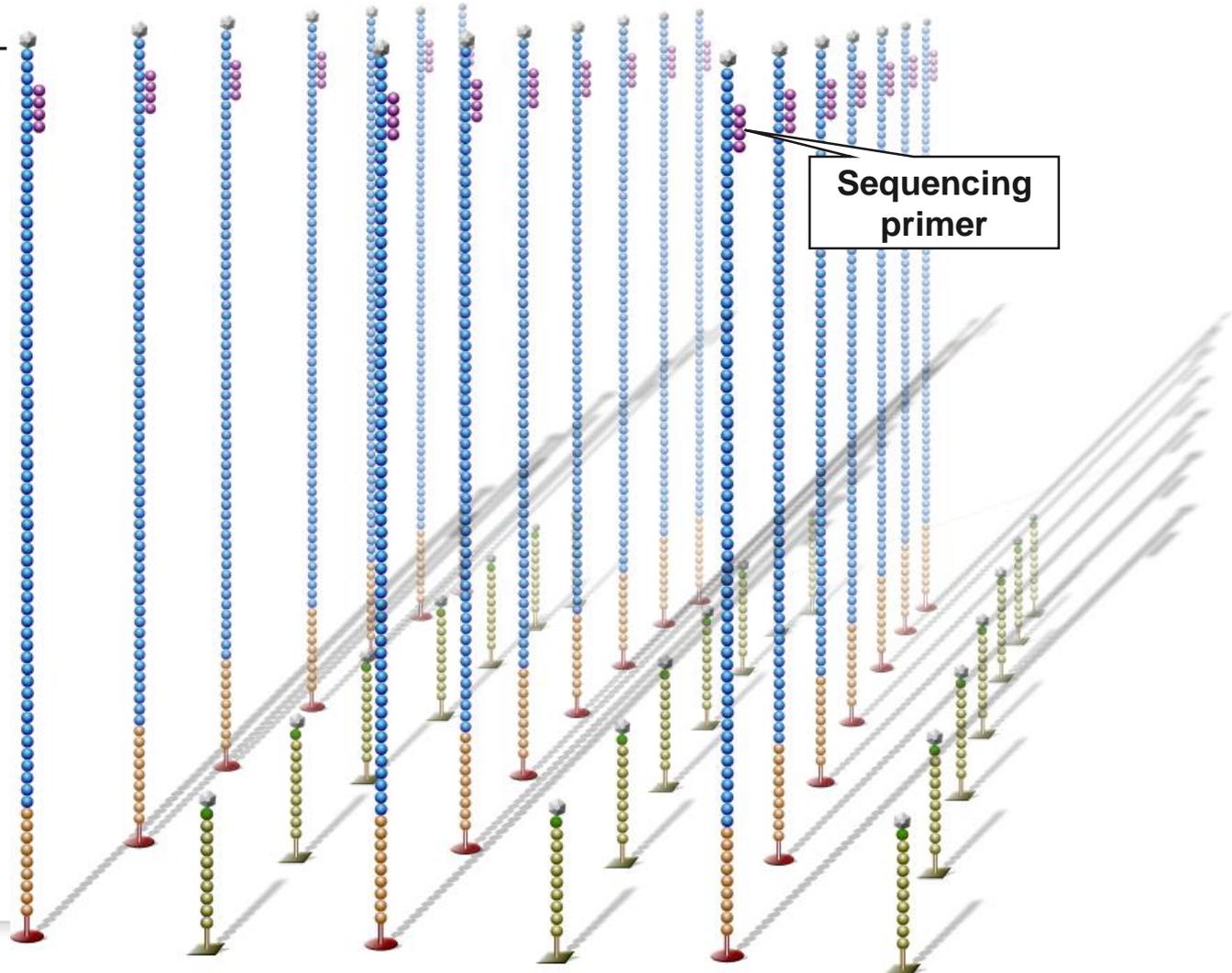
Primer Hybridization



シーケンスプライマー
がアダプタ配列に
ハイブリダイズする



シーケンス反応へ



MiSeq Sequencing Workflow

1

Library Preparation



2

Cluster Generation



3

Sequencing



4

Data Analysis

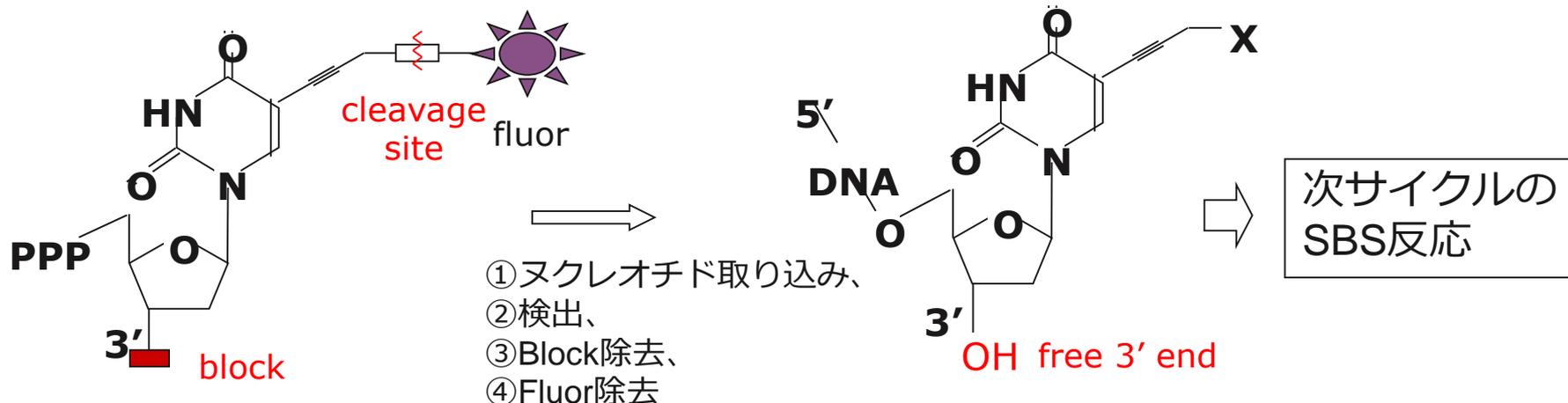




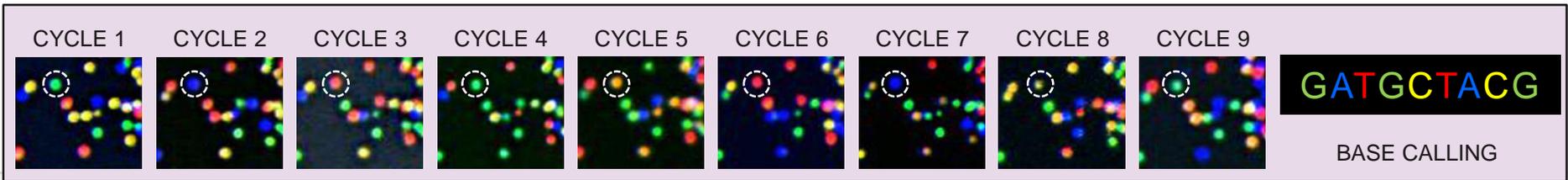
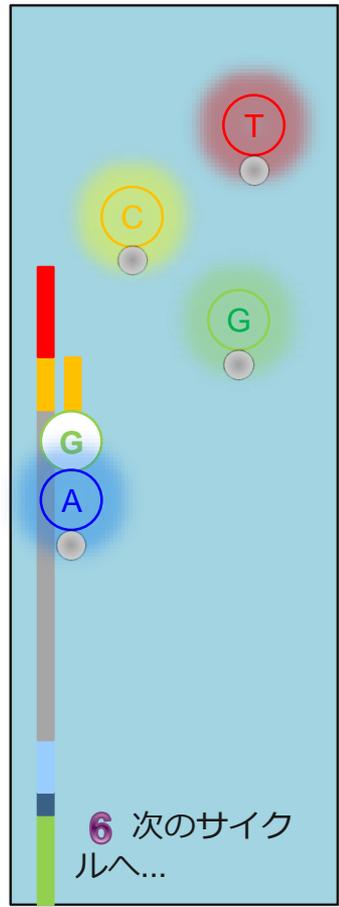
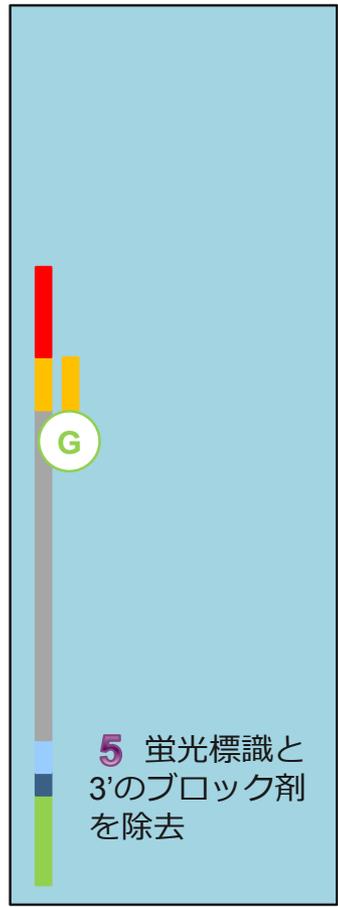
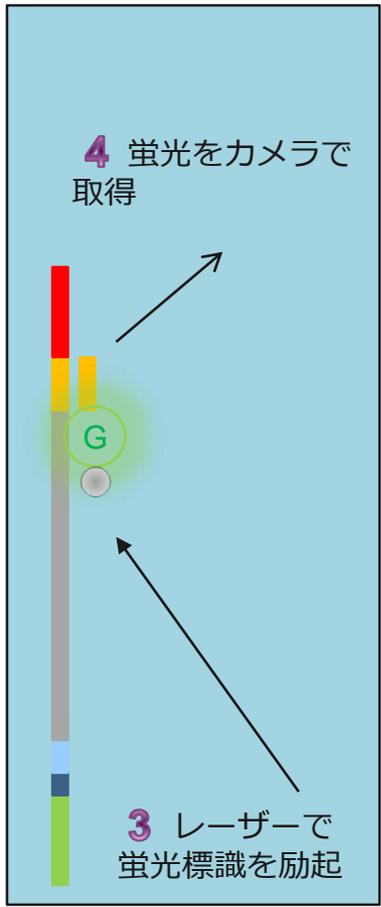
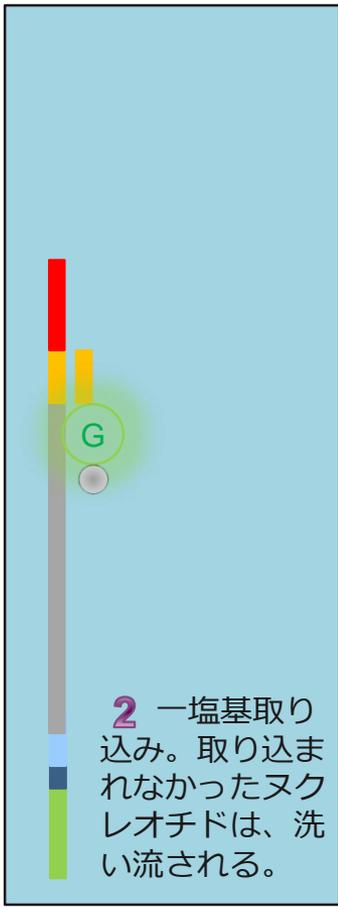
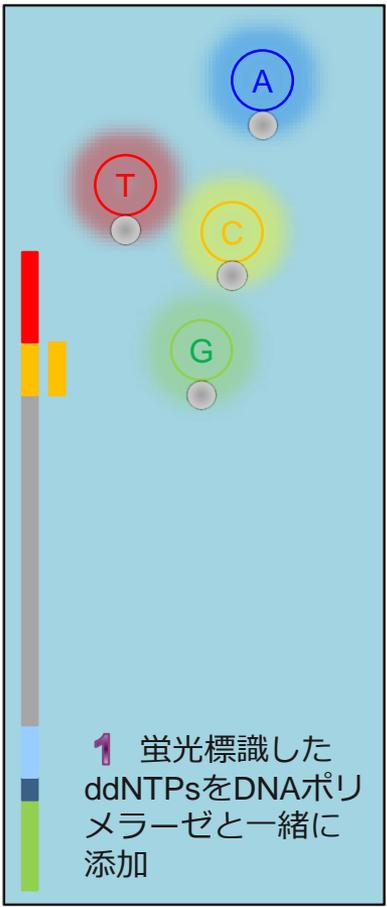
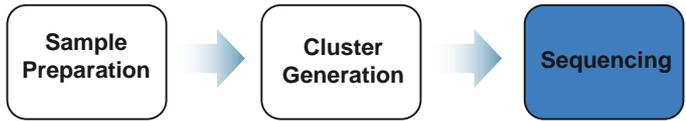
- ▶ MiSeq は可逆的ターミネーター法として信頼のあるイルミナ Sequencing By Synthesis (SBS) 反応を用いて DNA クラスタをシーケンスする:

Reversible Terminator Chemistry

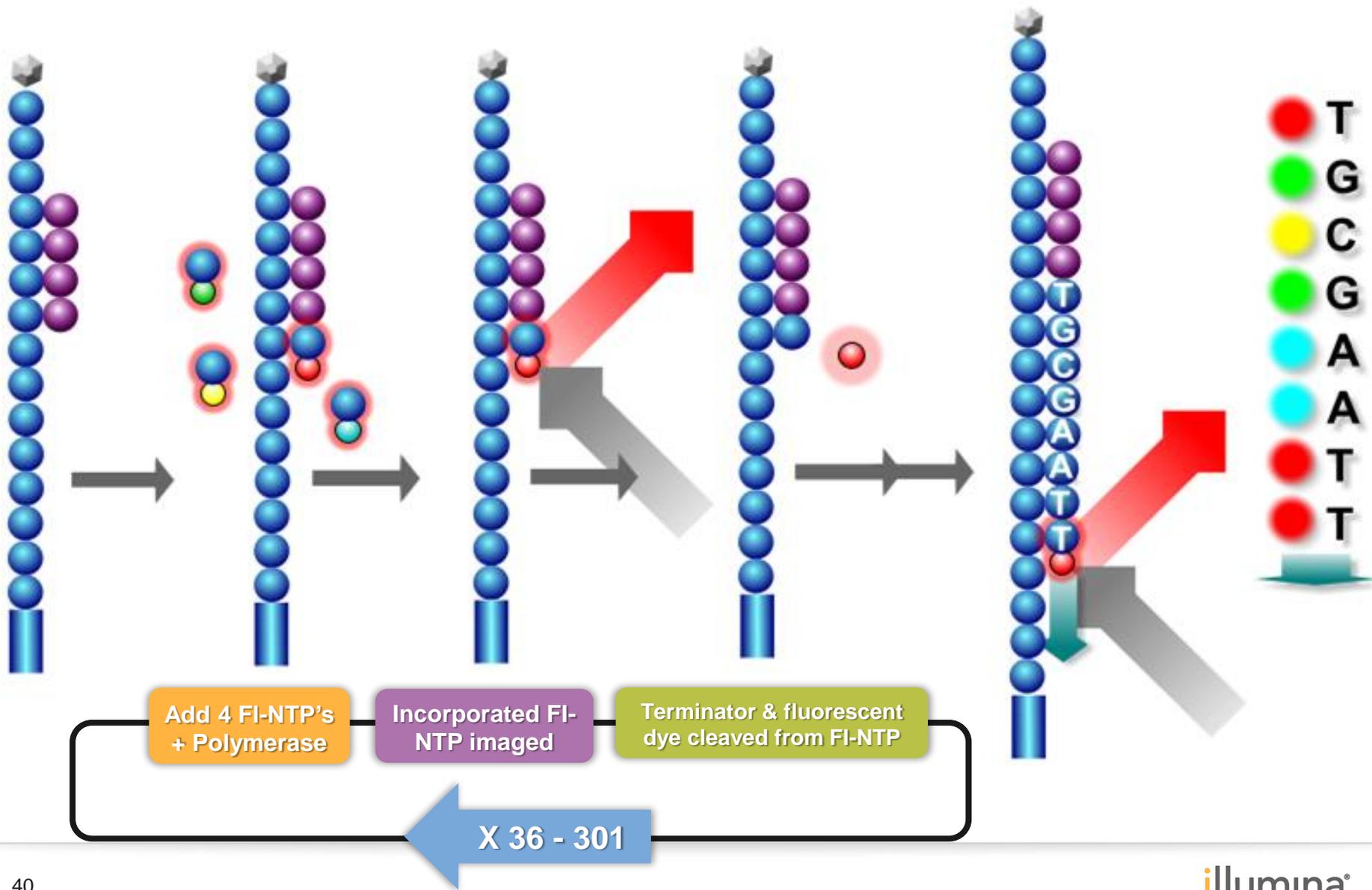
- ▶ 1 反応に蛍光ラベルされた 4 ヌクレオチドを加える
(蛍光の種類で塩基を特定)
- ▶ 高精度
- ▶ ホモポリマー領域も問題ない



SBS シーケンスケミストリー



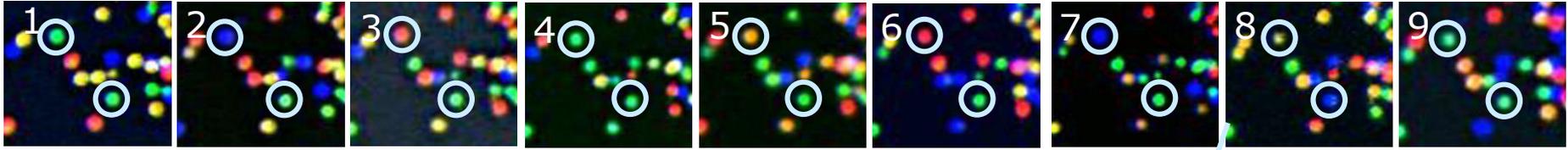
Sequencing by Synthesis (SBS) 反応



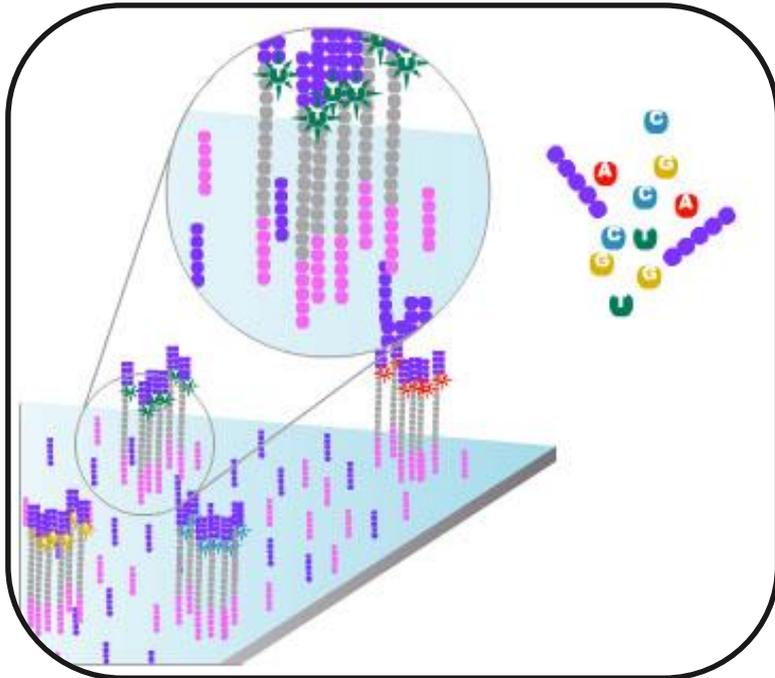
3 Sequencing



T G C T A C G A T ...



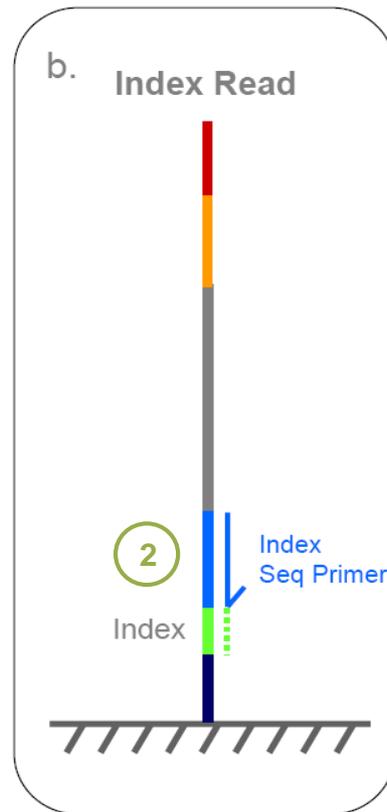
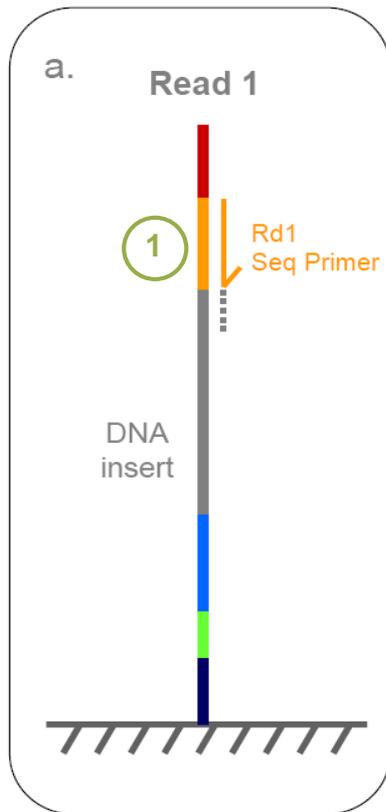
T T T T T T T G T ...



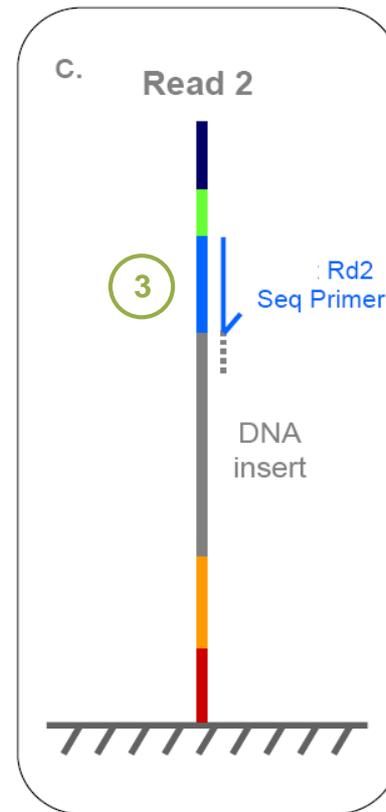
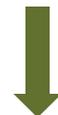
- 1サイクルごとにイメージの定点観測が行われる。
- 各クラスターの塩基配列はイメージから同定される。

ペアエンドシーケンス & シングルインデックスでシーケンスする場合

クラスター形成



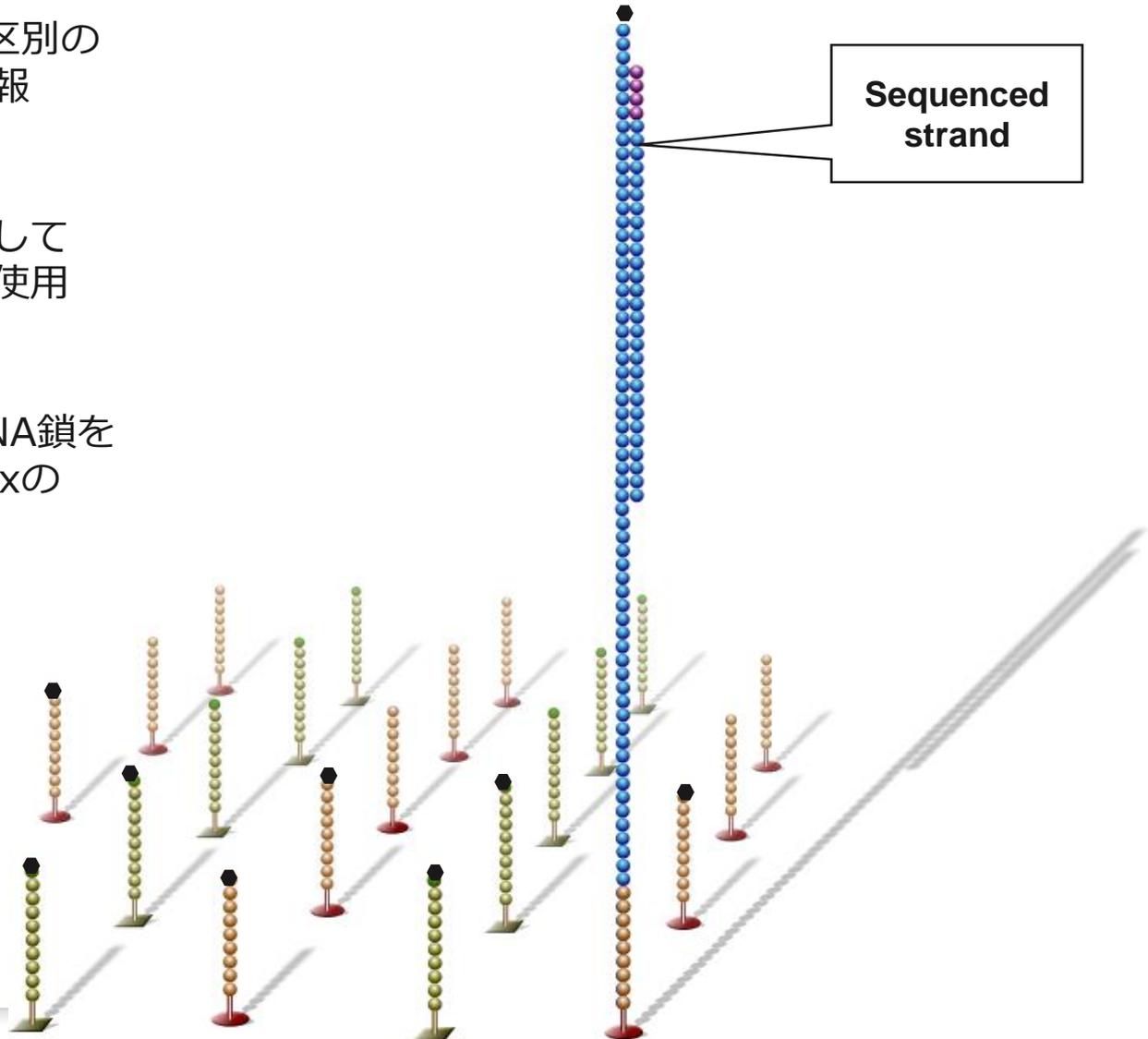
Read 2
Resynthesis



変性 & インデックスプライマーのハイブリ

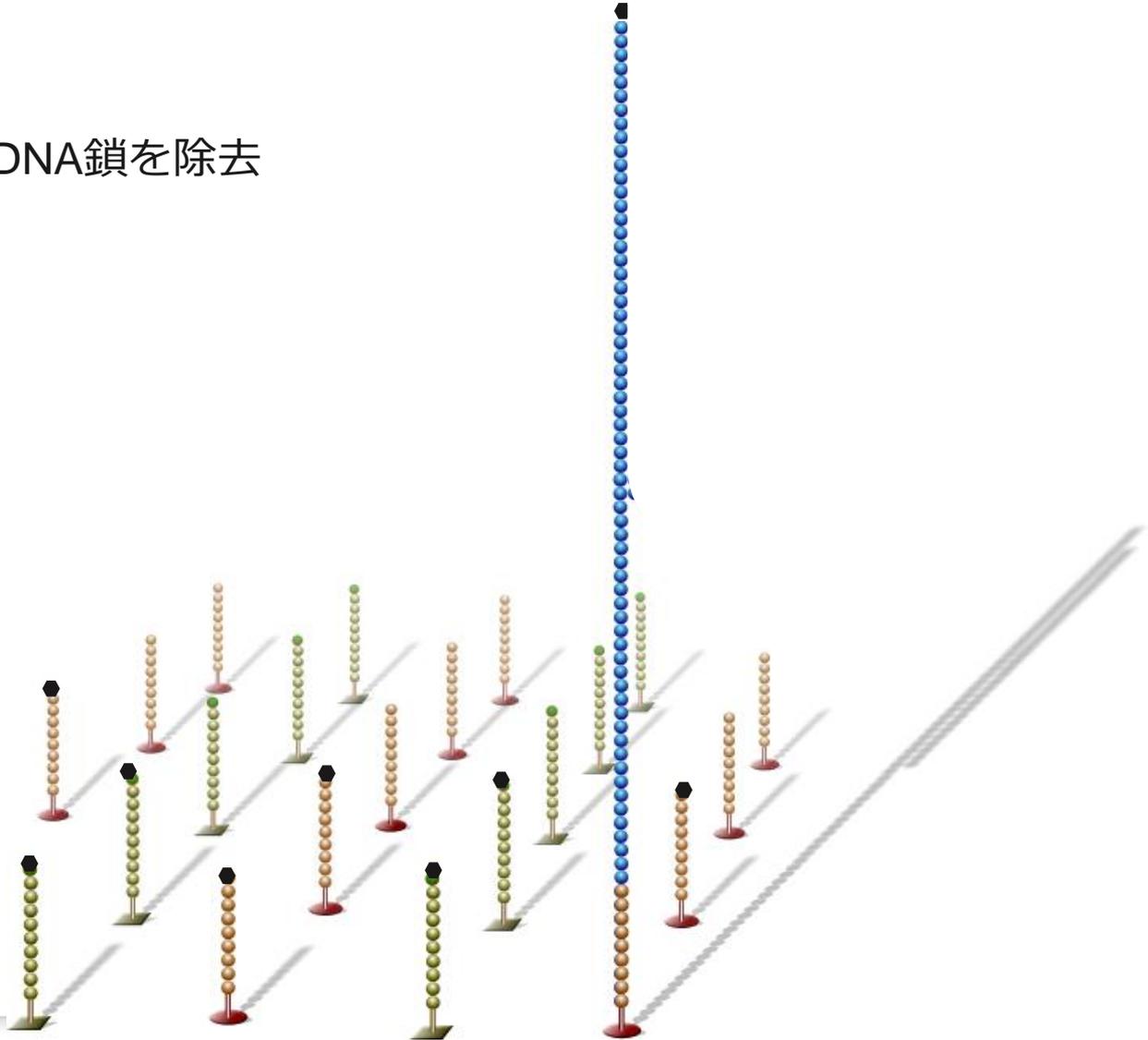


- ▶ Index : サンプル区別のためのバーコード情報
- ▶ 複数サンプルを混合して同時解析する場合に使用
- ▶ シーケンス済みのDNA鎖を剥がしてから、Indexの読み取りが行われる



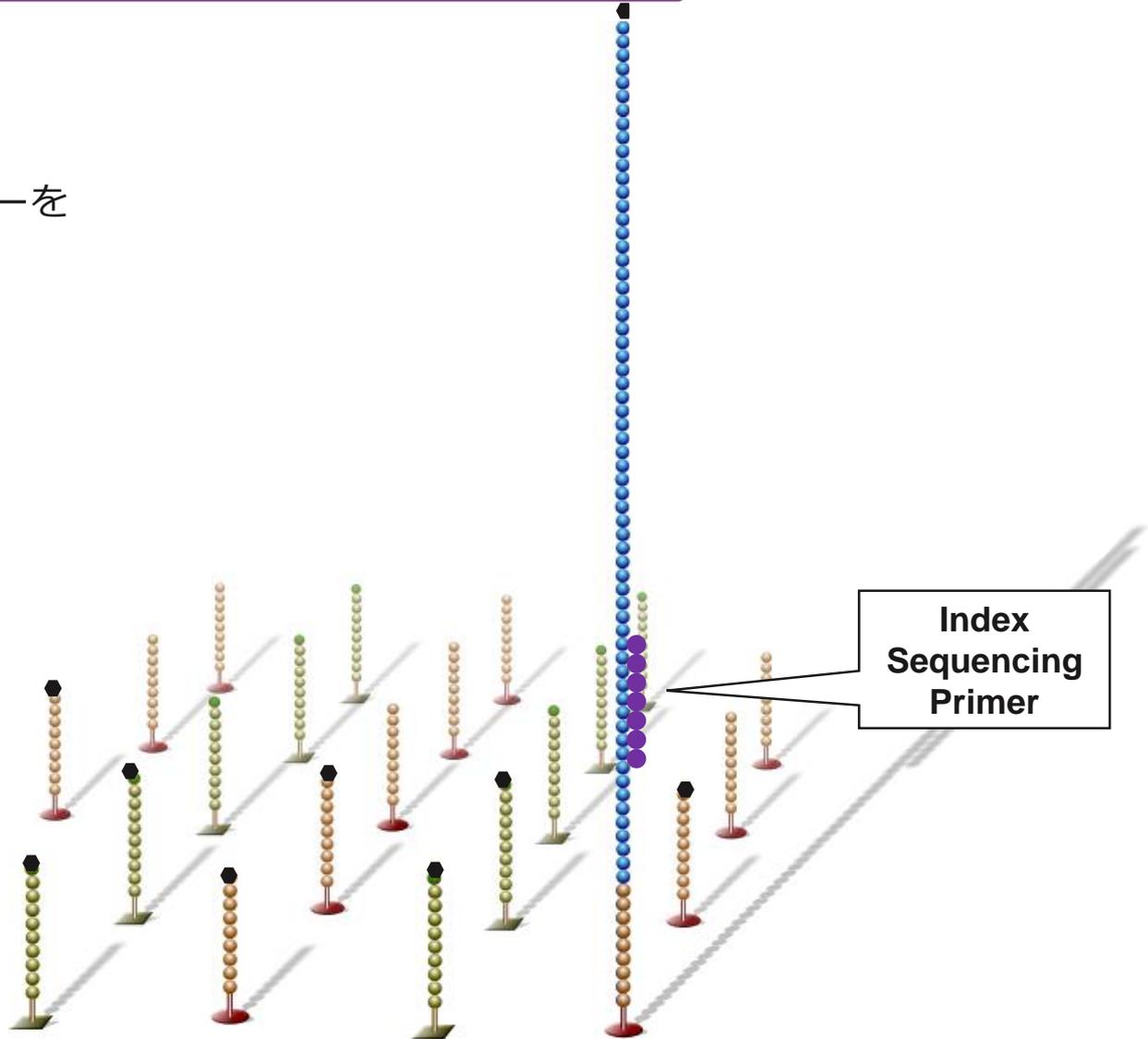


- ▶ シーケンス済みのDNA鎖を除去

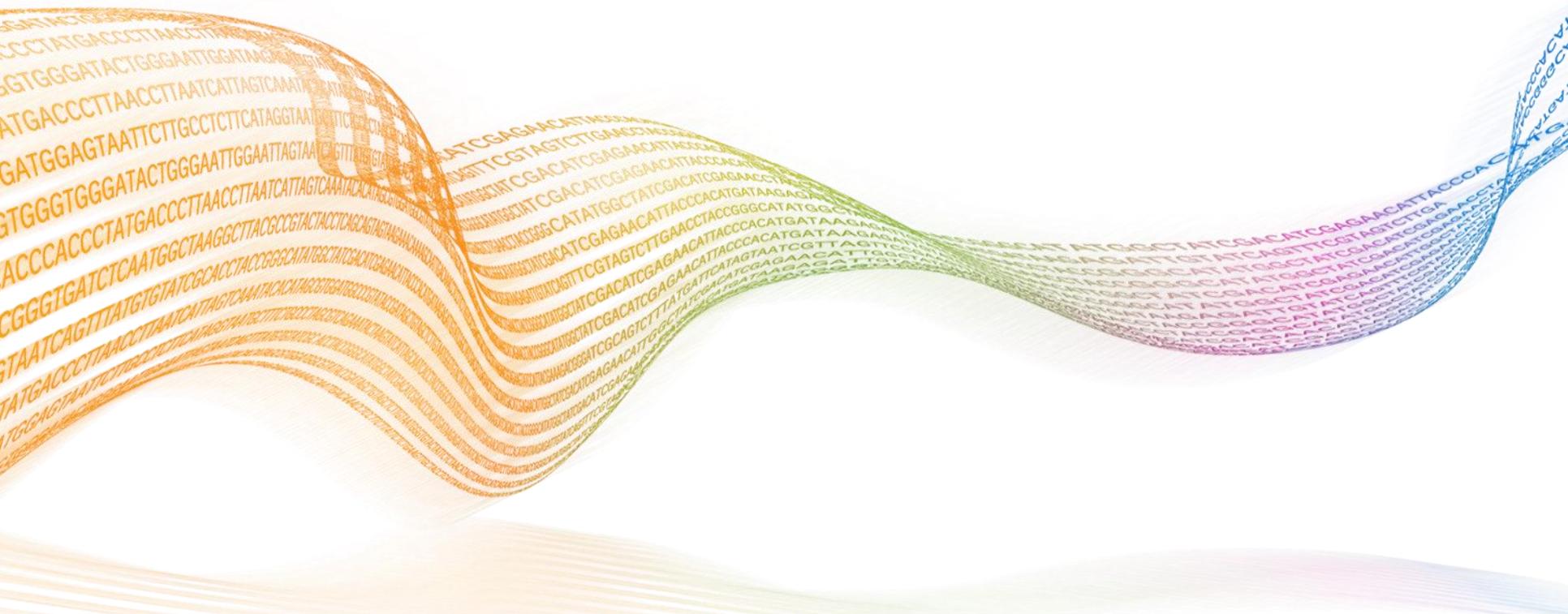




- ▶ Index読み取り用のシーケンスプライマーをハイブリ



Paired End Sequencing



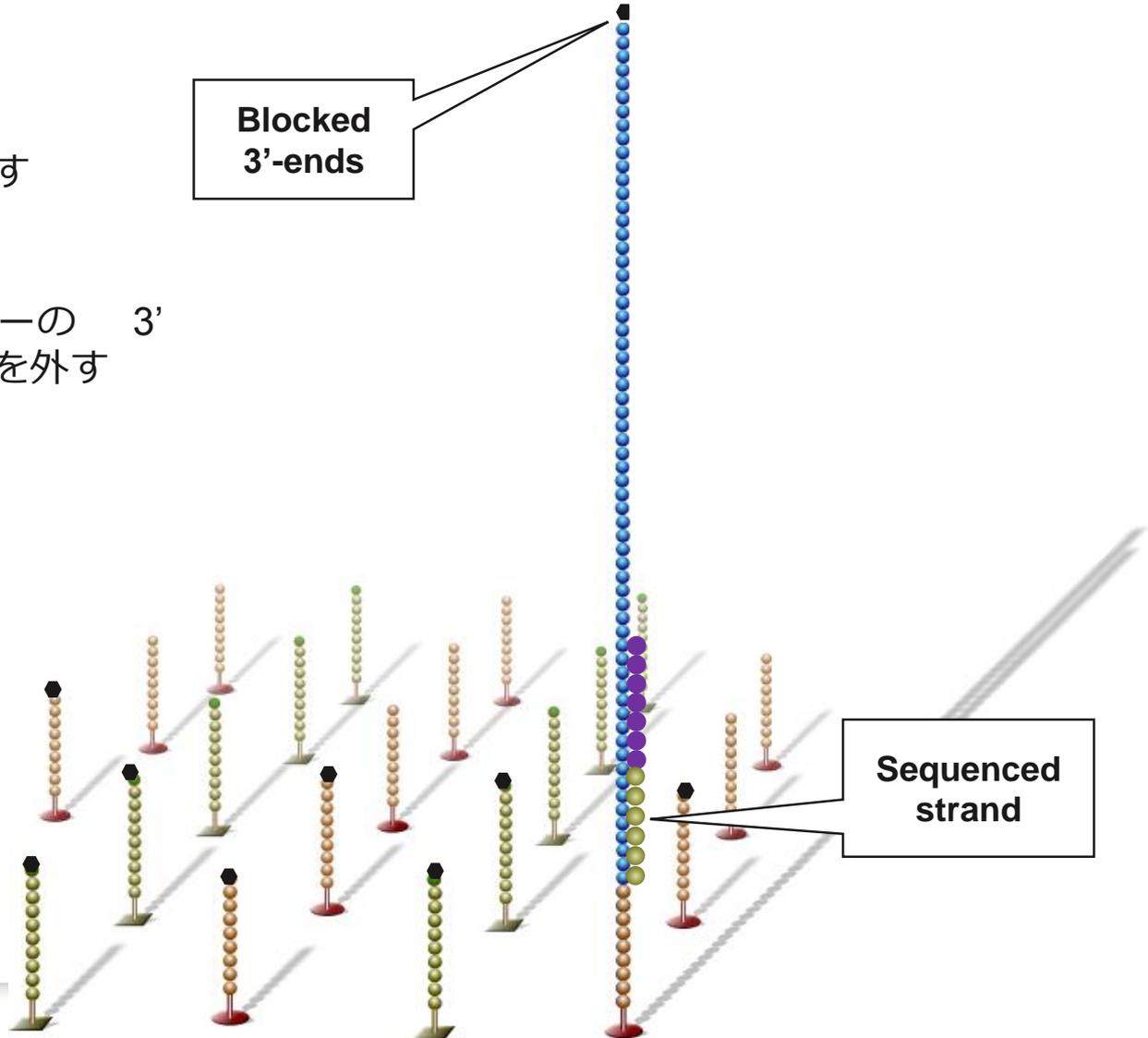
Paired End Sequencing



▶ シーケンス鎖を剥がす

▶ 鋳型と '芝' プライマーの 3' 末端のブロッキングを外す

Blocked
3'-ends

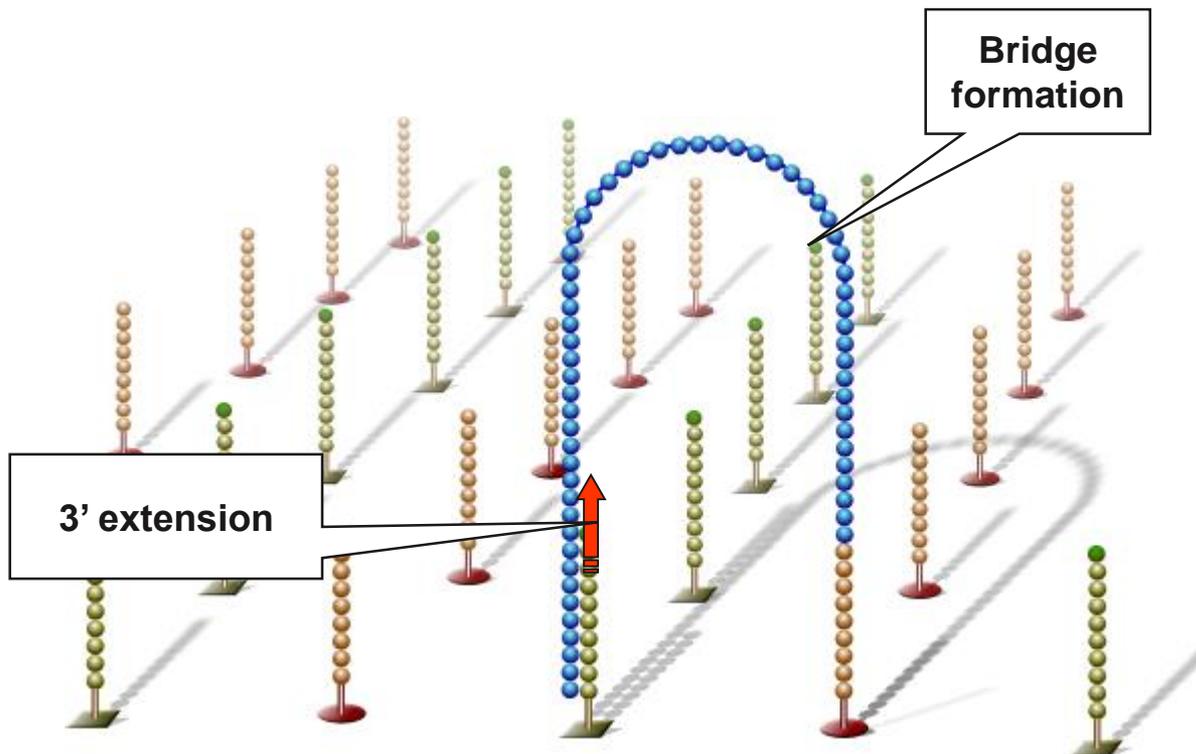


Sequenced
strand

Paired End Sequencing



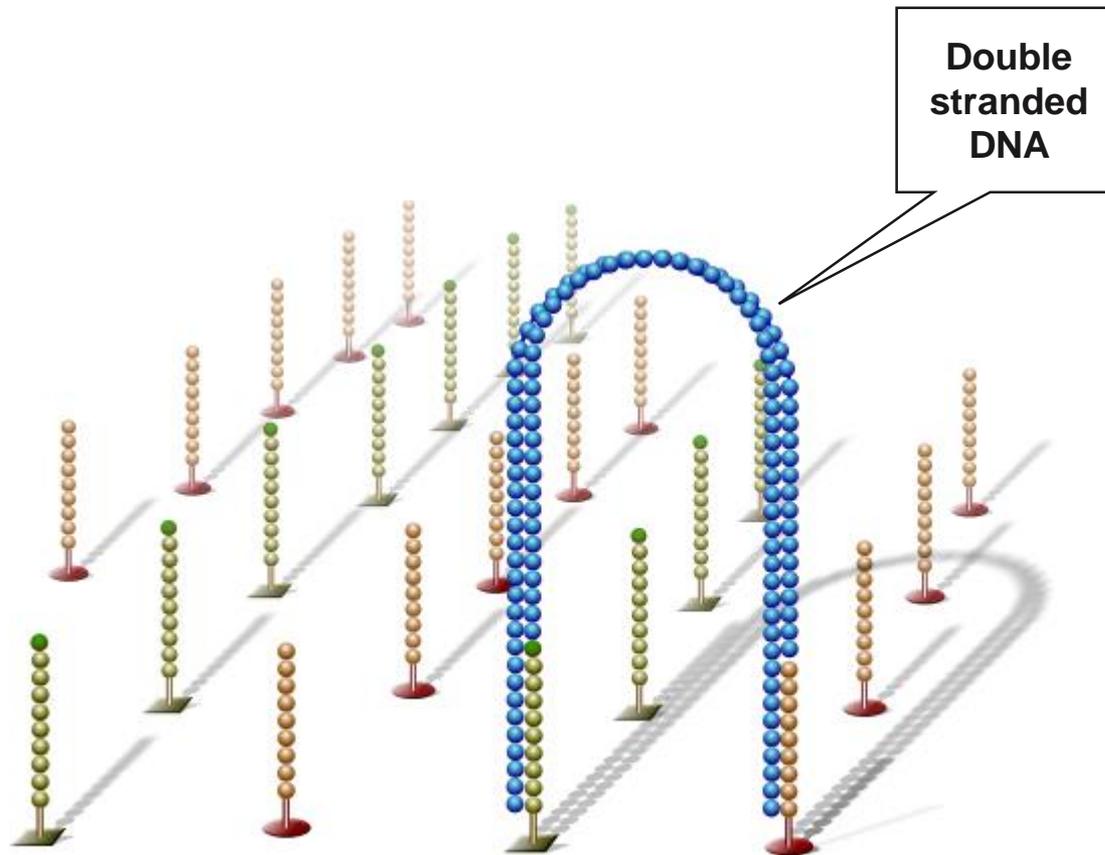
- ▶ 一本鎖の鋳型は曲がって'芝'のプライマーに結合してブリッジを形成する
- ▶ '芝' プライマーの 3' 末端から伸長する



Paired End Sequencing



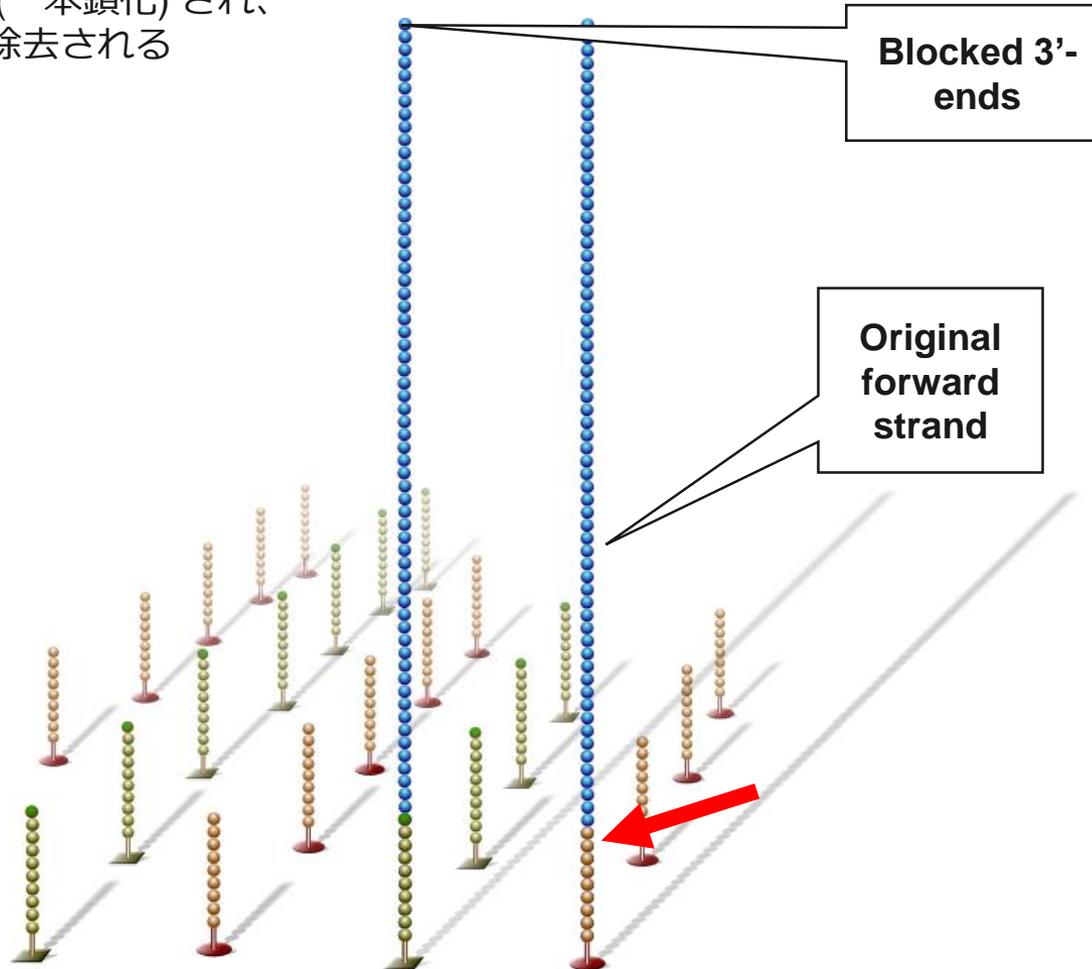
- ▶ '芝' プライマーの3'末端から一本鎖の鋳型相補的なDNA合成
⇒ 二重鎖DNAを形成



Paired End Sequencing



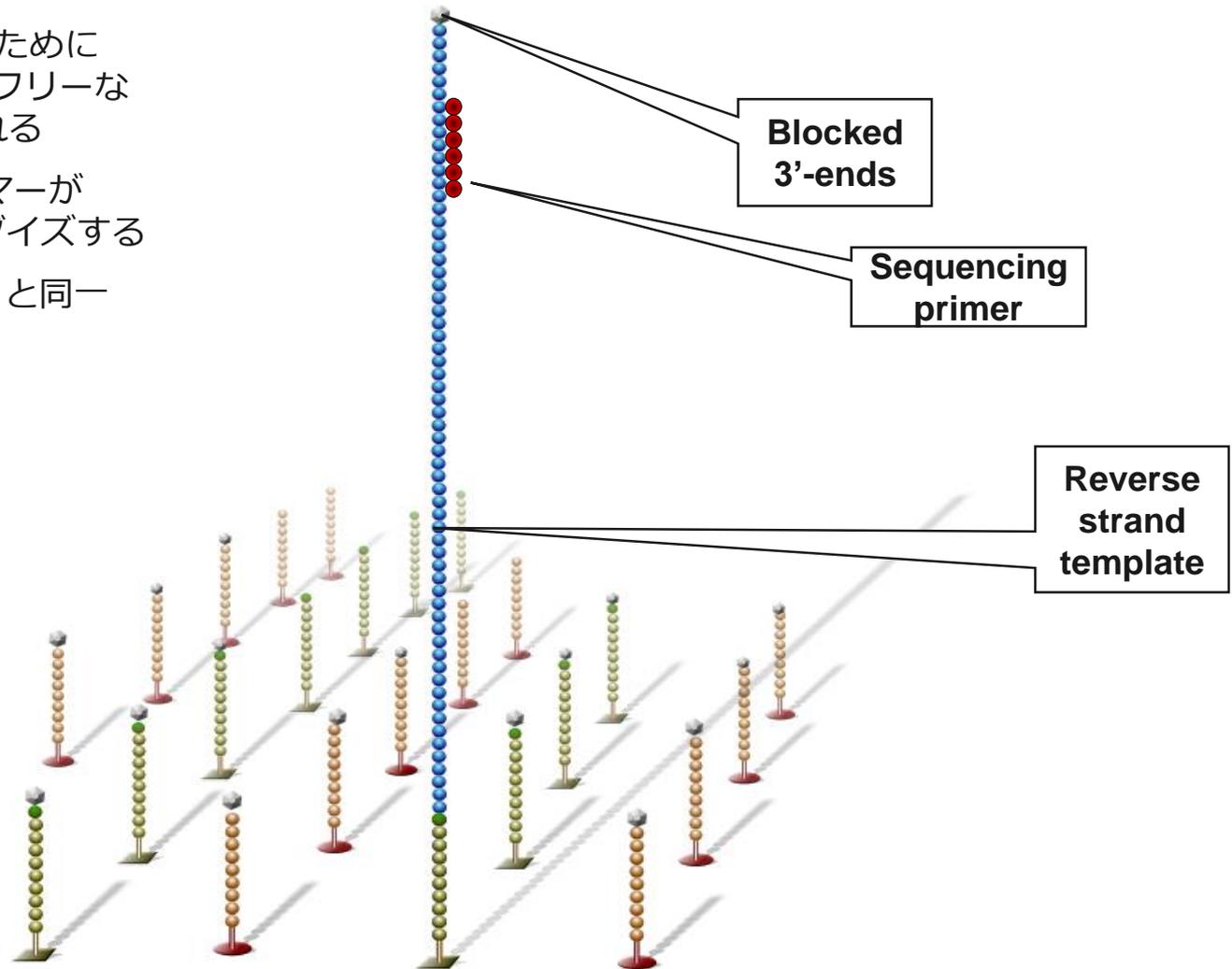
- ▶ ブリッジが直線化 (一本鎖化) され、元の順鎖の鋳型は除去される



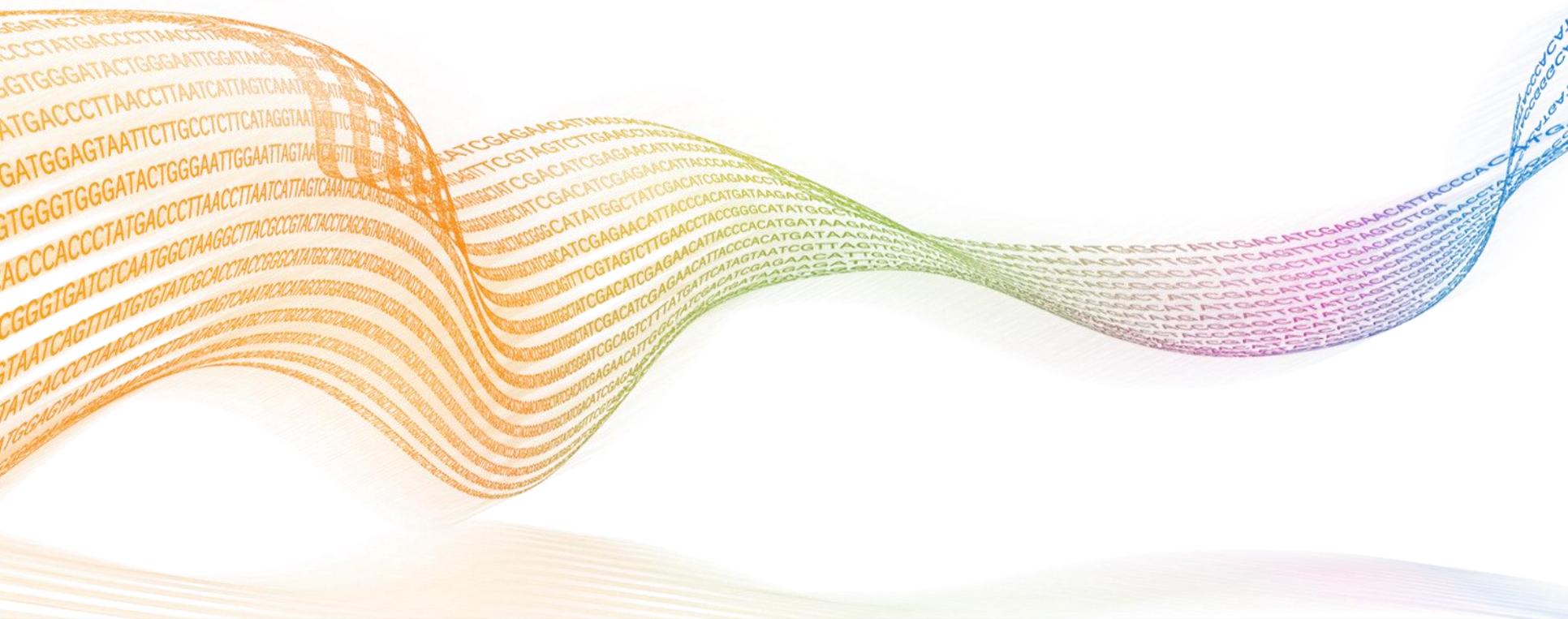
Paired End Sequencing



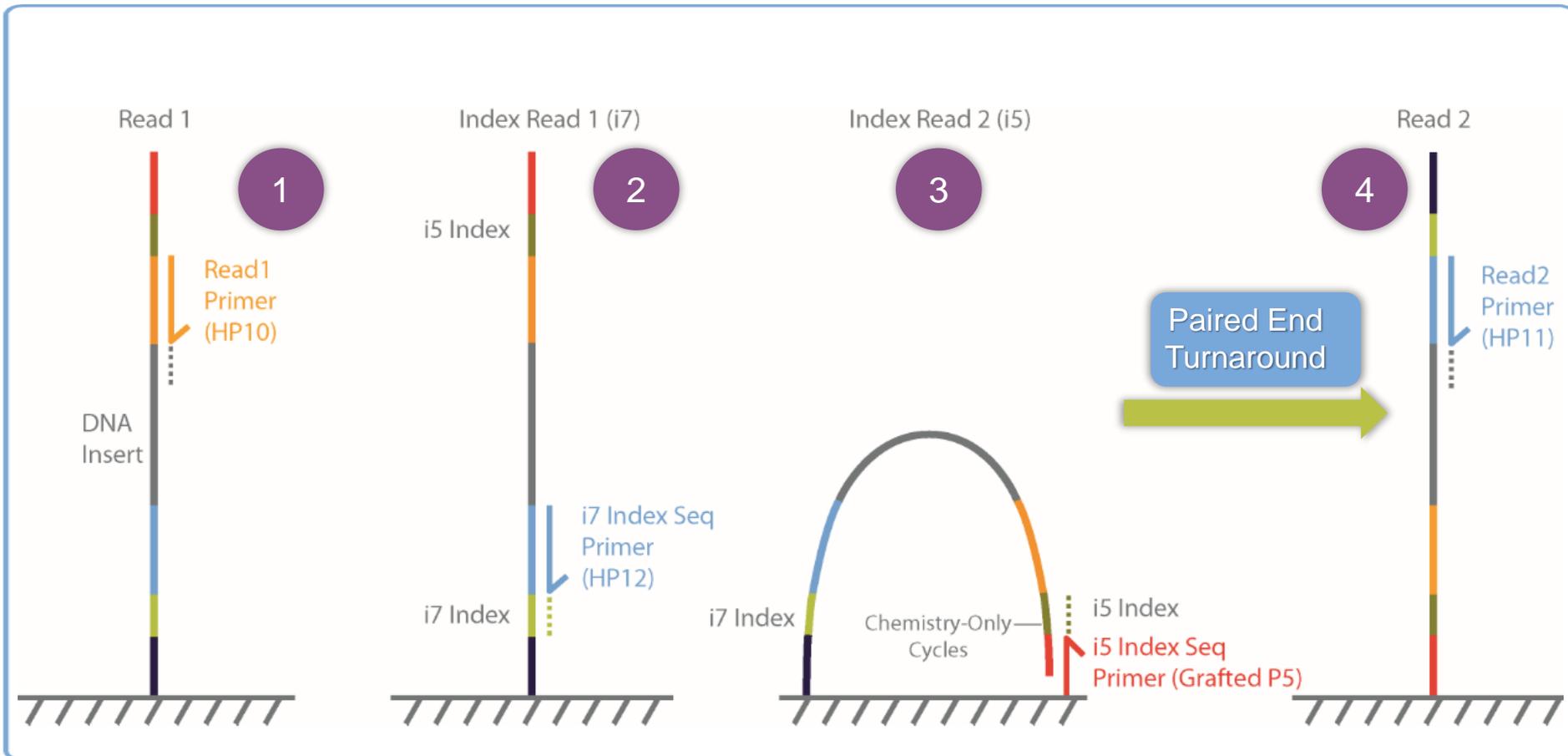
- ▶ 不要な DNA の結合を防ぐために逆鎖と'芝'のプライマーのフリーな3'末端がブロックされる
- ▶ Read2 シーケンスプライマーがアダプタ配列にハイブリダイズする
- ▶ シーケンスの手順はRead1と同一



Dual Indexシーケンスについて



Dual Index シーケンス (i5、i7の2インデックスを使用)



デュアルインデックス利用時には4ステップでシーケンスを実施

3

Dual Index Sequencing



Region complementary to P5 grafting primer

Index 2

P5 primer

DNA insert

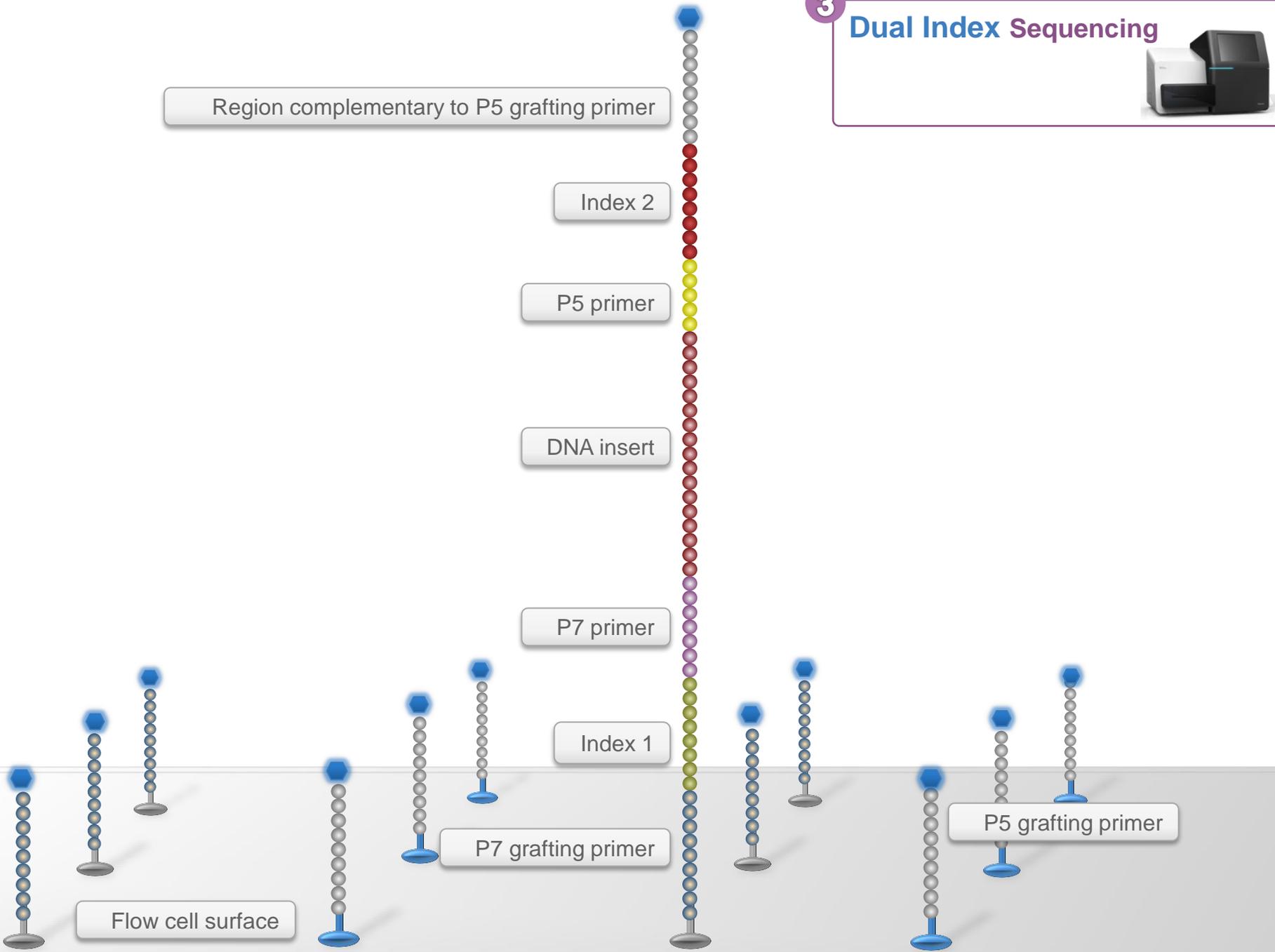
P7 primer

Index 1

P7 grafting primer

P5 grafting primer

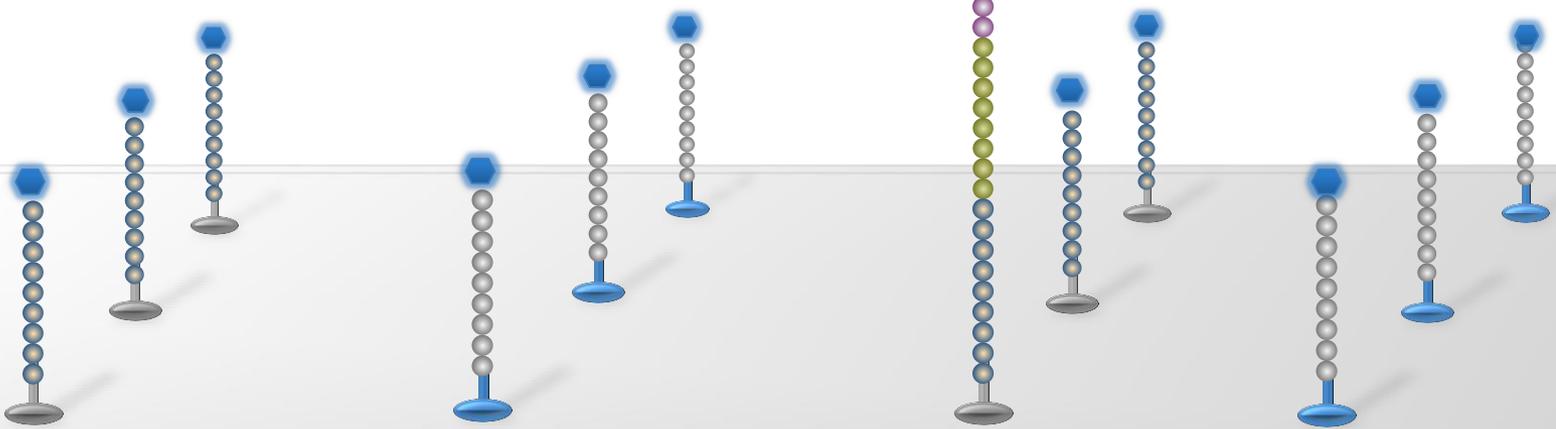
Flow cell surface





1. Read 1プライマーのハイブリ

SBS Sequencing Primer Hybridization

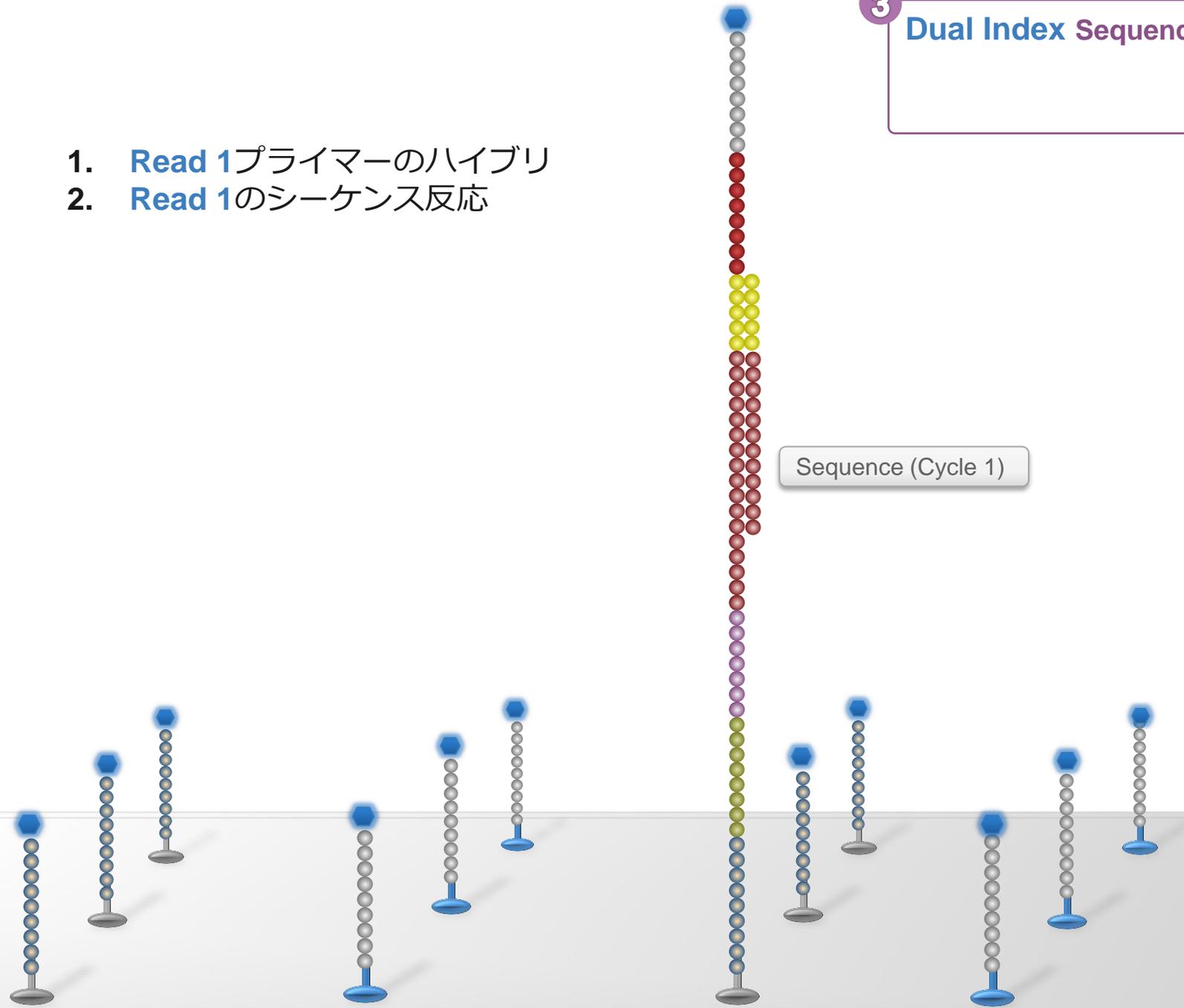


3

Dual Index Sequencing



- 1. Read 1プライマーのハイブリ
- 2. Read 1のシーケンス反応

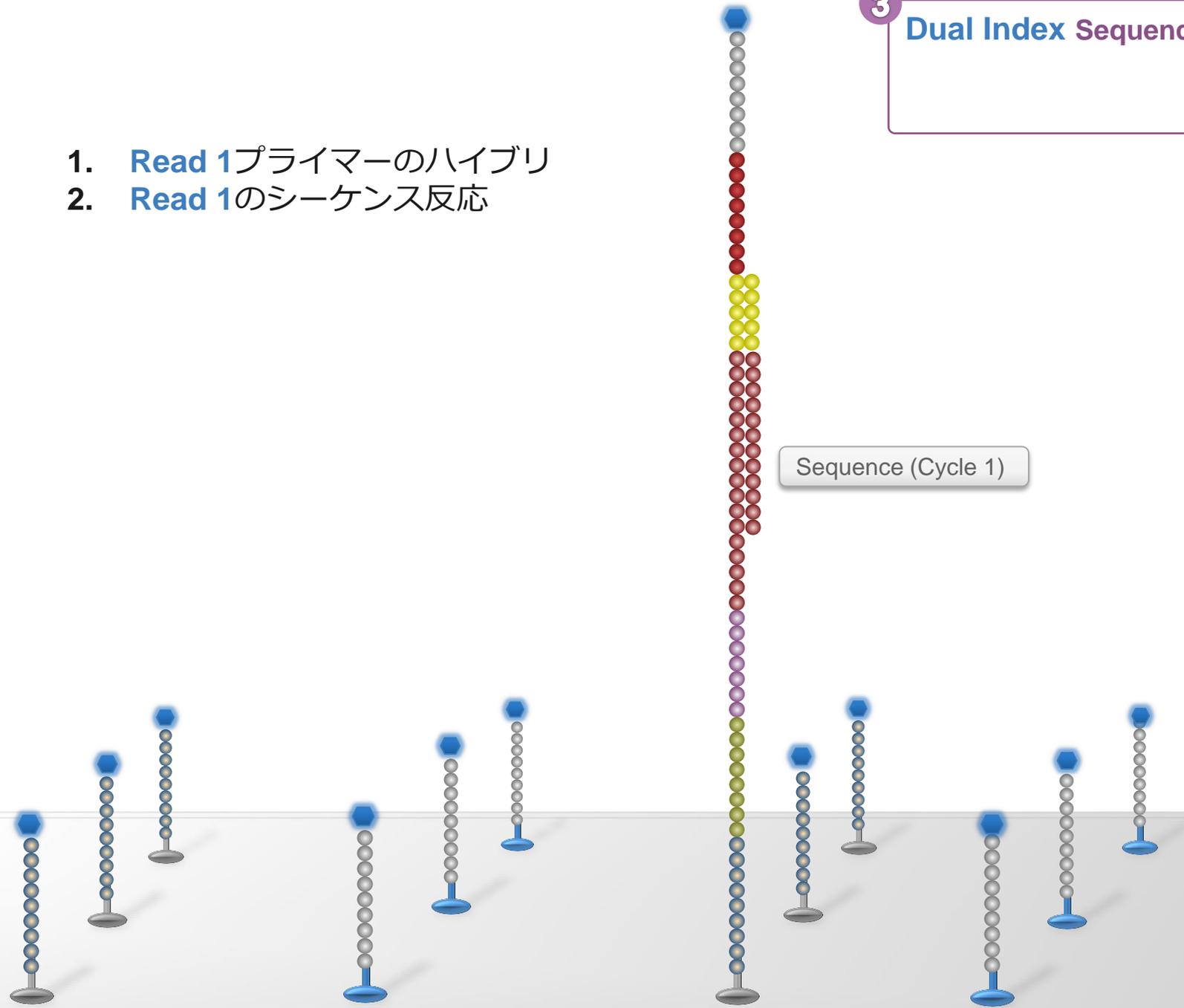


3

Dual Index Sequencing



- 1. Read 1プライマーのハイブリ
- 2. Read 1のシーケンス反応



Sequence (Cycle 1)

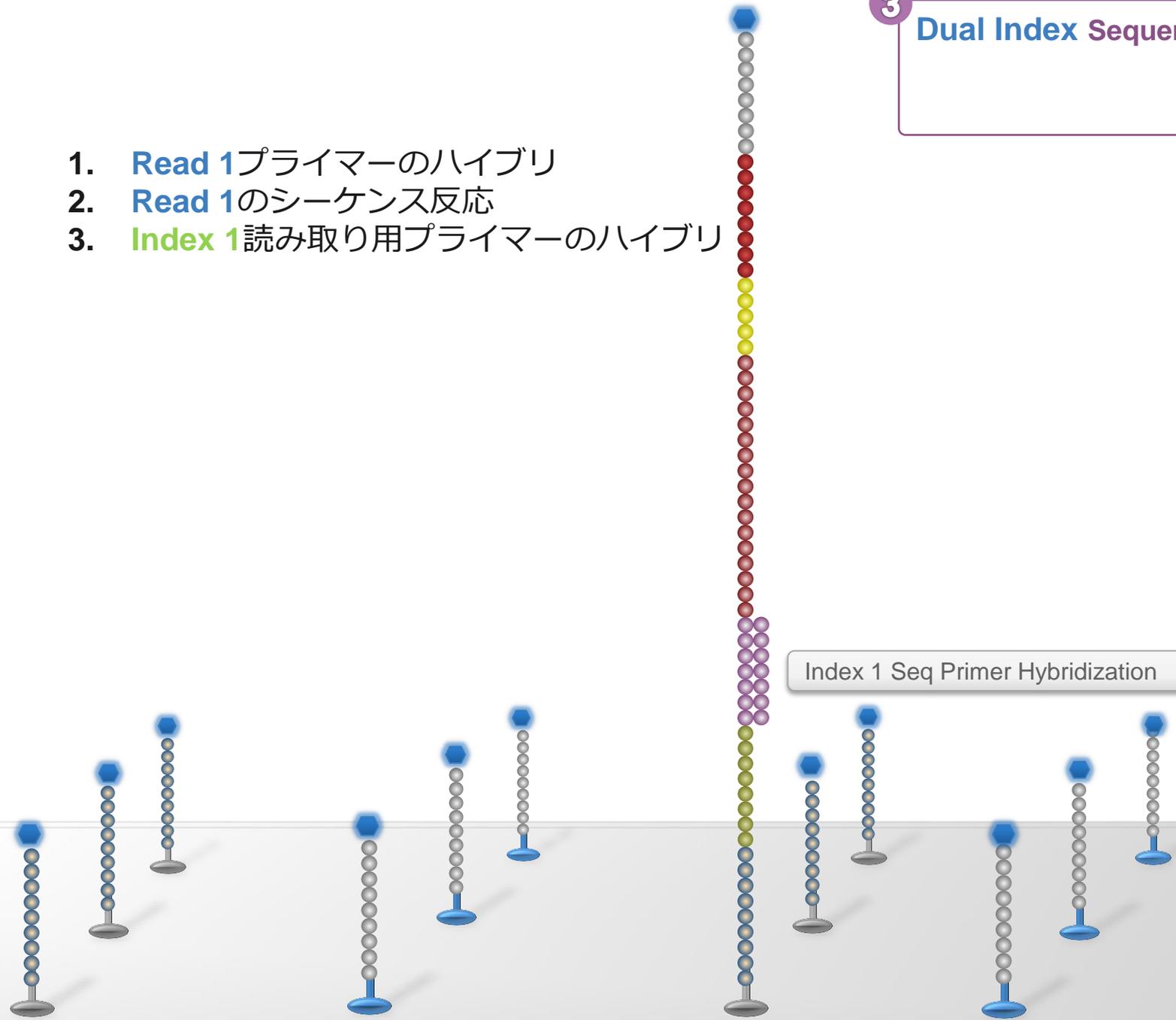
3

Dual Index Sequencing



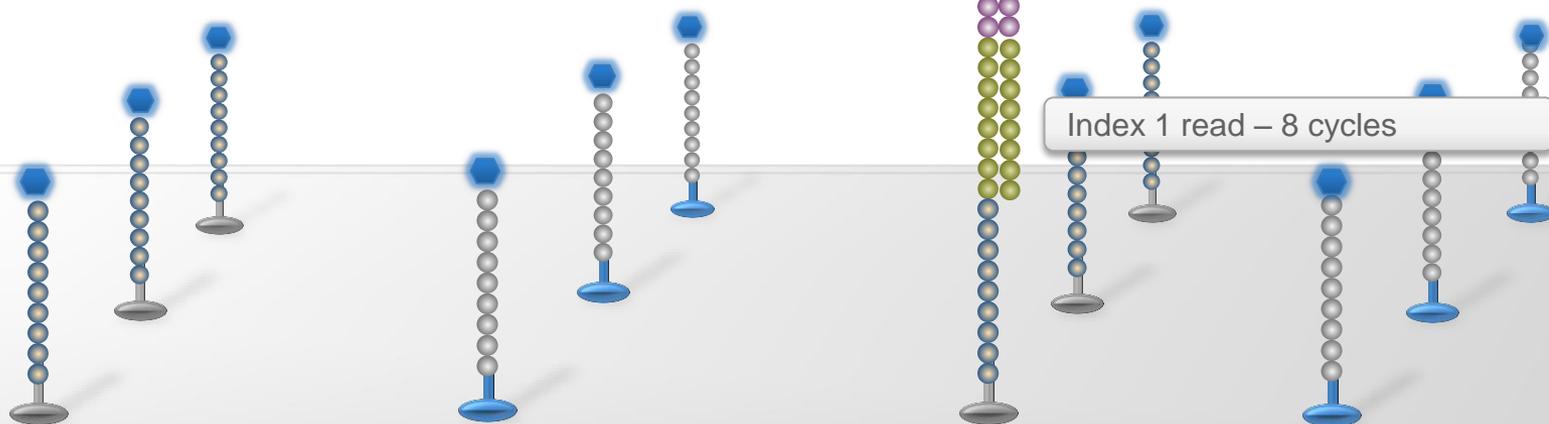
- 1. Read 1プライマーのハイブリ
- 2. Read 1のシーケンス反応
- 3. Index 1読み取り用プライマーのハイブリ

Index 1 Seq Primer Hybridization





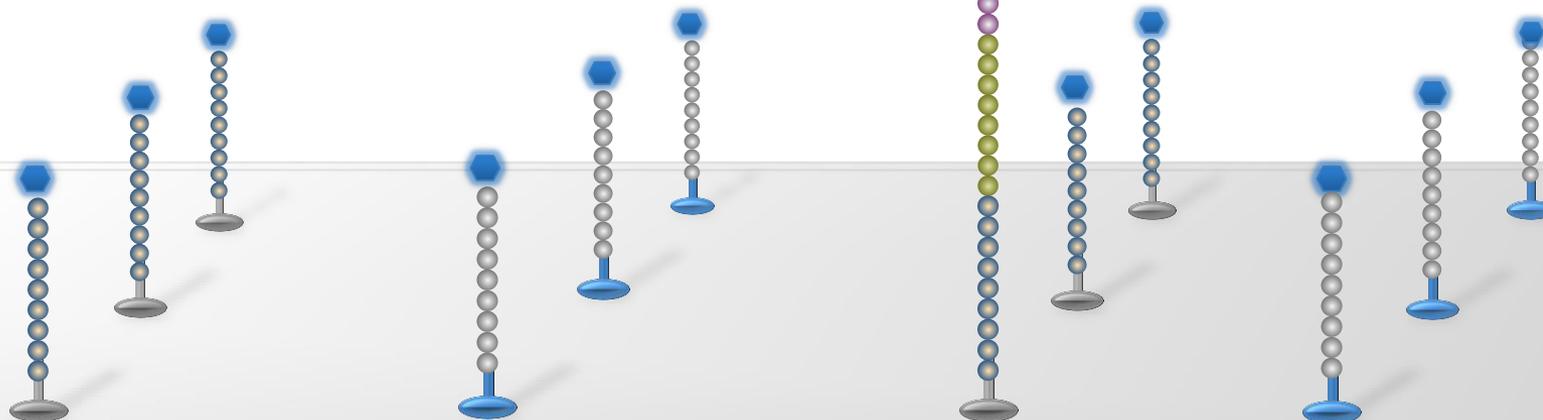
1. **Read 1** プライマーのハイブリ
2. **Read 1** のシーケンス反応
3. **Index 1** 読み取り用プライマーのハイブリ
4. **Index 1** の読み取り





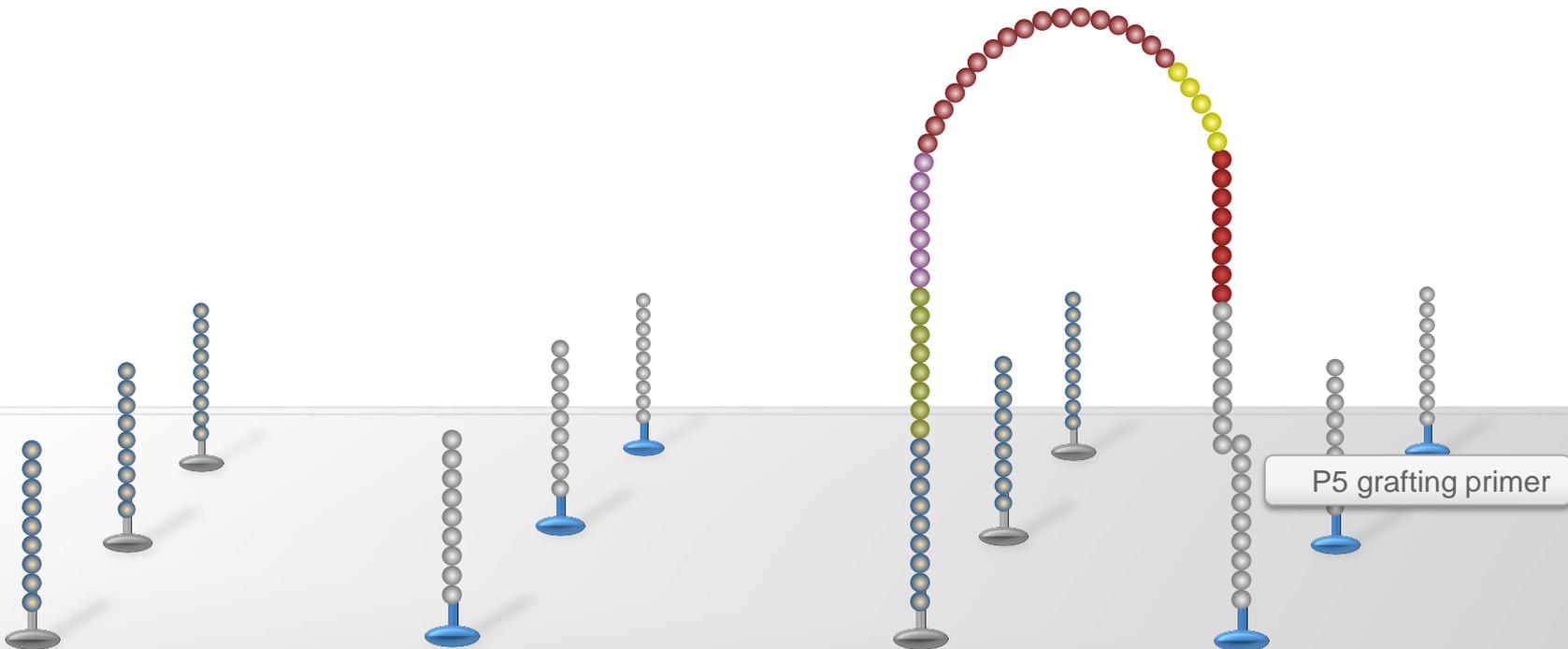
1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り

Unblock



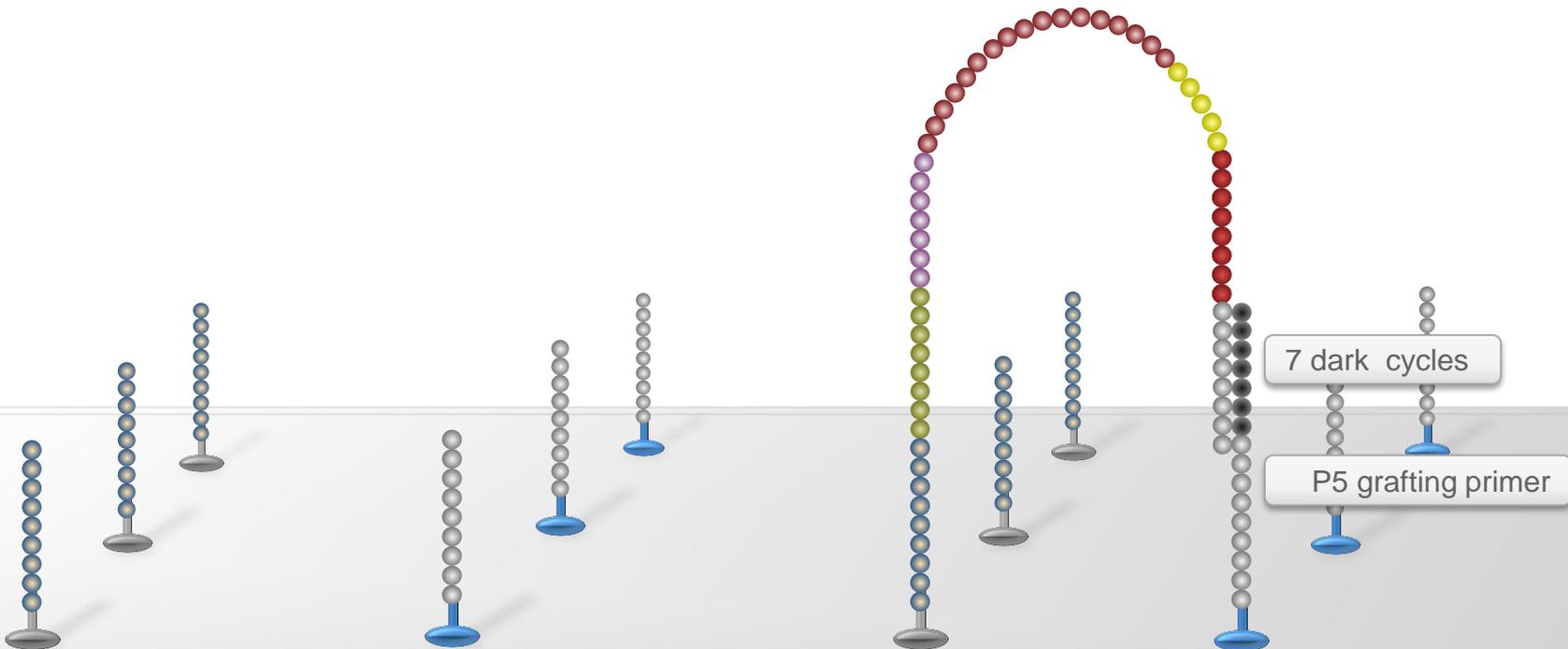


1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備



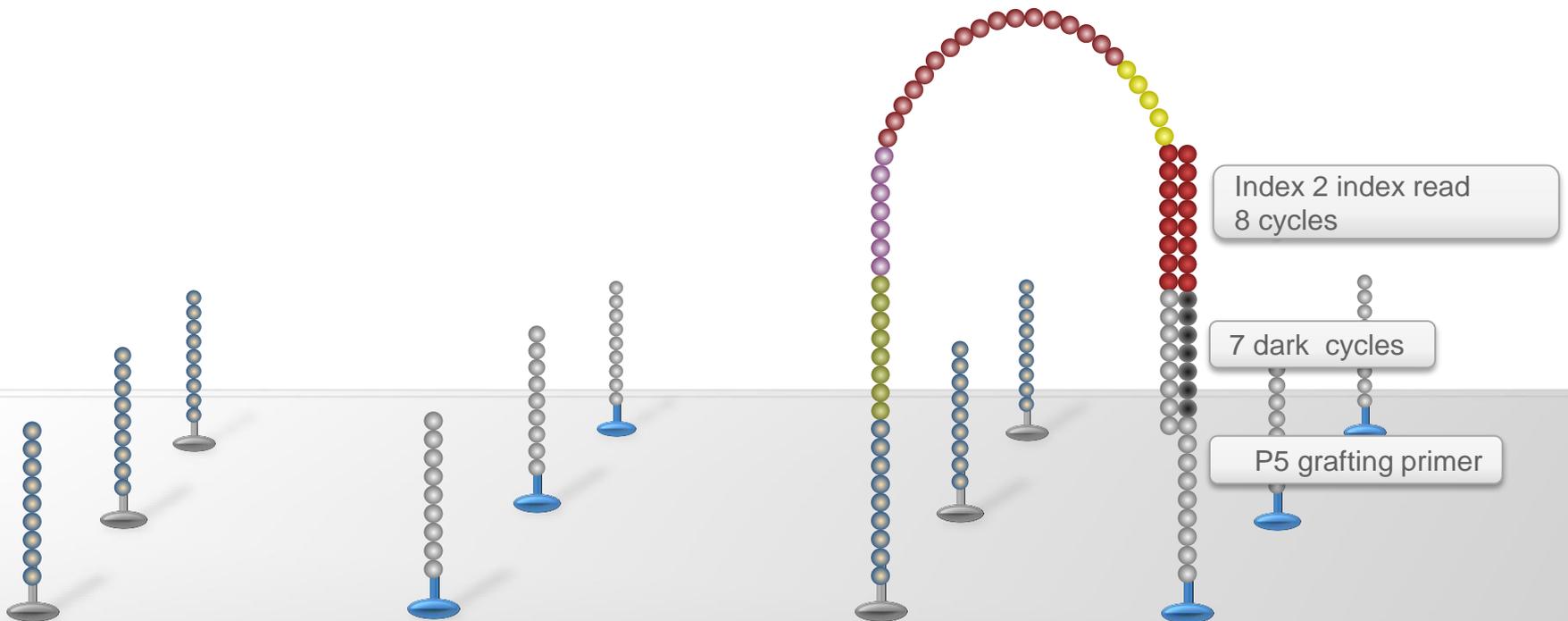


1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備





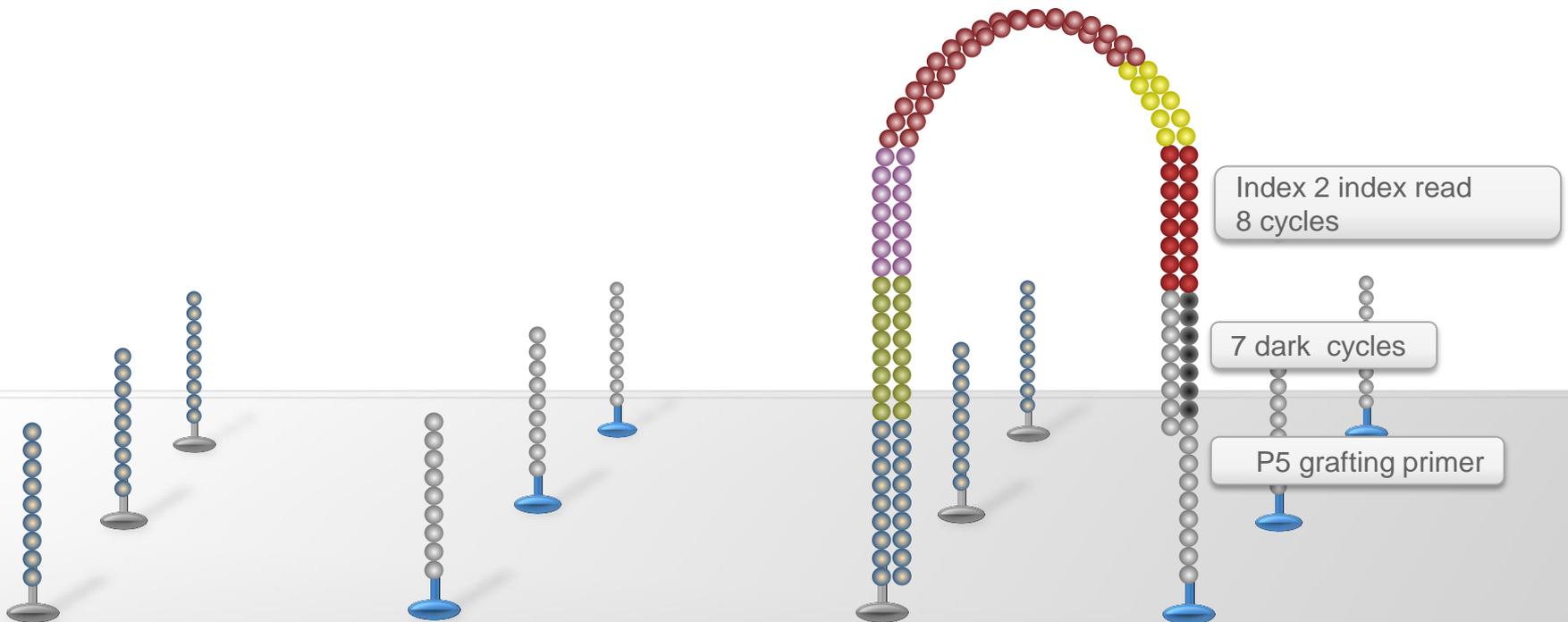
1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り



Dual Index Sequencing



1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り
7. **Read 2**向けに逆鎖合成



Dual Index Sequencing

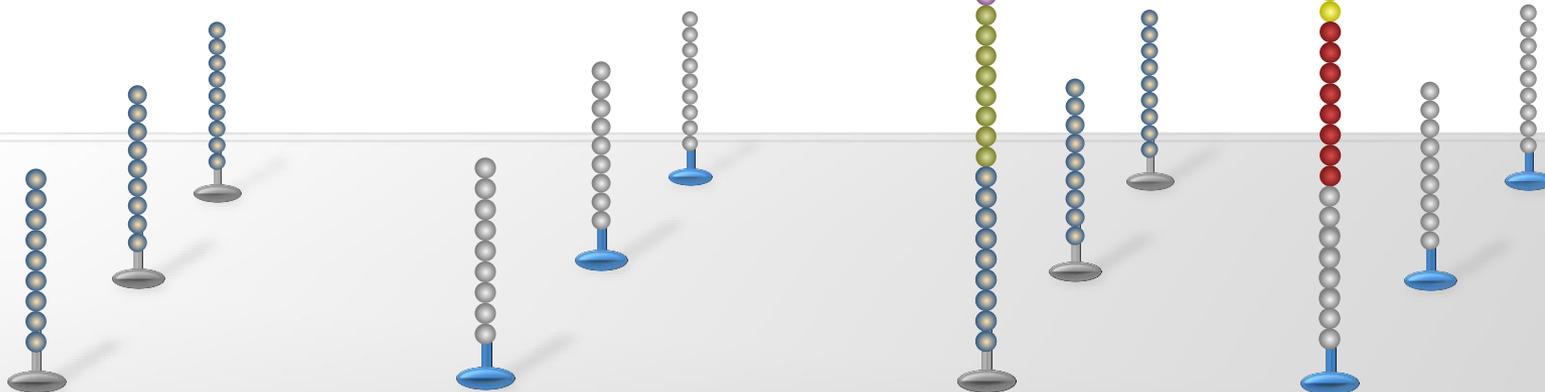


Linearization

1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り
7. **Read 2**向けに逆鎖合成

Original strand

New strand



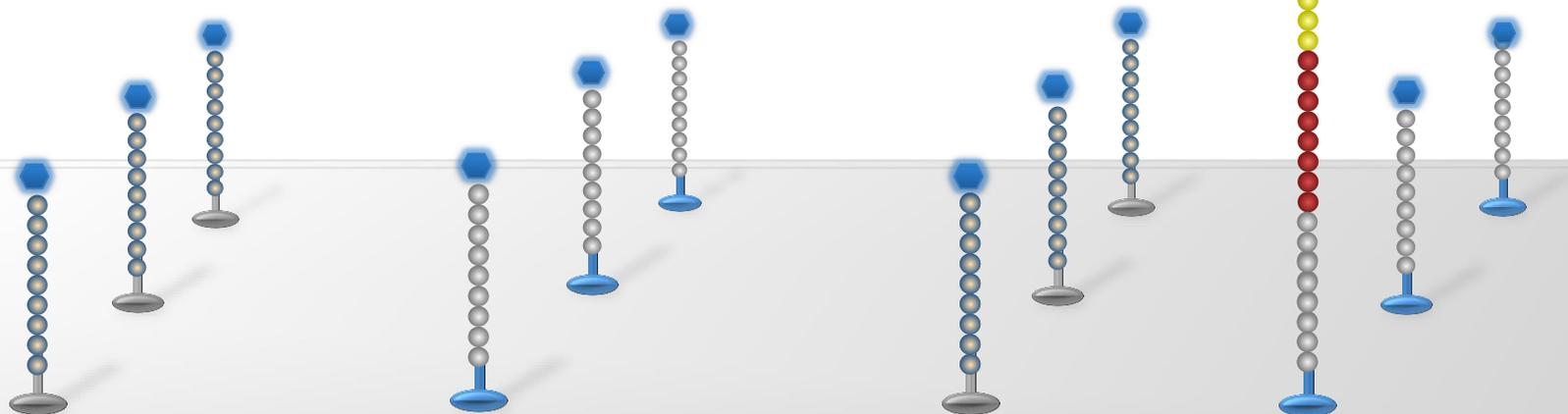
Dual Index Sequencing



Cleavage of Original Template

Reblocked

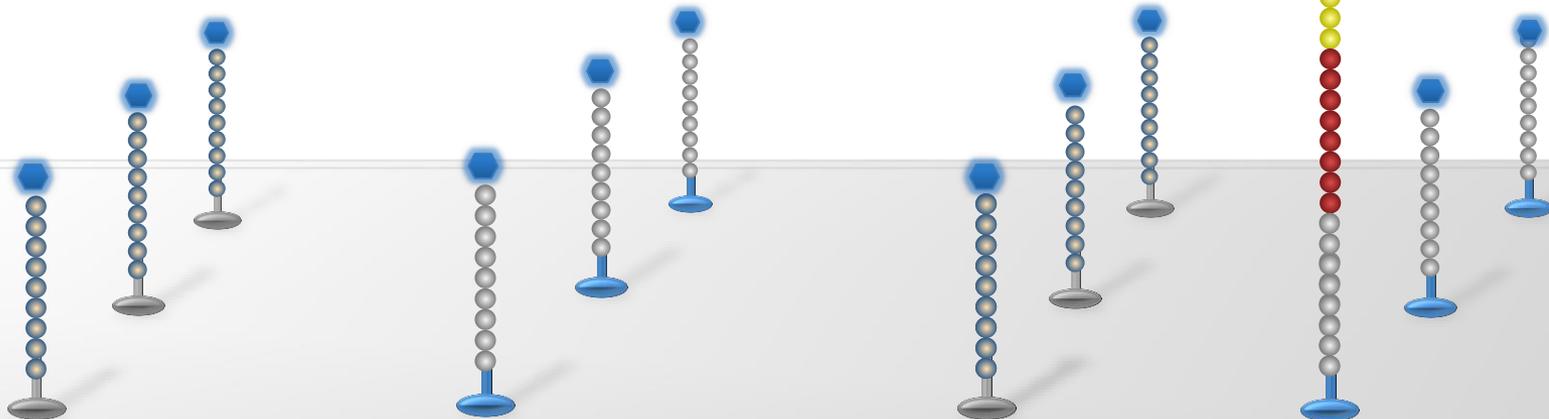
1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り
7. **Read 2**向けに逆鎖合成



Dual Index Sequencing



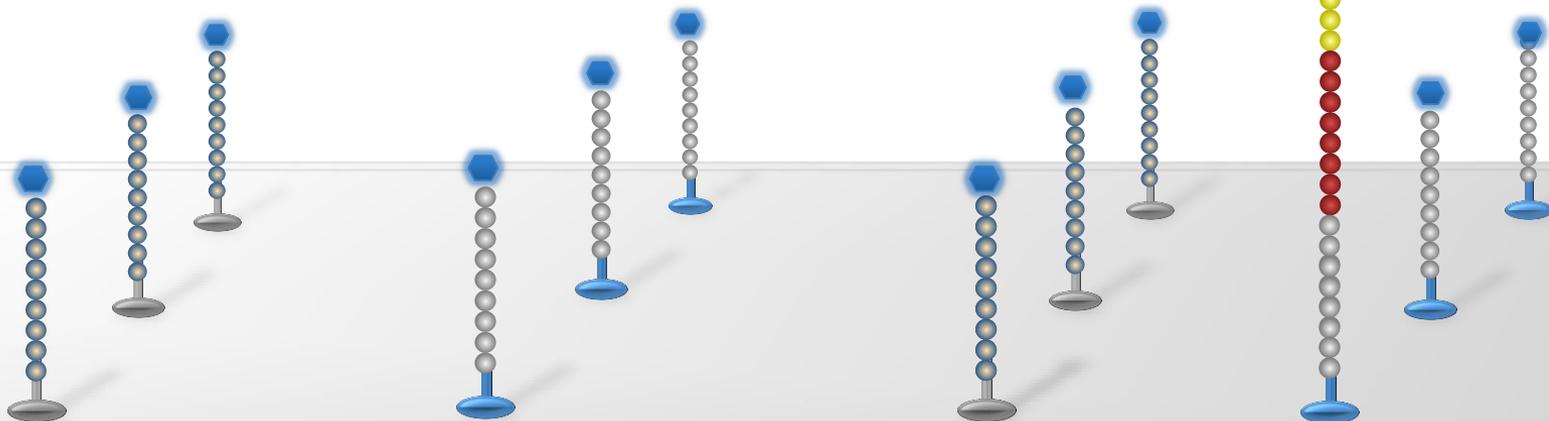
1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り
7. **Read 2**向けに逆鎖合成
8. **Read 2**プライマーのハイブリ



Dual Index Sequencing



1. **Read 1**プライマーのハイブリ
2. **Read 1**のシーケンス反応
3. **Index 1**読み取り用プライマーのハイブリ
4. **Index 1**の読み取り
5. **Index 2**の読み取り準備
6. **Index 2**の読み取り
7. **Read 2**向けに逆鎖合成
8. **Read 2**プライマーのハイブリ
9. **Read 2**のシーケンス反応



Sequence

シーケンス技術を詳しく学ぶには

- ▶ Technology Spotlight : イルミナシーケンステクノロジー
 - Pub. No. 770-2013-J001 20MAR2013
 - http://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/technote_illumina_sequencing_technology-j.pdf
- ▶ Technology Spotlight : イルミナ2色法SBS シーケンステクノロジー
 - Pub. No. 770-2013-J054 10APR2014
 - http://jp.illumina.com/content/dam/illumina-marketing/apac/japan/documents/pdf/technote_two-channel_sbs.pdf

2. 次世代シーケンサーの解読のしくみ

- 物理的、酵素、アンプリコンによって断片を生成し、アダプターを付加して、ライブラリーを調製
- 検出可能なシグナルを得るため、ライブラリーをフローセル上で増幅してクラスターを形成
- 1塩基ずつ伸長して、配列を解読

第1部 知っておきたいNGSの基礎

- ▶ 第1章 NGSの原理とワークフロー
- ▶ 第2章 NGSを扱う上で必要な基礎用語

第1部 知っておきたいNGSの基礎

▶ 第2章 NGSを扱う上で必要な基礎用語

1. 次世代シーケンサーを扱うための必須単語
2. アプリケーションに必要な用語

次世代シーケンサーデータ解析の必須用語

シングルリード/
ペアエンド リード

スループット

リード長

リード数

カバレッジ

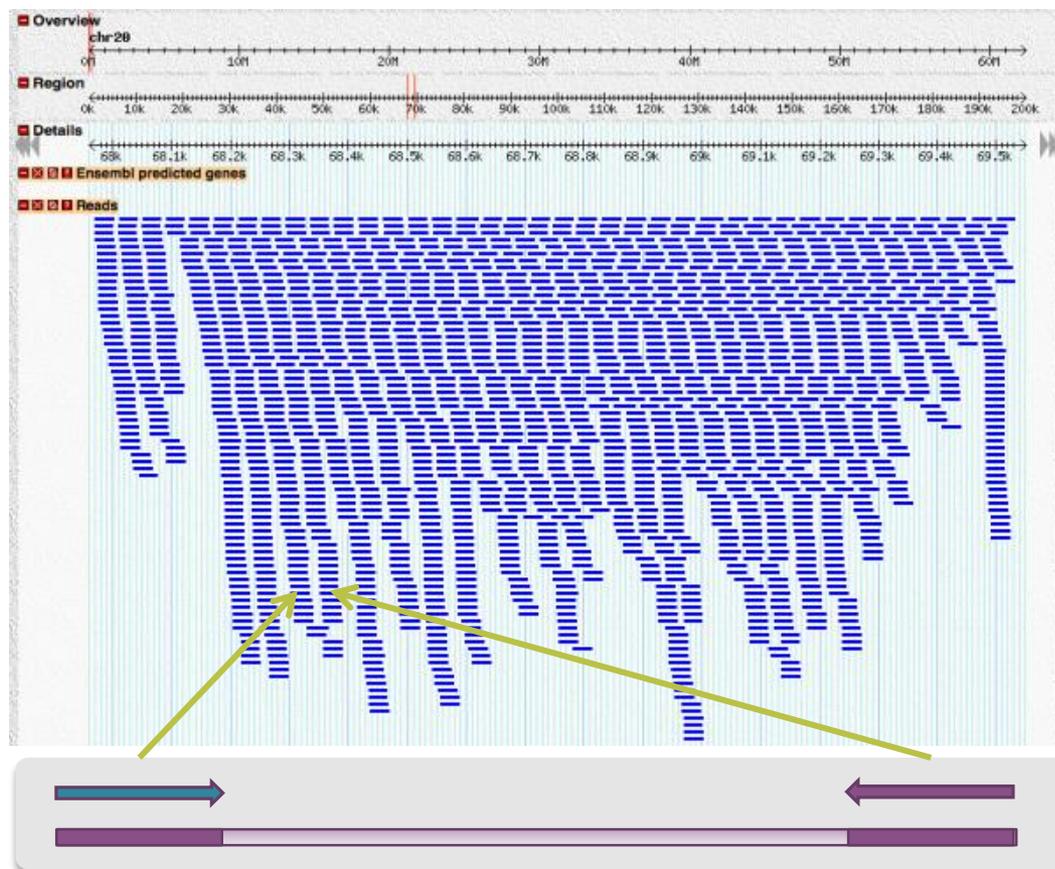
マッピング/
アライメント

バリエーション
コール

Qスコア

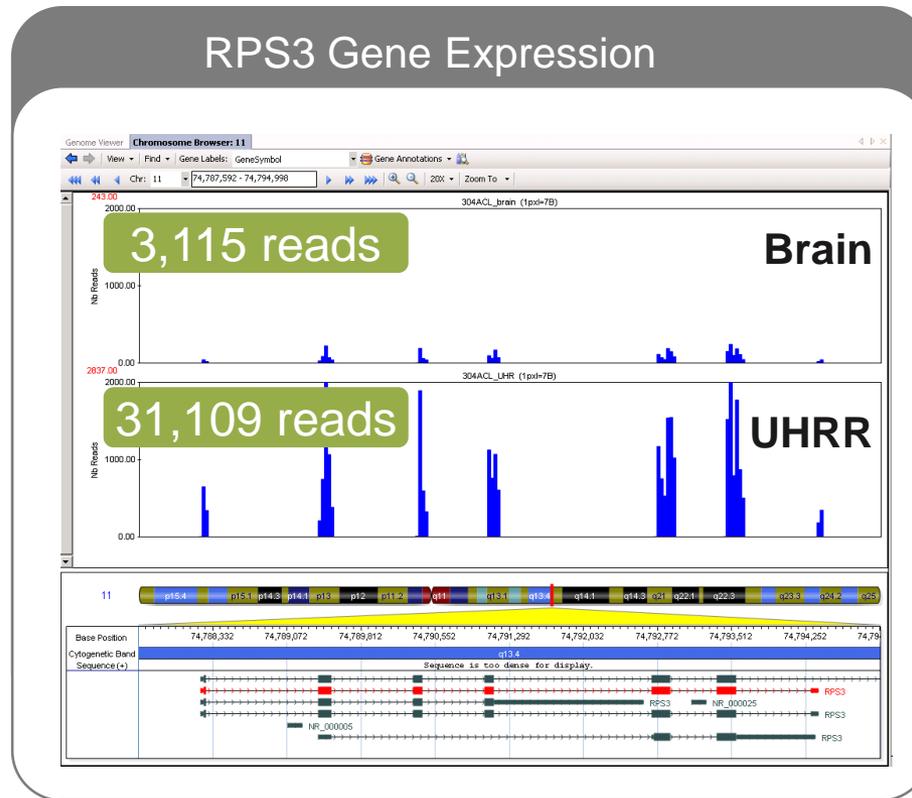
マッピング/アライメント

- ▶ 得られたリードを参照配列と照らし合わせることで位置を同定すること



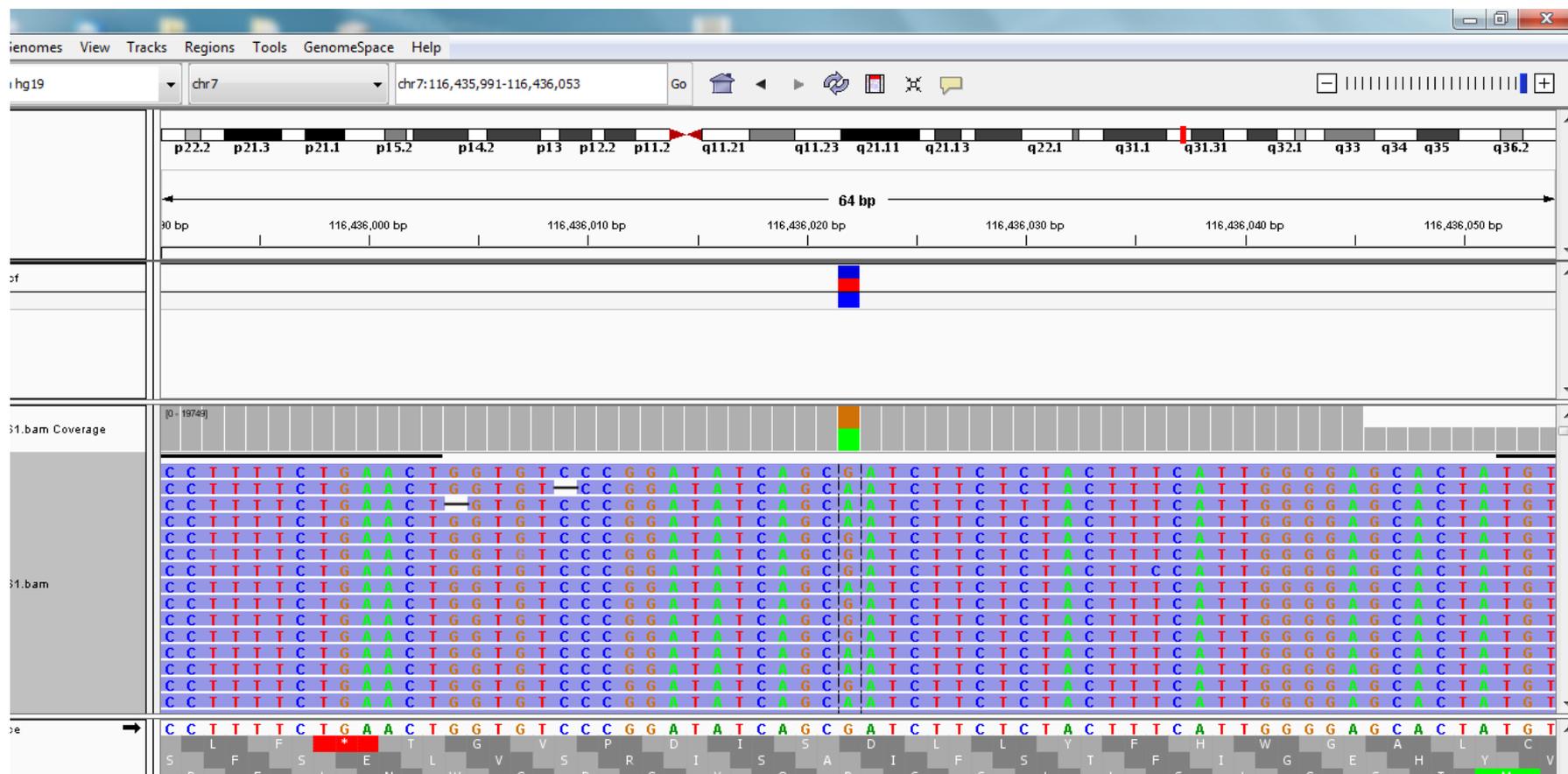
マッピング

- ▶ 参照配列の位置が同定されたリードから、頻度情報を得ること



バリエントコール

- ▶ 参照配列の位置が同定されたリードから、差異を同定すること



シーケンス方法

シングルエンド法: 調製したDNA断片のインサート部分の片側から読む方法。



ペアエンド法: 調製したDNA断片のインサート部分を両側から読む方法。

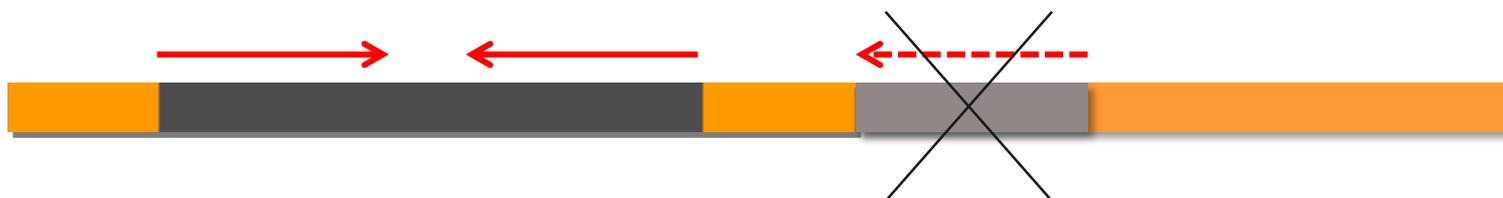
シングルエンド法よりも、より多くの情報を精確に得たい場合に活用。
得られるデータ量は、シングルエンド法の2倍になる。



メイトペア法: 長い(数kb)DNA断片の両端のみを読む方法。(第二部参照)



ペアエンドシーケンスの利点



誤った位置へのアライメント防止
de novoアセンブリの向上

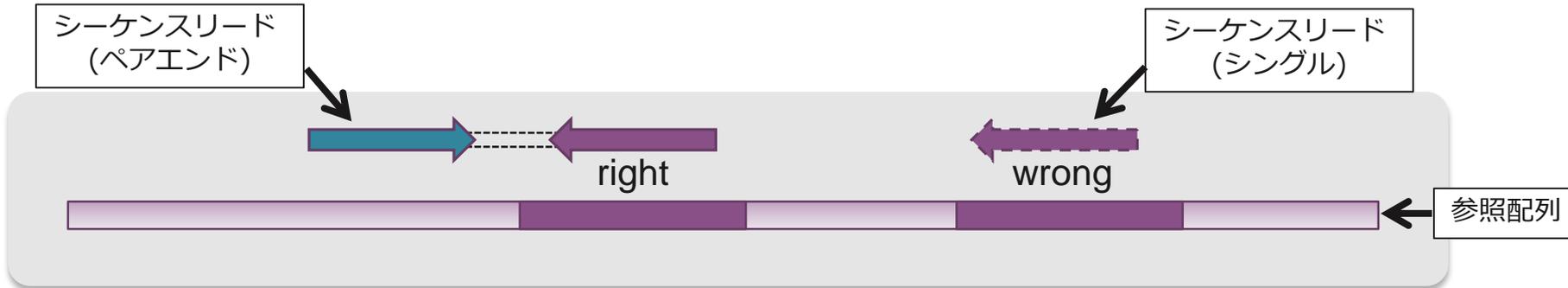


長いライブラリーの解読



高精度な解読
データ量の確保

Paired End Sequencing



“DNA 断片の両端をシーケンス”

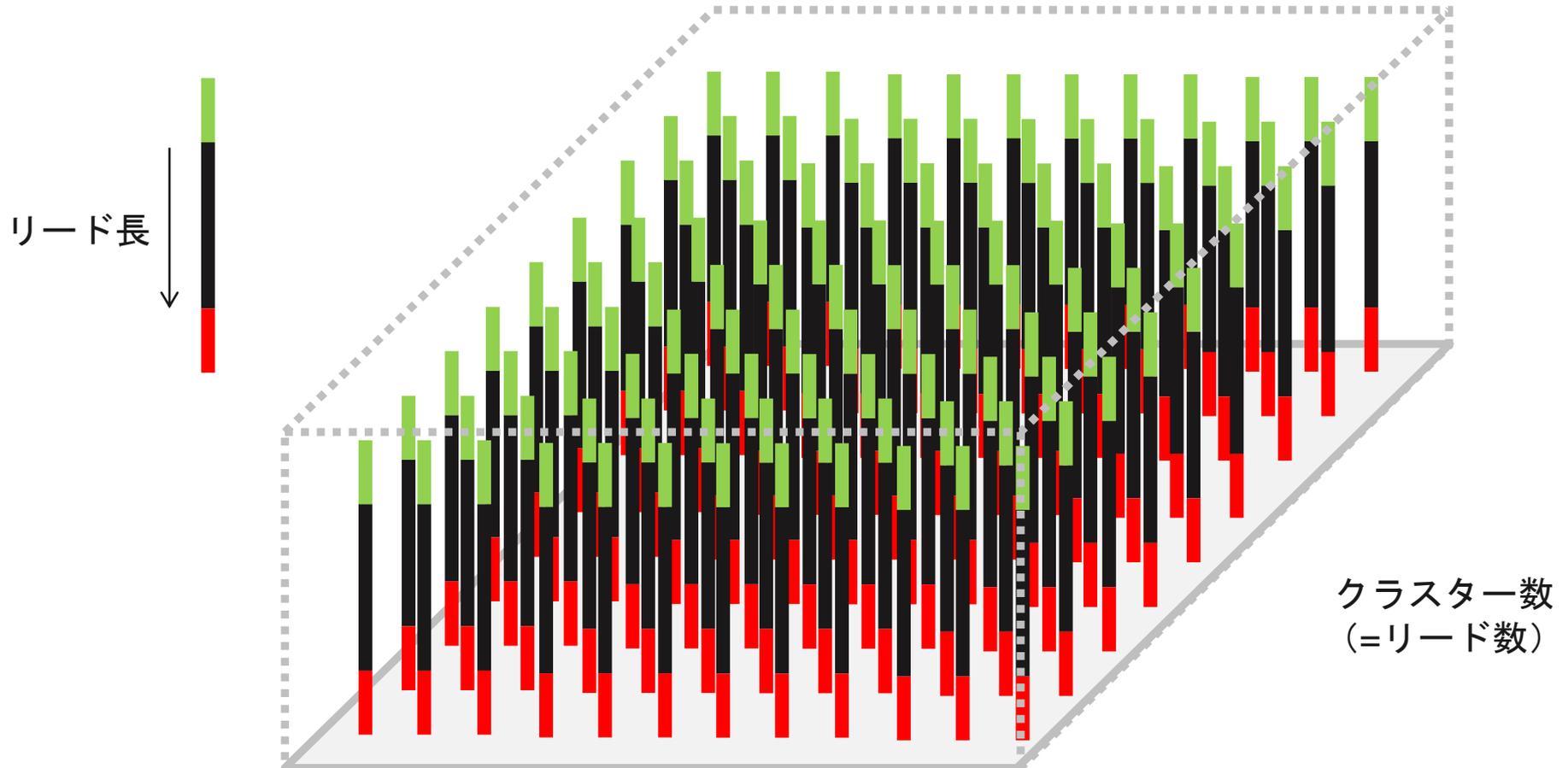
ライブラリ調製時に断片長を明確にしているため、片側末端がユニークヒットすることで、もう一端の曖昧な配置がはっきりとする

- ▶ 多くの短いリードのアプリケーションにとって重要；
 - Repeat sequences
 - Rearrangements
 - *De novo* assembly
 - Di-tag sequence cDNAs, ChIP, etc.
 - BAC-end sequencing

スループット \cong リード長 \times クラスタ数

▶ シングルリードの場合

スループット

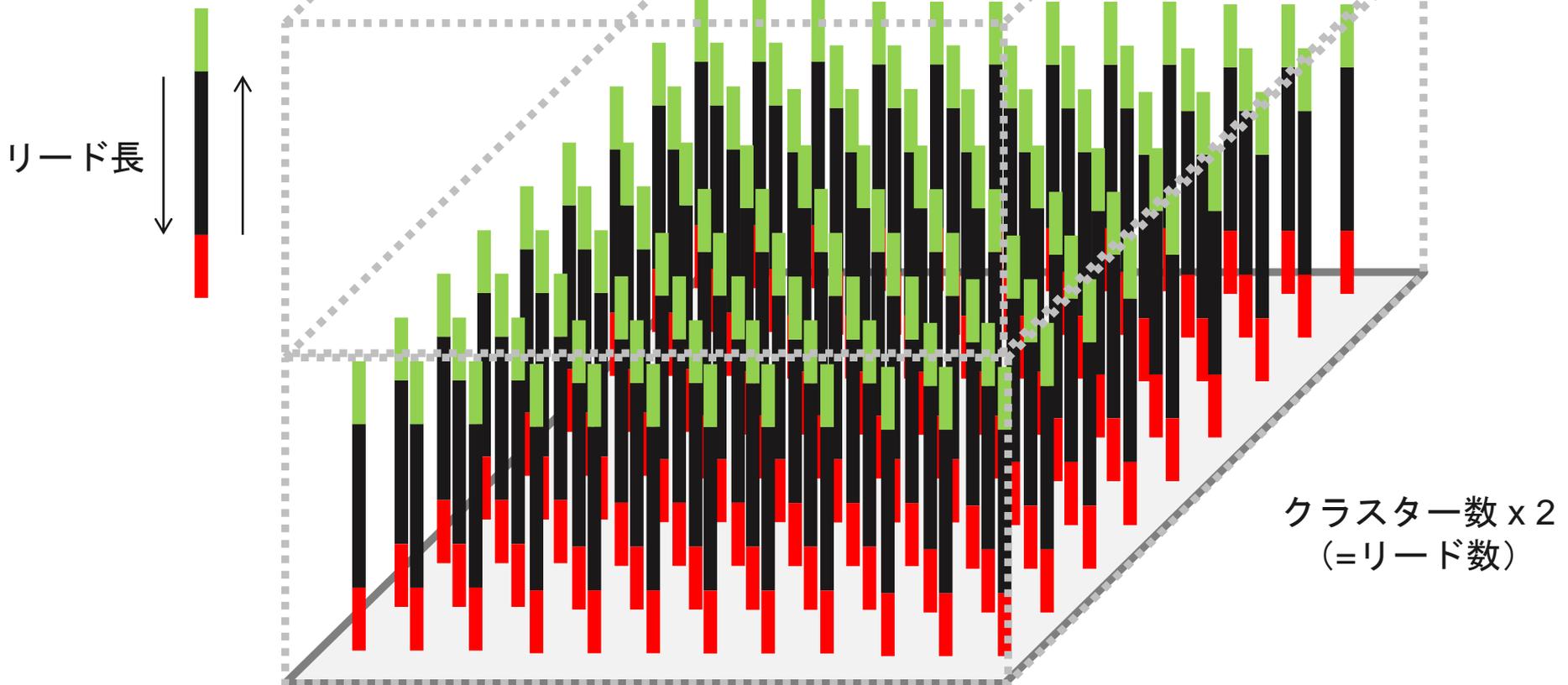


スループットとは、解読できる総データ量（塩基数）

スループット \equiv リード長 \times クラスタ数

▶ ペアエンドリードの場合

スループット



スループットとは、解読できる総データ量（塩基数）

リード長とクラスター数はどのように選ぶのか？

▶ MiSeqの場合

リード長：50~300bp

クラスター数：15Mまたは25M

MiSeq Reagenet Kit v3		150		600	
スループット		3.75 Gb		15 Gb	
リード長		75bp x 2		300bp x 2	
クラスター数	25 M				
MiSeq Reagenet Kit v2		50		300	500
スループット	750Mb		4.5Gb	7.5Gb	
リード長	50bp x1		150bp x2	250bp x2	
クラスター数	15 M				

リード長を選択する基準

短いリード長

25bp

50bp

100bp

150bp

250bp

300bp

10kb
(Synthetic)

長いリード長

アプリケーション
に合わせて選択

Small RNA

ChIP-Seq
mRNA

Exome
Whole Genome

De novo assembly
16S Metagenomics

Longer PCR products

注意点

○少ないデータ量で良い
○高品質なリードを生成
○ラン時間が短い
×アライメントの重複

○リードの重ね読みが可能
○より多くのデータが得られる
○良好なアライメント
△リードの後半で精度が低下
△より長いラン時間

Paired End Sequencing



Reference

This is really the best way to do sequencing

Single-reads

This is

...

is really

...

really the

...

the best

...

sequencing

Paired-reads

This is (----100 characters-----) sequencing

アセンブリが容易になる

[Skip Overview](#)

カバレッジ（デプス）とは？



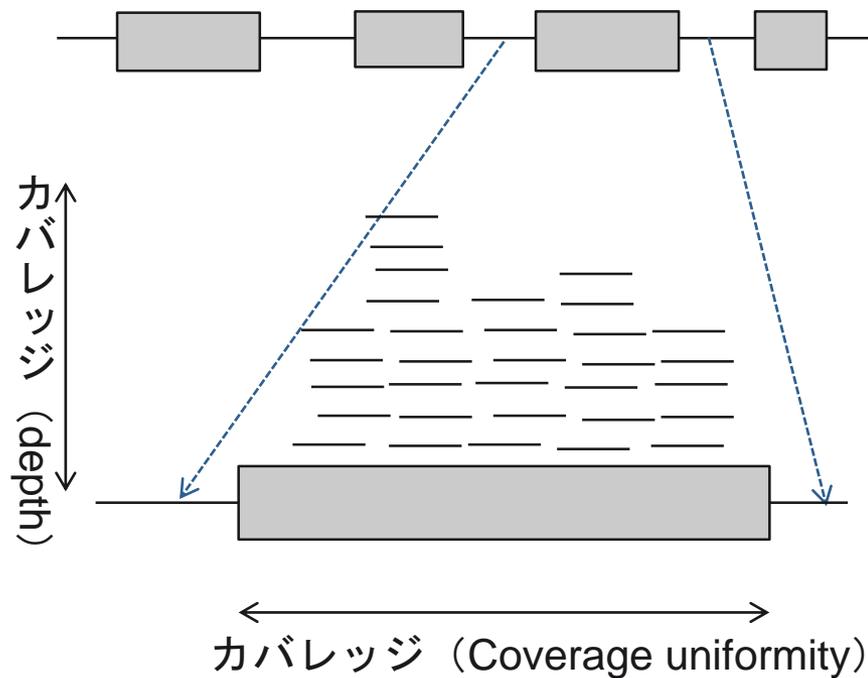
デプス：深度
(Read depth)
または
カバレッジ
(Fold coverage)

カバレッジ
(Coverage uniformity)

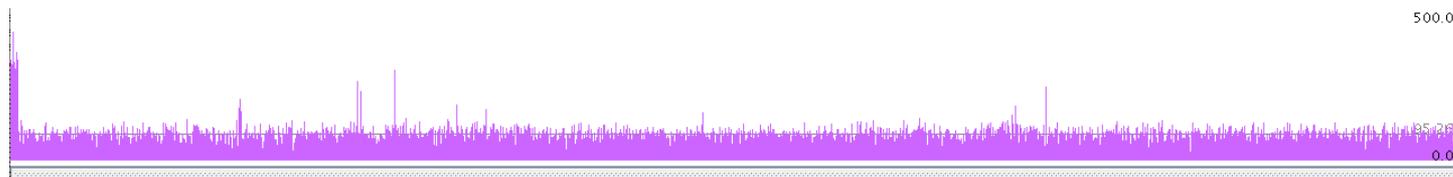
アプリケーション毎に異なる必要なカバレッジ

生物種	シーケンス手法	ライブラリ調製キット	検出目的	推奨平均カバレッジ (最少値)
微生物	Whole genome sequencing	TruSeq PCR Free	変異解析	30x
	Whole genome sequencing	Nextera XT	変異解析	40x
	<i>de novo</i> Assembly	Nextera MatePair	ゲノム配列決定	60x
ヒト	Whole genome sequencing	TruSeq PCR Free	生殖細胞変異	30x
	Exome Sequencing	Nextera Rapid Capture	生殖細胞変異	100x
	Amplicon Sequencing	TruSeq Custom Amplicon	生殖細胞変異	30x
	Amplicon Sequencing	TruSeq Custom Amplicon	体細胞変異 (アレル頻度 5%)	500x

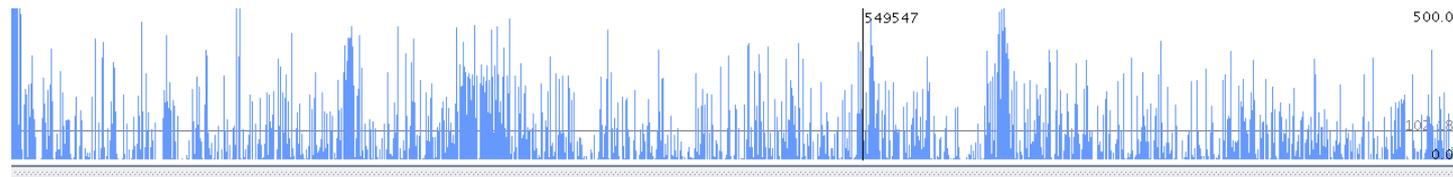
もう一つのカバレッジ (Coverage uniformity)



均一な例



不均一な例



シーケンス条件の決定 1回に処理できるサンプル数

どれくらいのサンプルを
解読するか？

MiSeq 150bp x2 = 4.5Gb
NextSeq 150bp x2 = 120Gb
HiSeq 100bp x2 = 50Gb/lane

$$\text{サンプル数} = \frac{\text{スループット (bp)}}{\text{ターゲットサイズ (bp)} \times \text{カバレッジ}}$$

Human = 3.3Gb
Ecoli = 4.6Mb
Exome Capture = 37Mb

どれだけリードを
重ねるか？

問 MiSeqを利用して、150bp x2で、平均50カバレッジとなるよう、
大腸菌E.coliのゲノムを解読した場合、何サンプル解読できるか？

答 $\frac{4,500,000,000}{4,600,000 \times 50} = 19$

次世代シーケンサーデータ解析の必須用語

シングルリード/
ペアエンド リード

スループット

リード長

リード数

カバレッジ

マッピング/
アライメント

バリエーション
コール

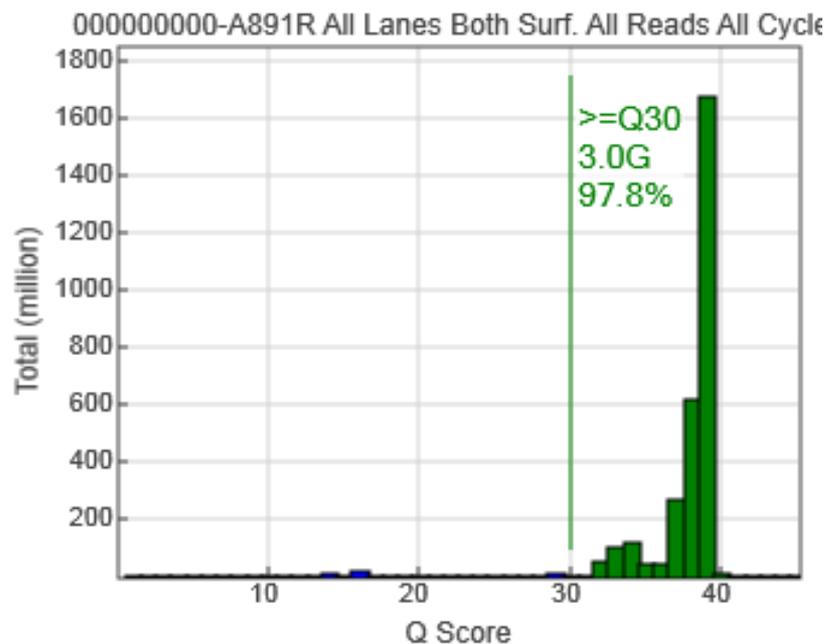
Qスコア

Qスコア

- ▶ Qスコア (Quality Score)
- ▶ = Phred scale
- ▶ 読み取られた塩基が間違いである確率

Phred Score	% Error	Error の確率
Q10	10%	1 in 10
Q20	1%	1 in 100
Q30	0.1%	1 in 1,000
Q40	0.01%	1 in 10,000

- ▶ Q30
 - 1ランで得られたリードのQスコアを総計
 - Q30以上の比率を見ることで、ランの成否を判定



NGSイントロダクション

- An Introduction to Next-Generation Sequencing Technology
http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
- Next-Generation Sequencing Systems
<http://www.illumina.com/content/dam/illumina-marketing/documents/products/brochures/ngs-buyers-guide.pdf>