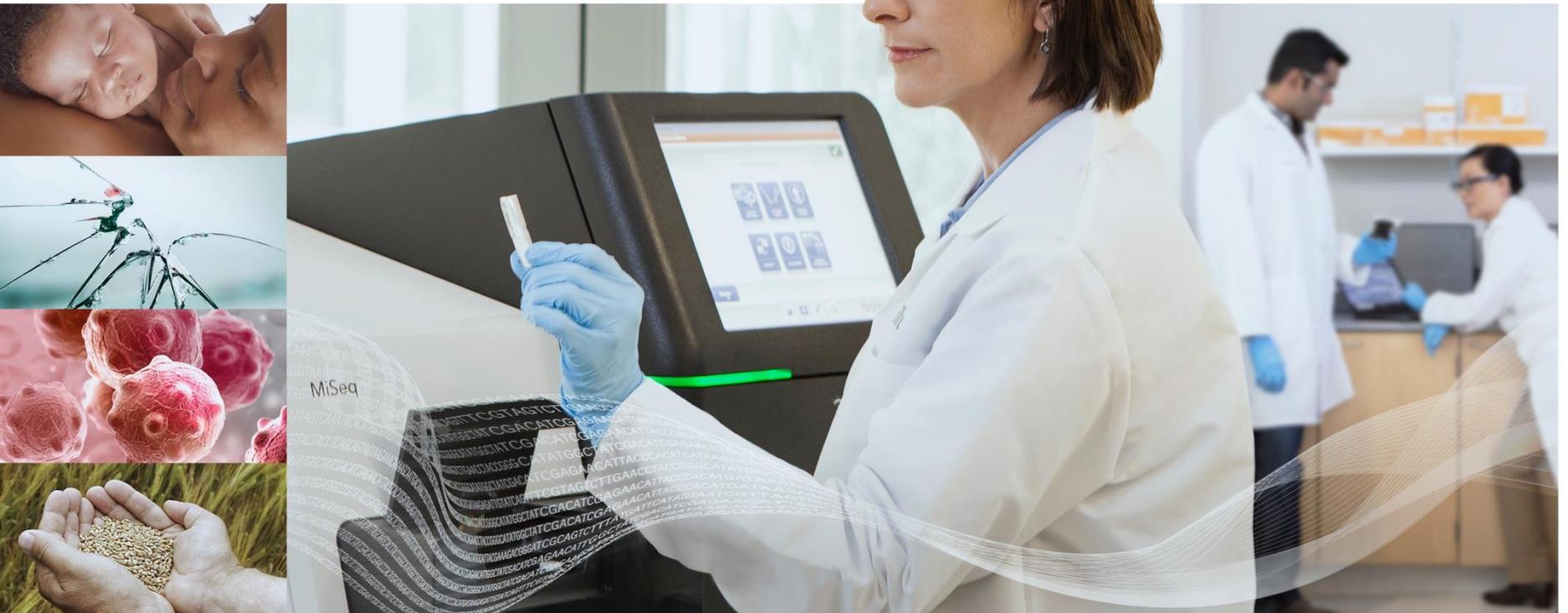


# PCRアンプリコンDNAライブラリーなどのLow Diversityサンプルのシーケンスランを成功させるコツ

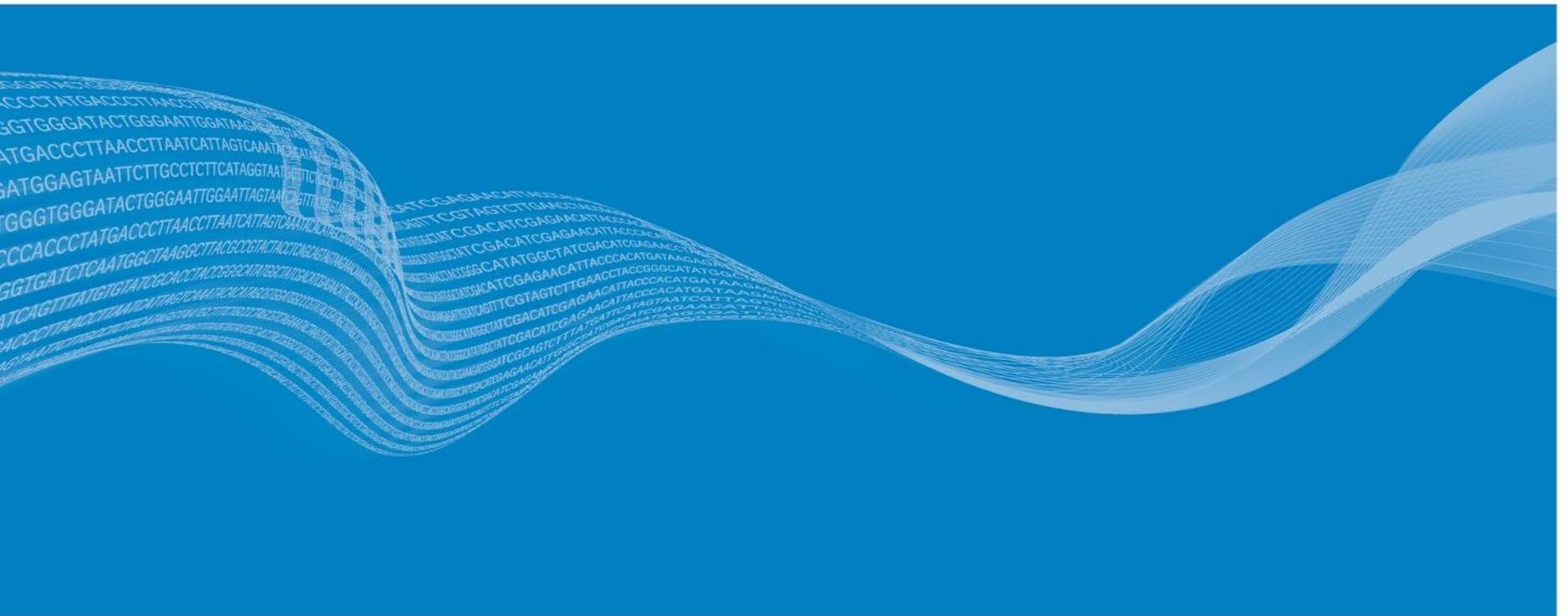
仁田原 翔太  
テクニカルアプリケーションサイエンティスト  
November 20, 2019



# 本日の内容

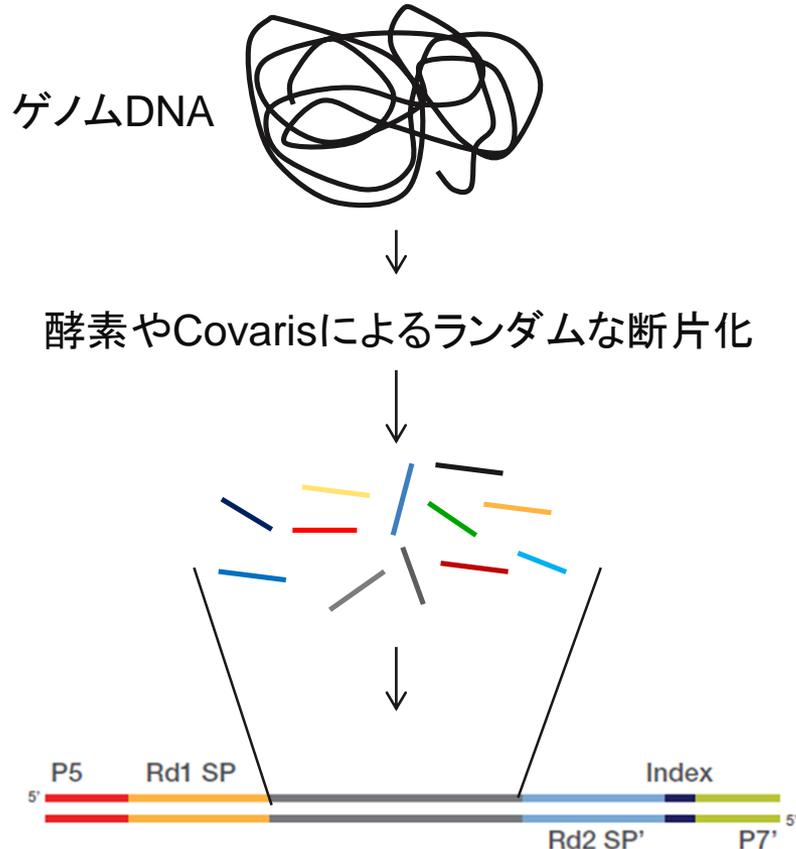
- **Low diversity、塩基多様性とはなにか？**
- **Low diversityサンプルのランはどのような影響を与えるか？**
  - Template Generation
  - Matrix Correction
  - Phasing and Pre-phasing Correction
  - Clusters Passing Filter
- **Low diversityサンプルでのランの所見**
- **Low diversityサンプルのシーケンスを最適化するためには？**

# Low Diversity, 塩基多様性とはなにか？



# 塩基多様性のあるライブラリー

- ホールゲノムシーケンス、エキソームシーケンス、RNA-Seq



- ライブラリーの各インサートはさまざまな配列を持つ
  - 各サイクルでのA, T, G, Cの各塩基の割合が等しく、“塩基の多様性が高い”、また“塩基の偏りが無い”ライブラリーとなる

ATCGTCT

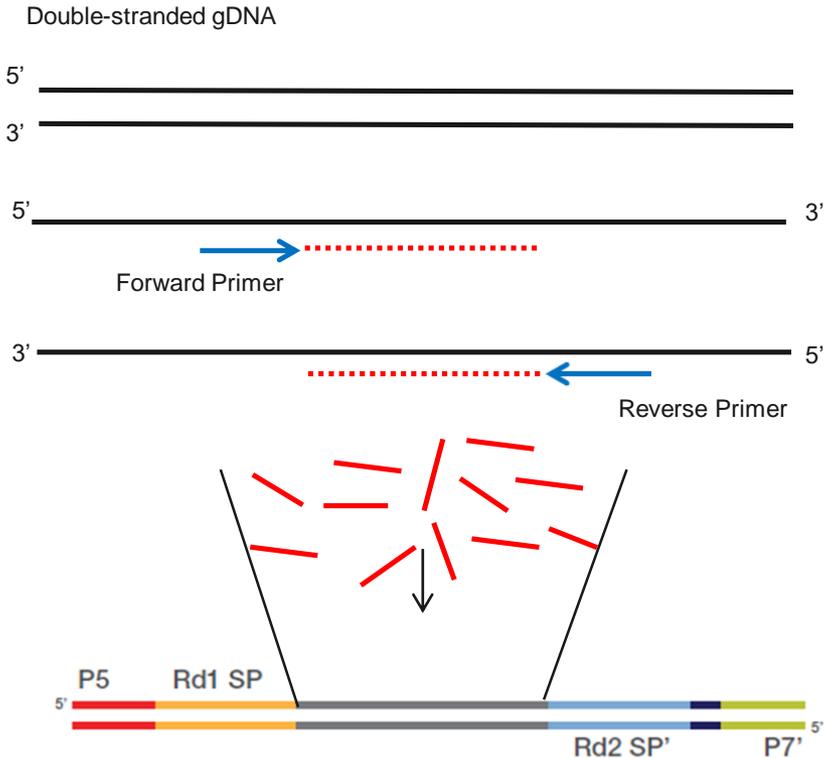
CCGTAAG

GTACGGC

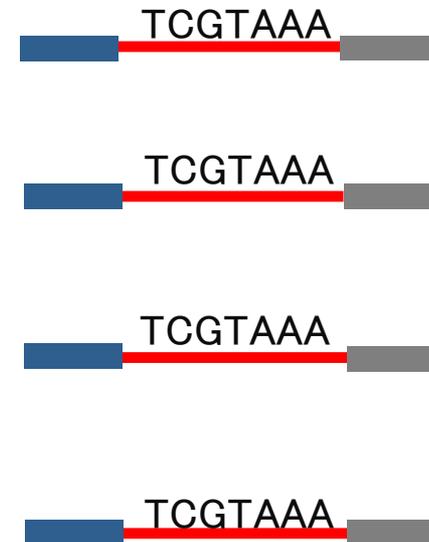
TCGTAAA

# Low Diversityなライブラリー

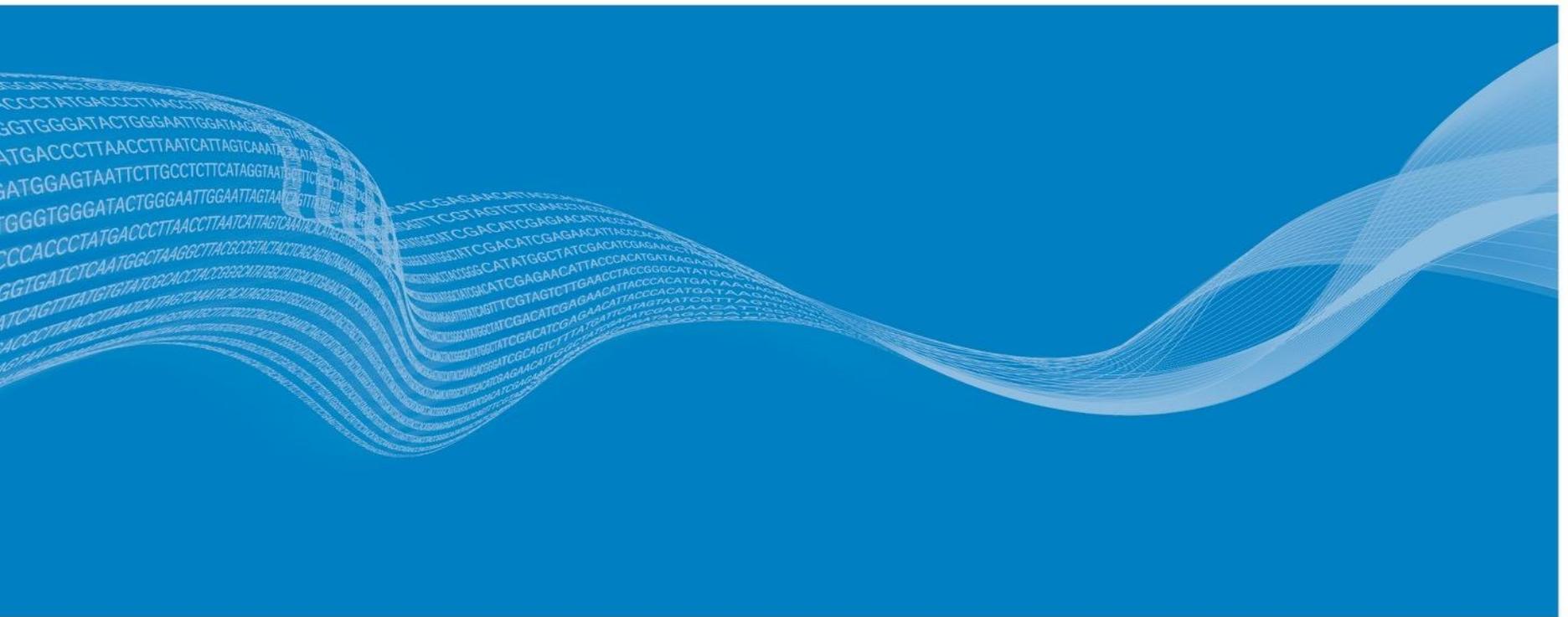
- 例) アンプリコンシーケンス
  - 16S metagenomics



- ライブラリーの各インサート同一の配列を持つ
  - 各サイクルでの特定の塩基の割合が高く、“塩基の多様性が低い”



# Low diversityサンプルのランはどのような影響を与えるか？



# Low diversityサンプルのランはどのような影響を与えるか？

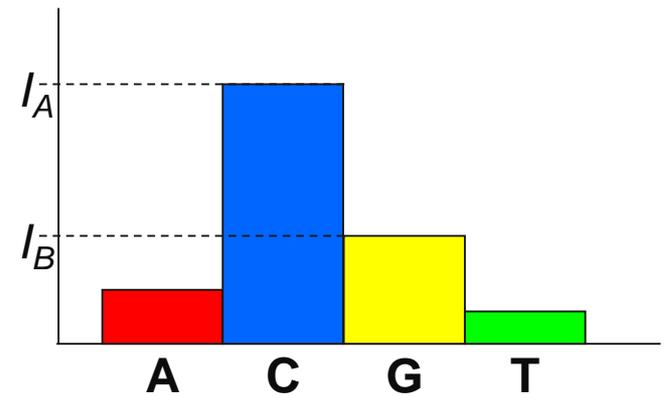
- Chastity Filter—クラスターの純度
- Template Generation—clusterの位置決定
- Matrix Calculation—どの塩基がイメージングされているか？
- Phasing Correction—SBS反応の正確さ

# Low diversityサンプルのランはどのような影響を与えるか？

- **Chastity Filter**—クラスターの純度
- Template Generation—clusterの位置決定
- Matrix Calculation—どの塩基がイメージングされているか？
- Phasing Correction—SBS反応の正確さ

# クラスターのChastity Filter

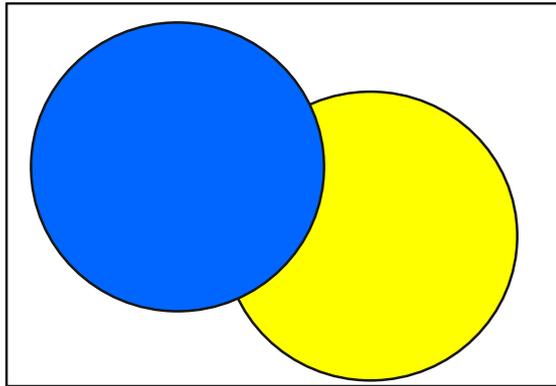
- シーケンサーはシグナルの純度によってクラスターをフィルタリングしている
- 純度の高いクラスターは“chastity filter”を通過する
- その計算に使われるのがChastityの計算式
- 最初の25サイクルで0.6以下となったサイクルが2回以上ある場合、そのクラスターはchastity filterを通過しない



$$C = \frac{I_A}{I_A + I_B}$$

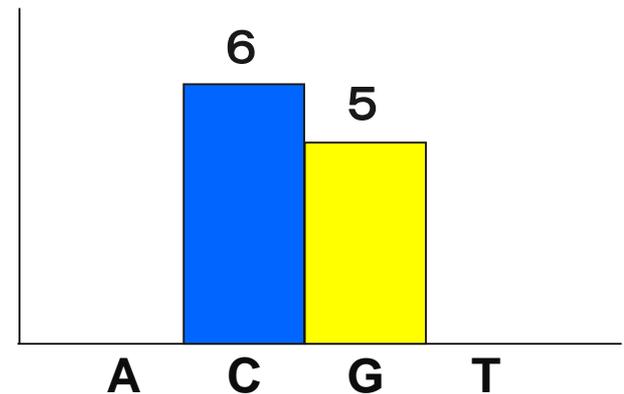
Chastityの計算式

# クラスターのChastity Filter



- 上記のように隣り合ったクラスターの場合、Chastityが0.6以下となり、pass filterを通過しない
- クラスターが正しく認識されない、密度が高い場合、重なりが大きくなる
  - データ量の低下につながる

Lane	Tiles	Density (K/mm <sup>2</sup> )	Cluster PF (%)	Legacy Phasing/Prephasing Rate
1	38	610 ± 17	88.83 ± 1.81	0.000 / 0.000



$$6 / (6 + 5) = 0.54$$

Chastityの 計算式

# Low diversityサンプルのランはどのような影響を与えるか？

- Chastity Filter—クラスターの純度
- **Template Generation—clusterの位置決定**
- Matrix Calculation—どの塩基がイメージングされているか？
- Phasing Correction—SBS反応の正確さ

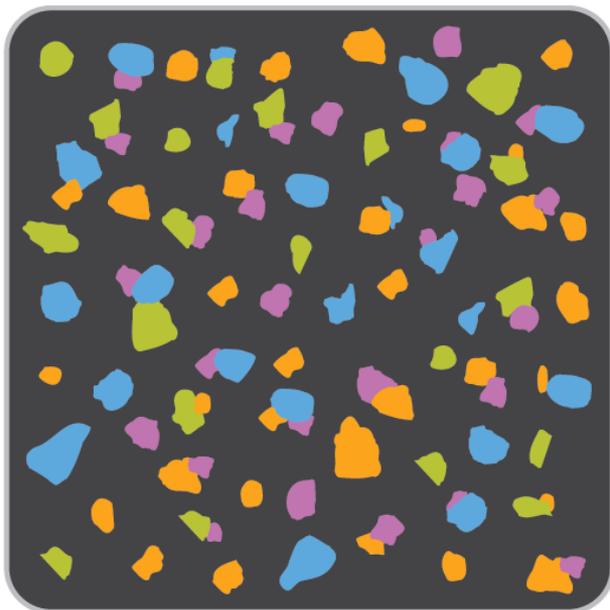
# Template Generation – clusterの位置決定

## ランダムフローセルとパターンフローセルの違い

### Template Generation

HiSeq™ 2500, NextSeq™ 500/550, MiniSeq™,  
MiSeq™

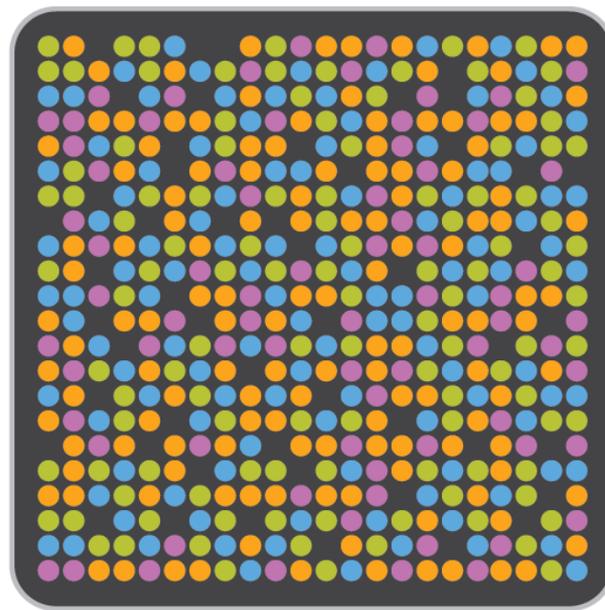
すべてのクラスターの位置を決め、  
マップを作成する



### Rigid Registration

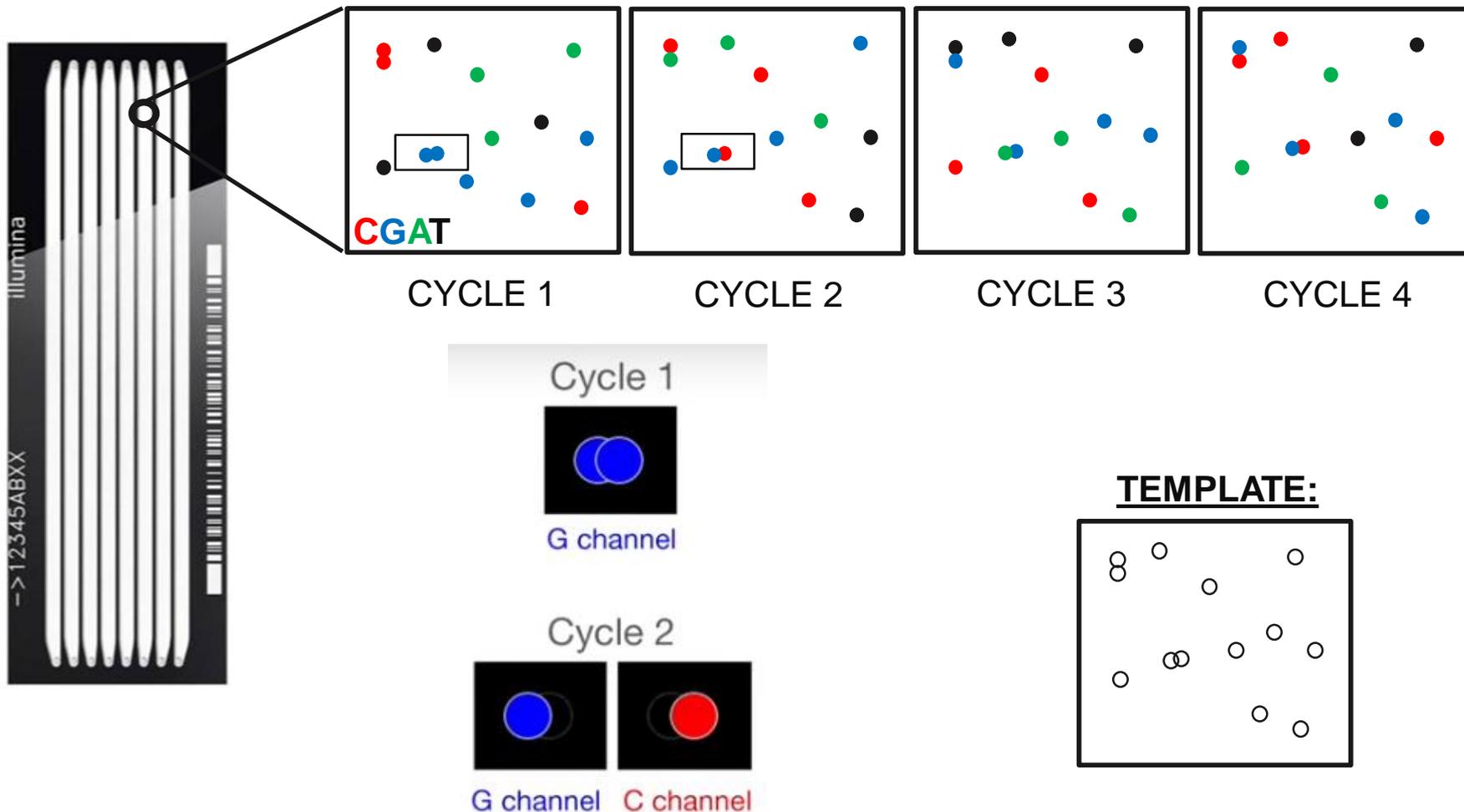
iSeq™ 100, HiSeq™ 3000/4000,  
HiSeq™ X, NovaSeq™ 6000

あらかじめクラスターが出来うる場所  
がわかっている



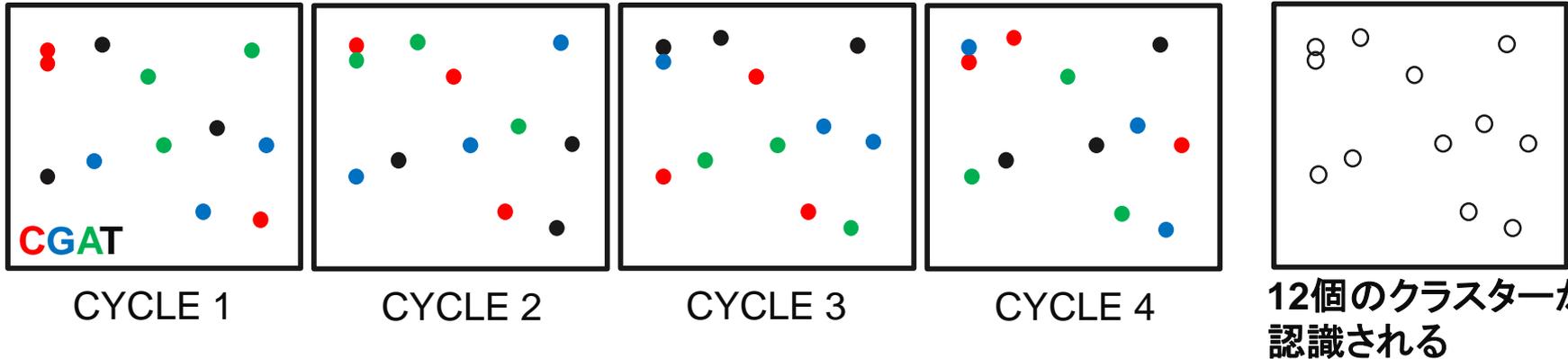
# ランダムフローセルのTemplate Generation

シーケンスの最初の4-7サイクルで得られた画像をもとにクラスターの認識を行なう。(サイクル数はシーケンサーの機種による)



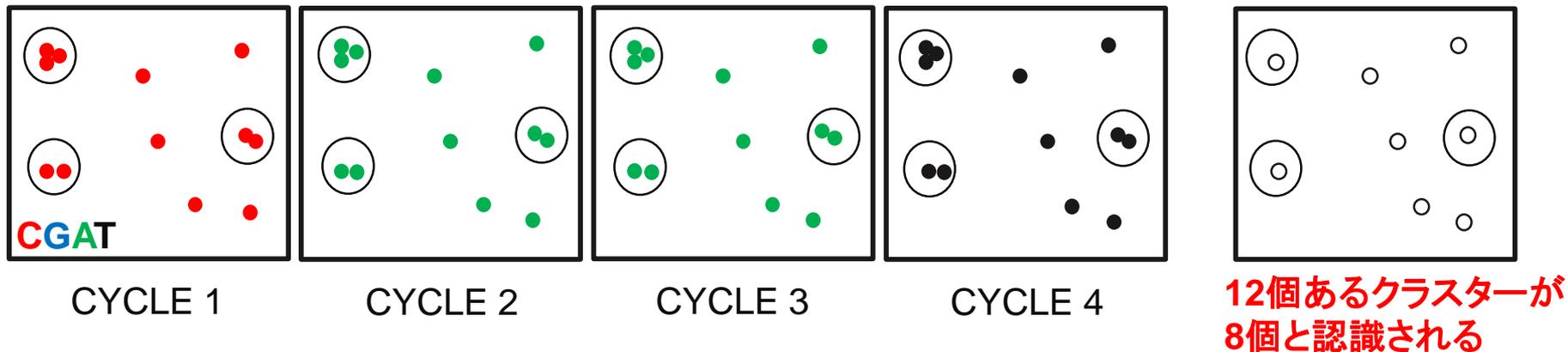
# Low Diversity サンプルの Template Generation

## High Diversity Samples



それぞれのクラスターが異なる配列を持つので、サイクルごとに異なるシグナルが得られる。  
隣り合ったクラスターを区別することができる。

## Low Diversity Samples



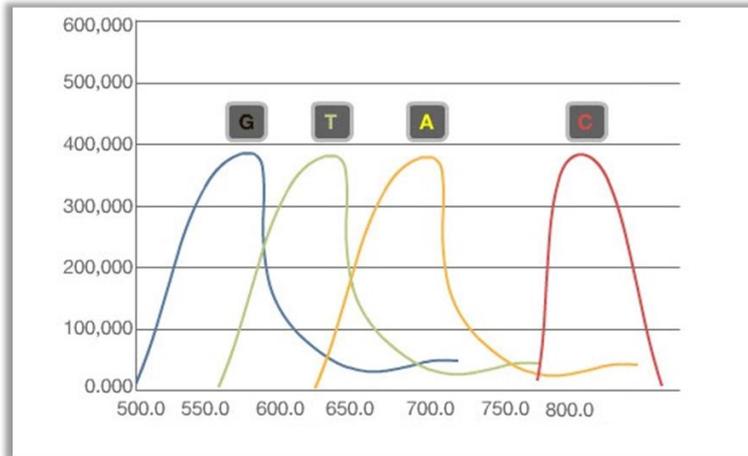
それぞれのクラスターが同じ配列を持つ(異なるシグナルが得られない)ので、  
隣り合った2つのクラスターを1つと認識してしまう

# Low diversityサンプルのランはどのような影響を与えるか？

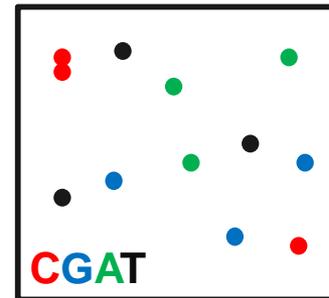
- Chastity Filter—クラスターの純度
- Template Generation—clusterの位置決定
- **Matrix Calculation—どの塩基がイメージングされているか？**
- Phasing Correction—SBS反応の正確さ

# Matrix Calculation for Low Diversity Samples

- 蛍光色素の読み取りの際、波長の重なりがあるため、それを考慮した補正 (color matrixの作成)をする必要がある

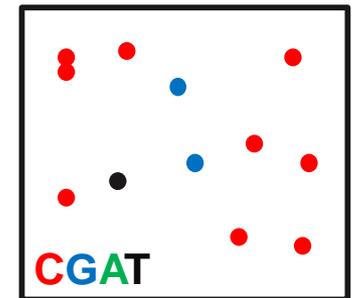


High Diversity



High Diversityサンプルでは参照するシグナルが十分にあり、color matrixを作成できる

Low Diversity



Low Diversityサンプルでは 参照するシグナルが乏しく、正しく補正できない

- 適切にmatrixが計算できないと、正しく補正ができずクオリティも低下する
- matrixを計算後のサイクルで塩基の組成に変化があると、それに合わない状況 となり、クオリティが低下しやすい

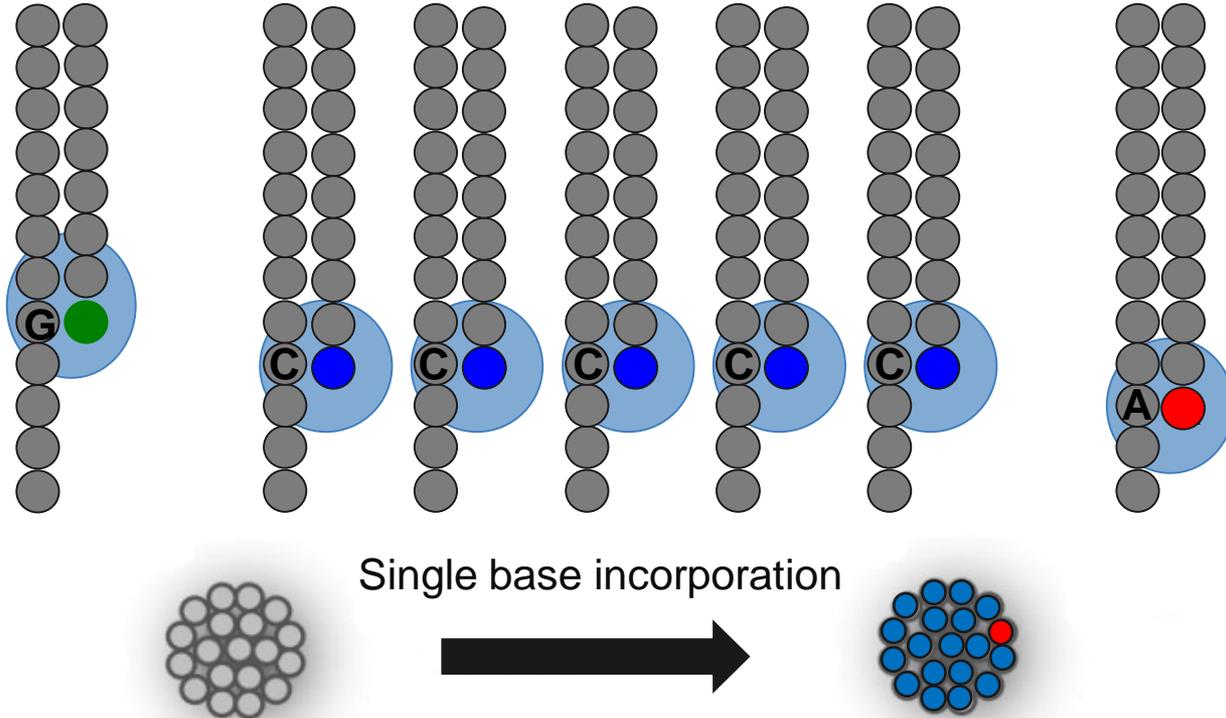
# Low diversityサンプルのランはどのような影響を与えるか？

- Chastity Filter—クラスターの純度
- Template Generation—clusterの位置決定
- Matrix Calculation—どの塩基がイメージングされているか？
- **Phasing Correction—SBS反応の正確さ**

# Phasing および Pre-phasing の計算

Phasing

Pre-phasing



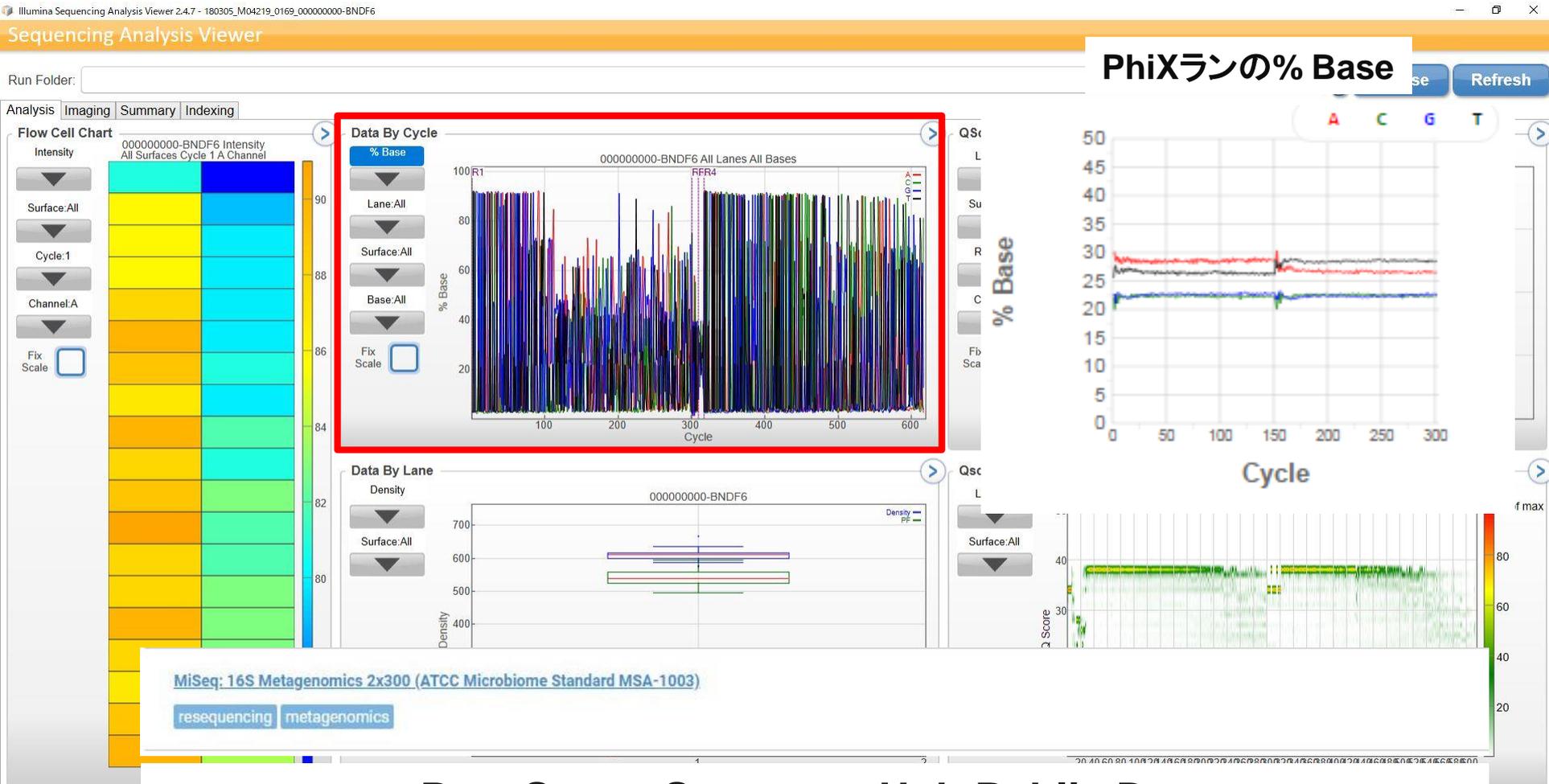
典型的なランでは、phasingおよびpre-phasingは0.5%以下となることが多い  
これらの値が高いことはサイクルごとでのSBS反応(取り込み/切り離し)の失敗を反映している

# Low diversityサンプルでのランの所見



# Low diversityサンプルでのランの所見, %Base

- Sequencing Analysis Viewer (SAV)でランデータを確認する

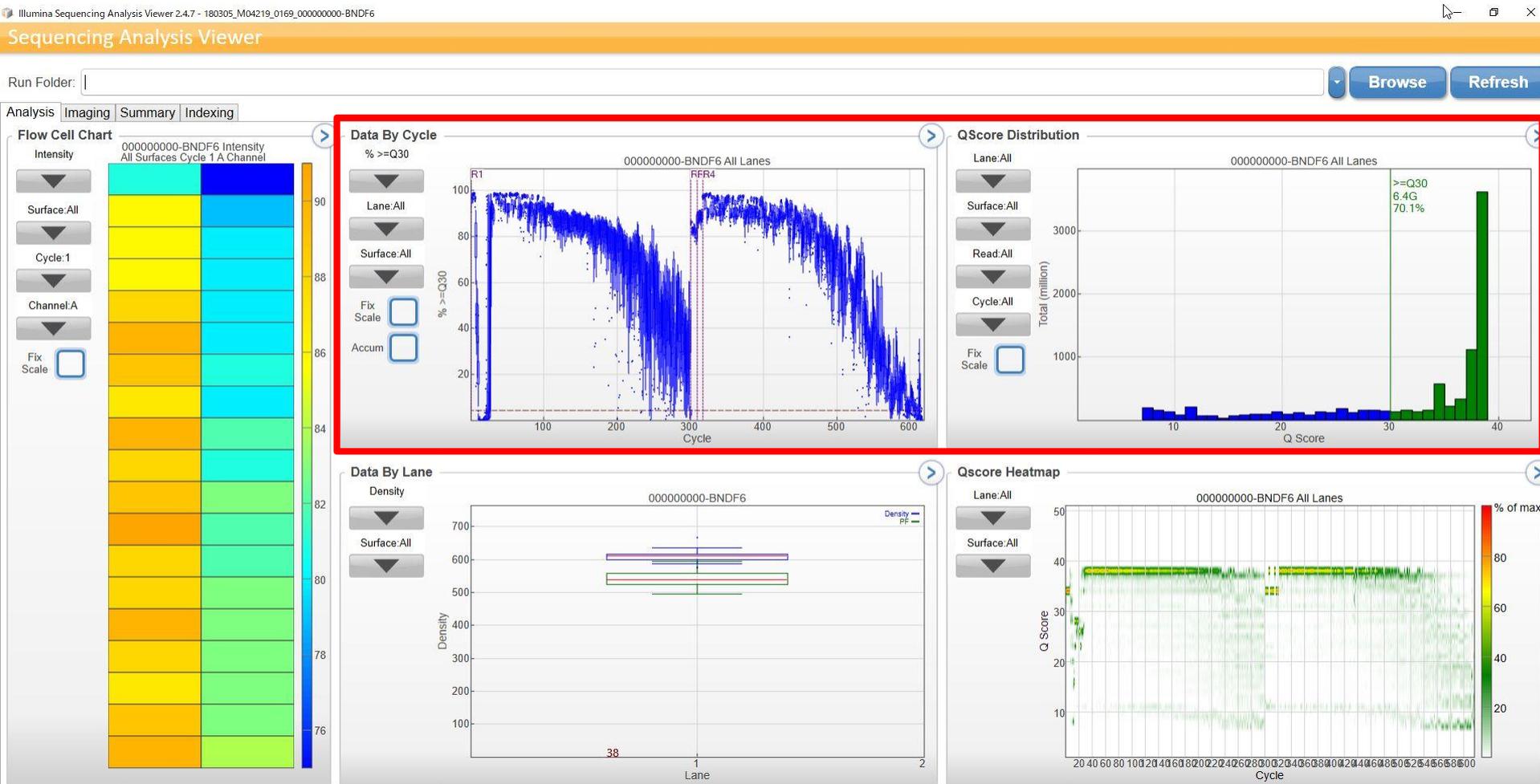


BaseSpace Sequence Hub Public Data

<https://basespace.illumina.com/datacentral>

# Low diversityサンプルでのランの所見, Q30%

- Sequencing Analysis Viewer (SAV)でランデータを確認する



# Low diversityサンプルでのランの所見

- Sequencing Analysis Viewer (SAV)でランデータを確認する

© Illumina Sequencing Analysis Viewer 2.4.7 - 180305\_M04219\_0169\_000000000-BNDF6

## Sequencing Analysis Viewer

Run Folder:

Analysis Imaging Summary Indexing

### Cycle Status

Extracted: 616    Called: 616    Scored: 616    Error Rated: 615

### Run Summary

Level	Yield Total (G)	Projected Total Yield (G)	Aligned (%)	Error Rate (%)	Intensity Cycle 1	% >= Q30
Read 1	4.46	4.46	9.36	2.03	85	72.41
Read 2 (I)	0.10	0.10	0.00	NaN	120	83.47
Read 3 (I)	0.10	0.10	0.00	NaN	533	88.42
Read 4	4.46	4.46	9.50	2.41	59	67.38
Non-Indexed Total	8.92	8.92	9.43	2.22	72	69.89
Total	9.13	9.13	9.43	2.22	199	70.26

### Read 1

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Legacy Phasing/Prephasing Rate	Phasing slope/offset	Prephasing slope/offset	Cluster Count Raw (M)	Cluster Count PF (M)	% >= Q30	Yield (G)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1
1	38	610 ± 17	88.83 ± 1.81	0.000 / 0.000	NaN / NaN	NaN / NaN	16.80	14.92	72.41	4.46	299	9.36 ± 0.36	2.03 ± 0.10	0.69 ± 0.12	0.48 ± 0.06	0.46 ± 0.07	85 ± 4

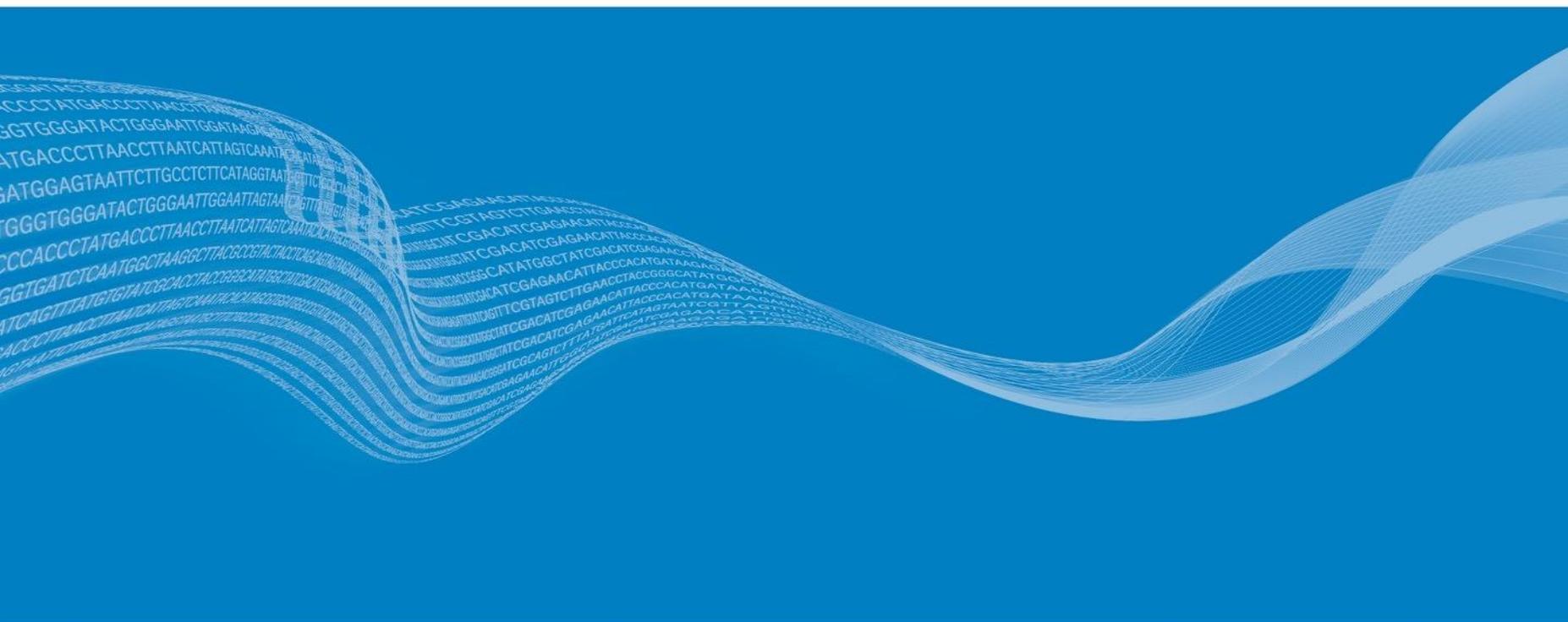
Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Legacy Phasing/Prephasing Rate	% >= Q30	Yield (G)	Cycles Err Rated	Aligned (%)	Error Rate (%)
1	38	610 ± 17	88.83 ± 1.81	0.000 / 0.000	72.41	4.46	299	9.36 ± 0.36	2.03 ± 0.10

クオリティスコア(Q30%)やCluster PF%、Yield(G)の低下にすることがある

SAV (Sequencing Analysis Viewer) でランの評価と改善を行う

[https://jp.illumina.com/events/webinar/2017/webinar\\_170614\\_j.html](https://jp.illumina.com/events/webinar/2017/webinar_170614_j.html)

# Low diversityサンプルのシーケンスを最適化するためには？

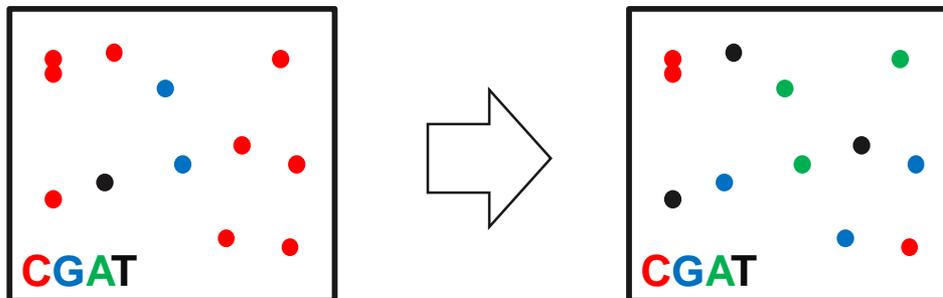


# low diversityのシーケンスを最適化するための方法

1. PhiX のSpike-in (添加)
2. クラスター密度を下げる
3. ライブラリーのデザインを考慮する
4. ライブラリーのQuality Control

# ポイント 1: PhiX のSpike-in (添加)

- ▶ PhiX はバクテリオファージ由来のDNAライブラリーのコントロール
  - ゲノムサイズが小さい (5kbase) = アライメントが速く行える
  - 塩基の多様性のあるライブラリー, GC含量 45%
  - 配列のすでに決まったゲノム
- ▶ PhiX Control Kit V3 として販売 (cat no. FC-110-3001)
  - インデックス配列は持たない
  - ライブラリーサイズ ~500 bp
  - 平均インサートサイズ ~ 375 bp



# ポイント 1: PhiX のSpike-in (添加)

- 塩基の多様性が低い、あるいは偏りのあるライブラリーの補正に推奨されるPhiXの添加量 (ライブラリーによってはより多く必要とする場合もございます)

Platform	PhiX Aligned (%)†
iSeq 100	minimum 5%
MiniSeq	10-50%*
MiSeq (MCS 2.2 or higher)	minimum 5%
NextSeq	10-50%*
HiSeq 2500 (HCS 2.2.38 or higher)	minimum 10%
HiSeq 3000/4000 (HCS 3.3.76 or lower)	10-50%*
HiSeq 3000/4000 (HCS 3.4.0 or higher)	5-20%*
NovaSeq	minimum 10%

How much PhiX spike-in is recommended when sequencing low diversity libraries on Illumina platforms?  
[https://jp.support.illumina.com/content/illumina-support/apac/ja\\_JP/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html](https://jp.support.illumina.com/content/illumina-support/apac/ja_JP/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html)

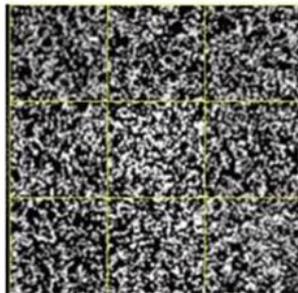
# ポイント 2: クラスター密度を下げる

- バランスのとれたライブラリーでの推奨クラスター密度は下記の通り

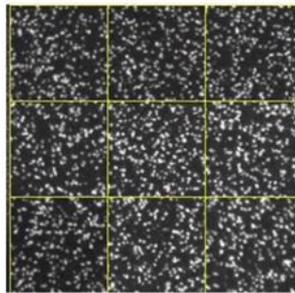
Table 1: Optimal Raw Densities for Illumina sequencing systems

	MiniSeq™	MiSeq		NextSeq®	HiSeq 2500 Rapid Run (RR)	HiSeq 2500 High Output (HO)	
Versions	High and Mid Output	v2	v3	v2 High and Mid Output	v1 and v2	v3	v4
Raw Density (K/mm <sup>2</sup> )	170–220	1000–1200	1200–1400	170–220	850–1000	750–850	950–1050

- 塩基の多様性が低いサンプルでは、上記**推奨よりも低め(6-8割程度)**に設定することをお奨めしている
- ラン結果(クオリティなど)をもとに、クラスター密度は調整する



1200 k/mm<sup>2</sup>



850 k/mm<sup>2</sup>

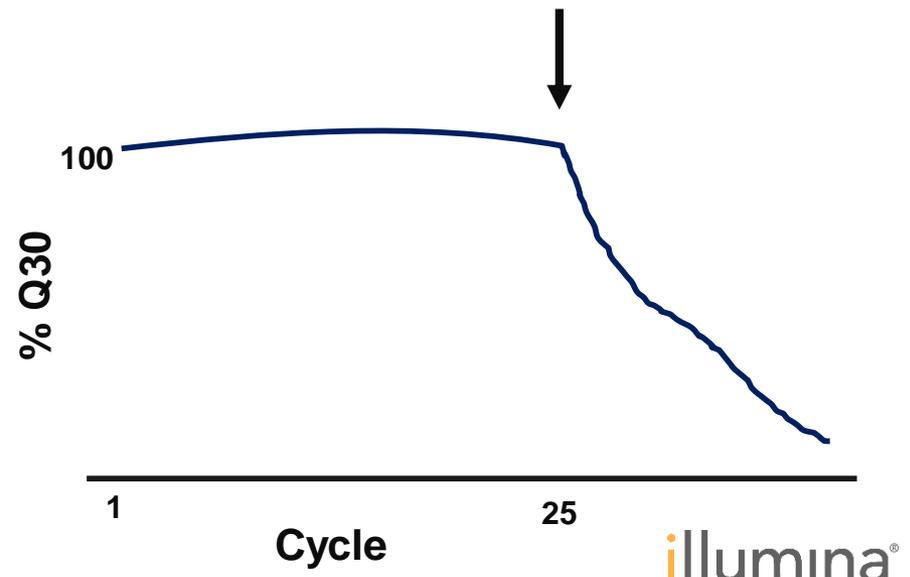
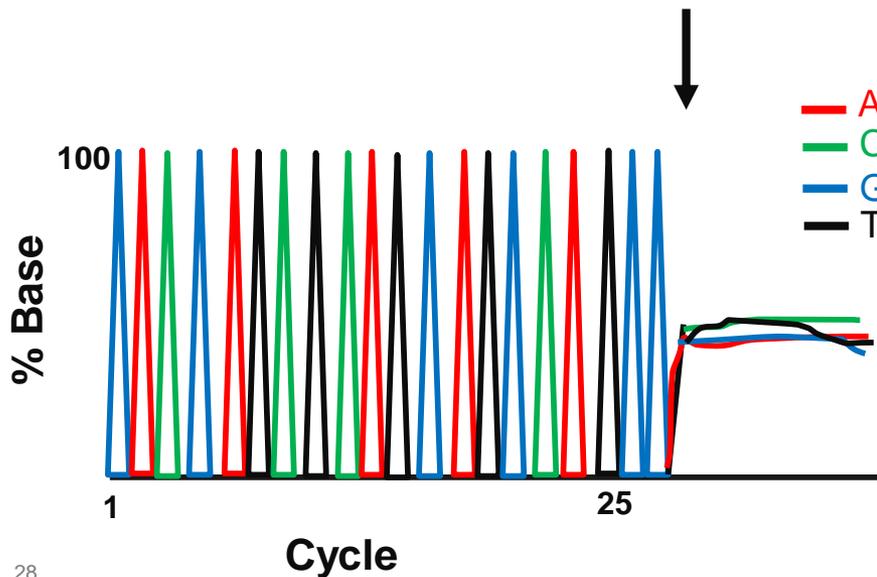
- クラスター同士が重なるリスクが低くなる
- S/N比の改善

Cluster Optimization Overview Guide

[https://support.illumina.com/content/dam/illumina-support/documents/documentation/system\\_documentation/cluster-optimization-overview-guide-1000000071511-00.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cluster-optimization-overview-guide-1000000071511-00.pdf)

# ポイント 3: ライブラリーのデザイン

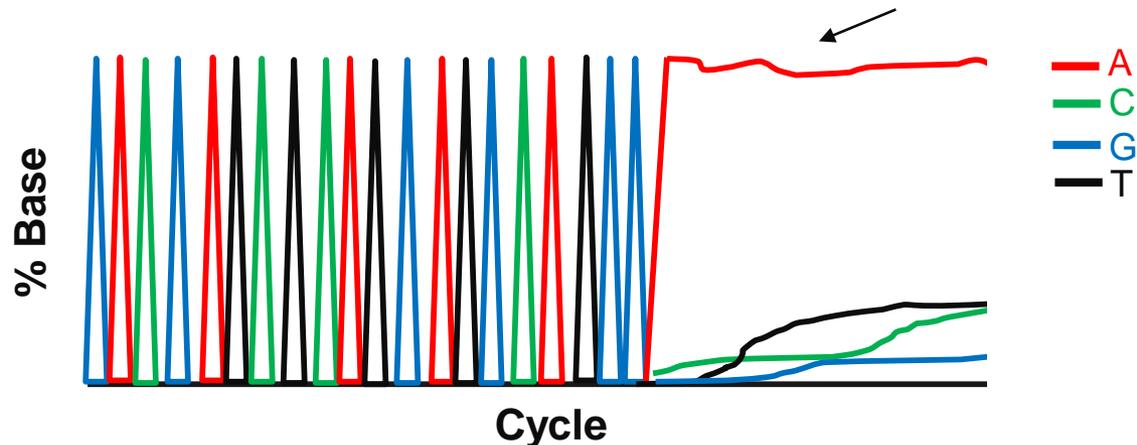
- 読み取りの最初25サイクルに塩基の多様性を増やす
- シーケンスするアンプリコンの種類を増やす (> 16 amplicons)
  - 塩基の多様性を上げるようにする
- リードの途中から塩基の多様性が大きく変わるようなデザインは避ける
  - Color matrixによる補正が変わり、クオリティが落ちる原因となる



# ポイント 4: ライブラリーのQuality Control

- ライブラリーのクオリティをBioanalyzerなどで確認する
  - アダプターダイマーやプライマーダイマーの混入を避ける
  - リード長より短いライブラリーは避ける
  - 短いライブラリーはアダプターまで読み切ってしまう、クオリティが低下する

ライブラリーを読み切ると、  
MiSeq™, HiSeq™ 2500ではAの割合が高くなる。  
iSeq™ 100, NextSeq™ 500/550, MiniSeq™, NovaSeq™ 6000  
ではGの割合が高くなる



# 本日のまとめ

- **Low Diversityサンプルとは、配列に相同性がある、もしくは塩基の割合に偏りがあるサンプルのこと**
  - データの出力やクオリティスコアに影響することがある
- **塩基の多様性はSAVの結果をみることで確認できる**
- **Low Diversityサンプルのランを成功させるには**
  - PhiXのスパイクインを実施する (例. 16Sでは20-30%程度)
  - クラスタ密度を低くする (推奨の6-8割程度)
  - アプリケーション、シーケンサーによっては条件検討が必要となる

# 參考資料

- **Technical Notes:**

[Low-Diversity Sequencing on the Illumina MiSeq Platform](#)

[Low-Diversity Sequencing on the Illumina HiSeq Platform](#)

[Cluster Optimization Overview Guide](#)

[Using a PhiX Control for HiSeq Sequencing Runs](#)

- **Application Notes:**

[Whole-Genome Bisulfite Sequencing on the HiSeq 3000/HiSeq 4000 Systems](#)

[16S Metagenomics Studies with the MiSeq System](#)

[16S metagenomics sequencing with the iSeq™ 100 System](#)

- **Support Bulletins:**

[Best Practices for Low-Diversity Sequencing on the NextSeq and MiniSeq Systems](#)

[How much PhiX spike-in is recommended when sequencing low diversity libraries on Illumina platforms?](#)

[Cluster density guidelines for Illumina sequencing platforms](#)

# 參考資料

- **Support Bulletins:**

[Library quantification and quality control quick reference guide](#)

[Cluster Density Specifications for Illumina Sequencing Platforms](#)

[Considerations when migrating non-Illumina libraries between sequencing platforms](#)

[Cluster density considerations when migrating Illumina libraries between sequencing platforms](#)

[How to achieve more consistent cluster density on Illumina sequencing platforms](#)

[Nextera Library Validation and Cluster Density Optimization](#) – Lists some guidelines for loading concentrations based on average size of library

- **Additional Programs:**

[Sequencing Analysis Viewer \(SAV\) Guide & Download](#)

ご参加頂き、ありがとうございました。  
セッション終了後のご質問は  
[techsupport@illumina.com](mailto:techsupport@illumina.com)で承ります。

