



# Low Diversityサンプルを解析するための テクニック

小林 孝史  
イルミナ株式会社

テクニカルアプリケーション  
サイエンティスト

塩基の偏りの大きいサンプルとは？



# Low diversityサンプルとは？

1. ライブラリー（クラスター）内の偏り:  
偏った配列がReadの初めのサイクルにあったり、%が高かったりすると解析しにくい  
他のシーケンサーより精度が良い

1つのリードの配列 —

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

2. ライブラリー（クラスター）間の偏り:  
イルミナシーケンサーで解析が難しいパターン、Low diversityサンプルと位置付ける

Low Diversity

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

# なぜLow diversityサンプルが読みにくいのか？

4サイクル目の画像では

High Diversity

AA**C**GAGACGCATTC  
TT**C**TGACGAGTAAC  
ACA**C**AGGCTACTAC  
GCC**A**TACCATGATG

G  
T  
C  
A

それぞれの塩基のシグナルが  
均等に取得される

画像を取得するサイクル 1 2 3 4 5 6 7 8 9.....



Low Diversity

ACA**T**AGGCTACTAC  
ACA**T**AGGCTACTAC  
ACA**T**AGGCTACTAC  
ACA**T**AGGCTACTAC

T  
T  
T  
T

Tのみのシグナルが取得される  
取得シグナルに偏りがある

画像を取得するサイクル 1 2 3 4 5 6 7 8 9.....

# なぜLow diversityサンプルが読みにくいのか？

たとえば、一つ一つのチョコレートを、一つのライブラリーから得られるシグナルとする。



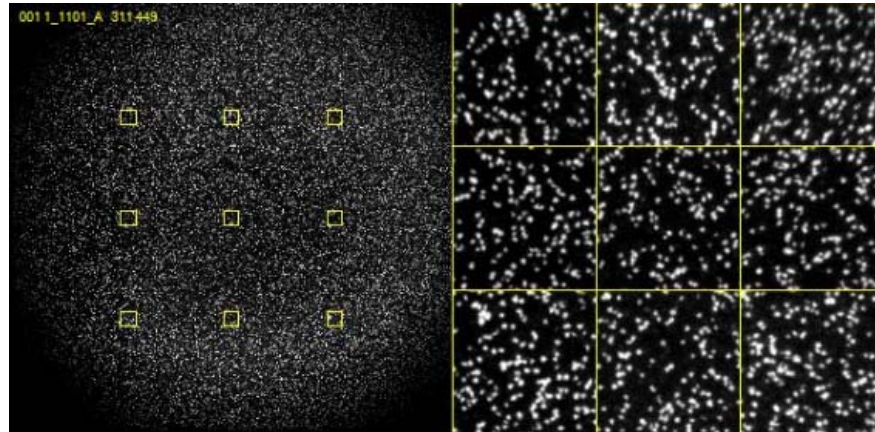
それぞれの塩基のシグナルが均等に取得される  
→隣のシグナルとの境界を判別しやすい。  
それぞれのクラスターの大きさを正常に判別できる



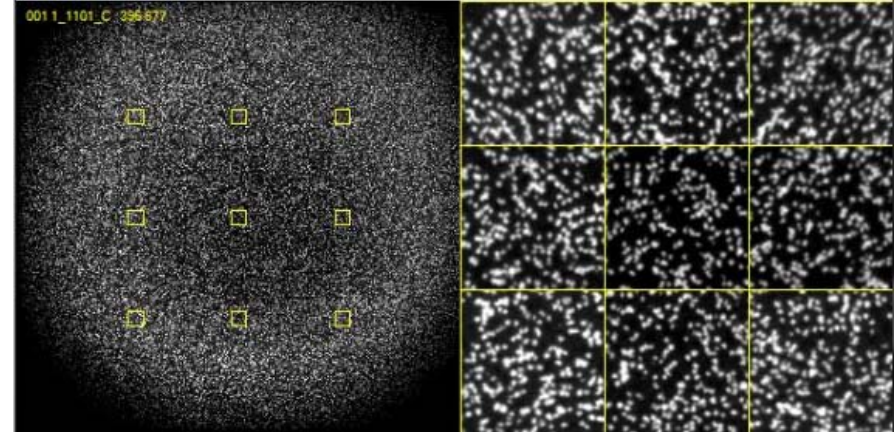
1つの塩基のシグナルしか取得されない  
→隣のシグナルとの区別ができない。  
2つのクラスターを一つのクラスターとして認識してしまう。  
他の塩基の取得(Registration)が適正にできない。

# 具体的な解析データの例: コントロールが難しい!

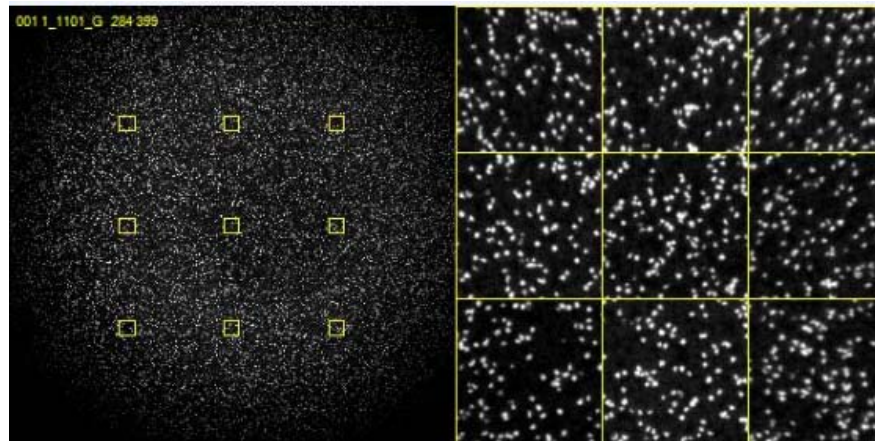
**A**



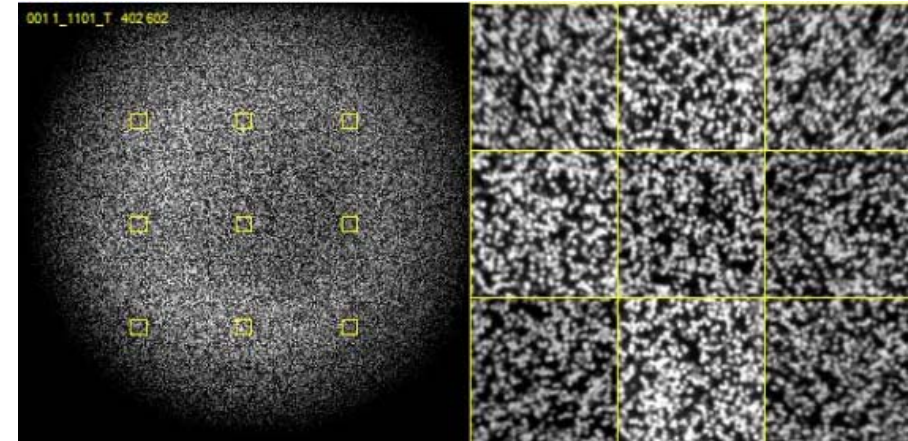
**C**



**G**: 若干少なめなクラスター密度



**T**: 明らかなオーバークラスター



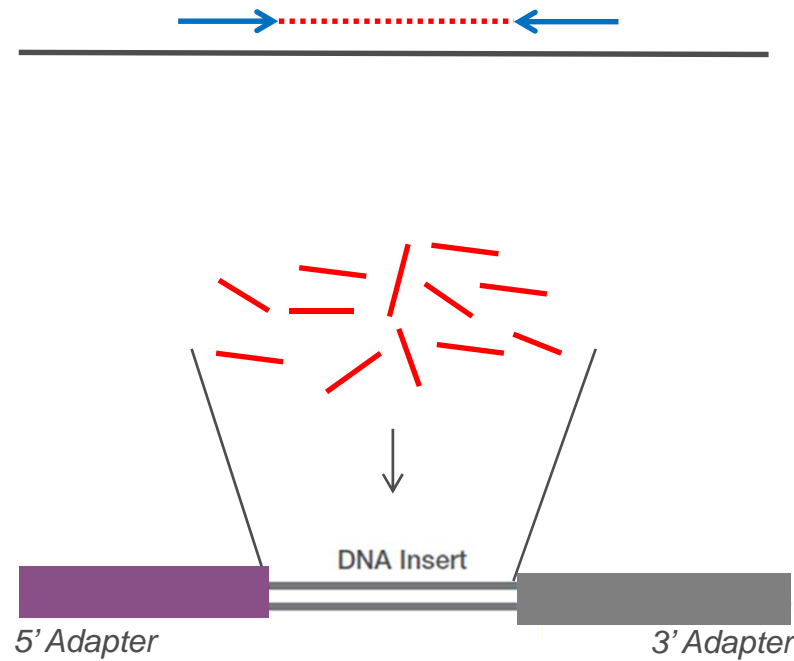


Low diversityサンプルを解析するアプリケーションは？



# Low diversityサンプルを解析するアプリケーションは？

## ▶ ①PCRアンプリコン



- ▶ シーケンス中の各配列は、同一の配列を持つ。
  - 異なる配列はターゲット箇所のみ
  - 多様性の低いサンプル

解析される配列の例

 ACATAGGCTACTAC

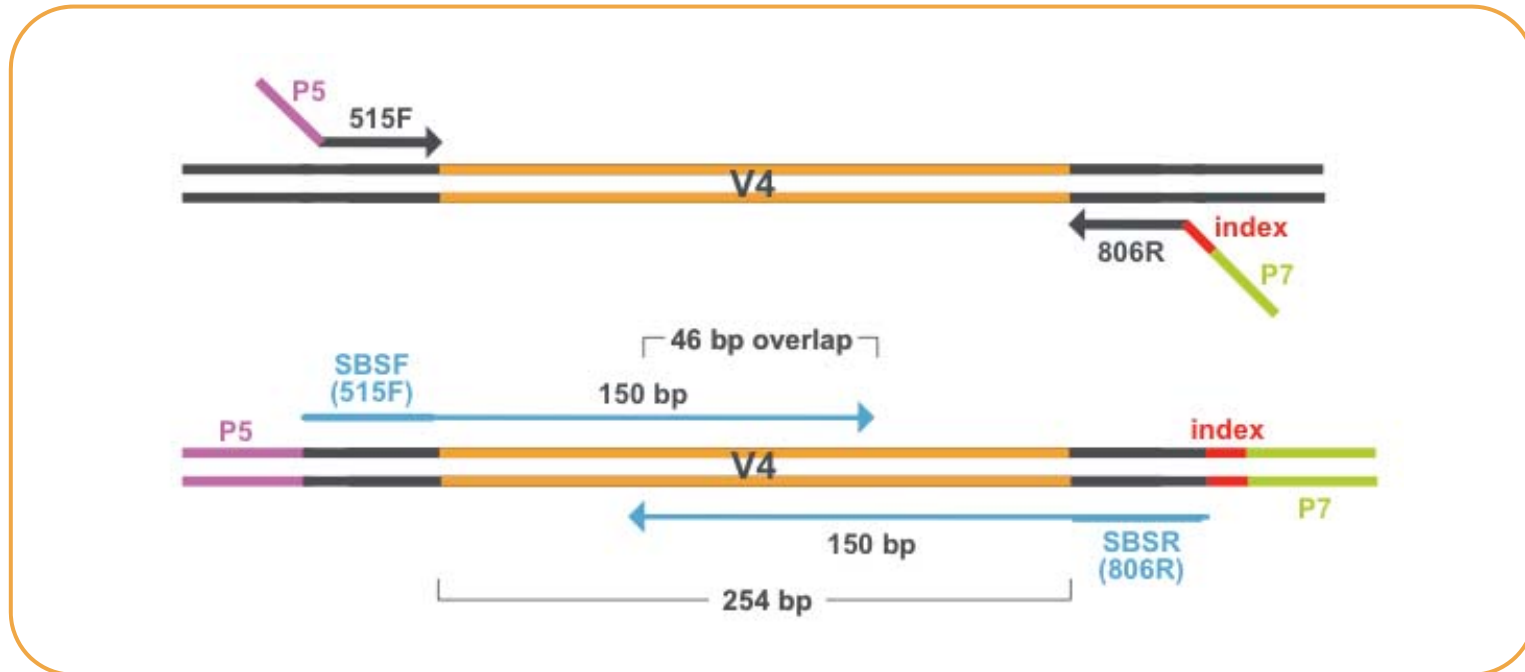
 ACATAGGATACTAC

調べたい変異の割合も小さい

 ACATAGGCTACTAC

 ACATAGGCTACTAC

# 16S rRNAサンプルはLow diversityサンプル



- 環境サンプルなどから16S ribosomal RNAのV4領域をPCRにより増幅し、個体間のマイナーな変異をモニターする。
- ライブラリー間で相動性が非常に高い。⇒Low Diversityである。

# 16S rRNAサンプルはLow diversityサンプル

- イルミナで公式にサポートする16S metagenomeシーケンシングのプロトコールのリリース
- [http://support.illumina.com/downloads/16s\\_metagenomic\\_sequencing\\_library\\_preparation.ilmn](http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.ilmn)

## 16S Metagenomic Sequencing Library Preparation

*Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*

Introduction	2
16S Library Preparation Workflow	5
Amplicon PCR	6
PCR Clean-Up	8
Index PCR	10
PCR Clean-Up 2	13
[Optional] Validate Library	15
Library Quantification, Normalization, and Pooling	16
Library Denaturing and MiSeq Sample Loading	17
MiSeq Reporter Metagenomics Workflow	20
Supporting Information	21

# Low diversityサンプルを解析するアプリケーションは？

## ▶ ②Index Read

例) Nextera XT userguideから抜粋。

イメージを取る際に、赤蛍光のシグナル(C/A)と緑蛍光のシグナル(G/T)の2つのシグナルが同時に必要。

These represent only some of the acceptable combinations. Alternatively, please check the real sequences of each index in the table above to make sure each base position will have signal in both color channels for the index read:

Good				Bad			
Index 1		Index 2		Index 1		Index 2	
705	GGACTCCT	503	TATCCTCT	705	GGACTCCT	502	CTCTCTAT
706	TAGGCATG	503	TATCCTCT	706	TAGGCATG	502	CTCTCTAT
701	TAAGCGA	504	AGAGTAGA	701	TAAGCGA	503	TATCCTCT
702	CGTACTAG	504	AGAGTAGA	702	CGTACTAG	503	TATCCTCT
	√√√√√√√√		√√√√√√√√		√√√√√√√√		√√√√xxxx

√=signal in both color  
x=signal missing in one color channel

# Index Readは塩基の偏りが大きくなりやすいので注意

1) Indexの数が多ければ多いほどIndex ReadのQ scoreは安定する。  
→可能であればサンプルの数を増やしてIndex ReadのQ scoreを安定させる

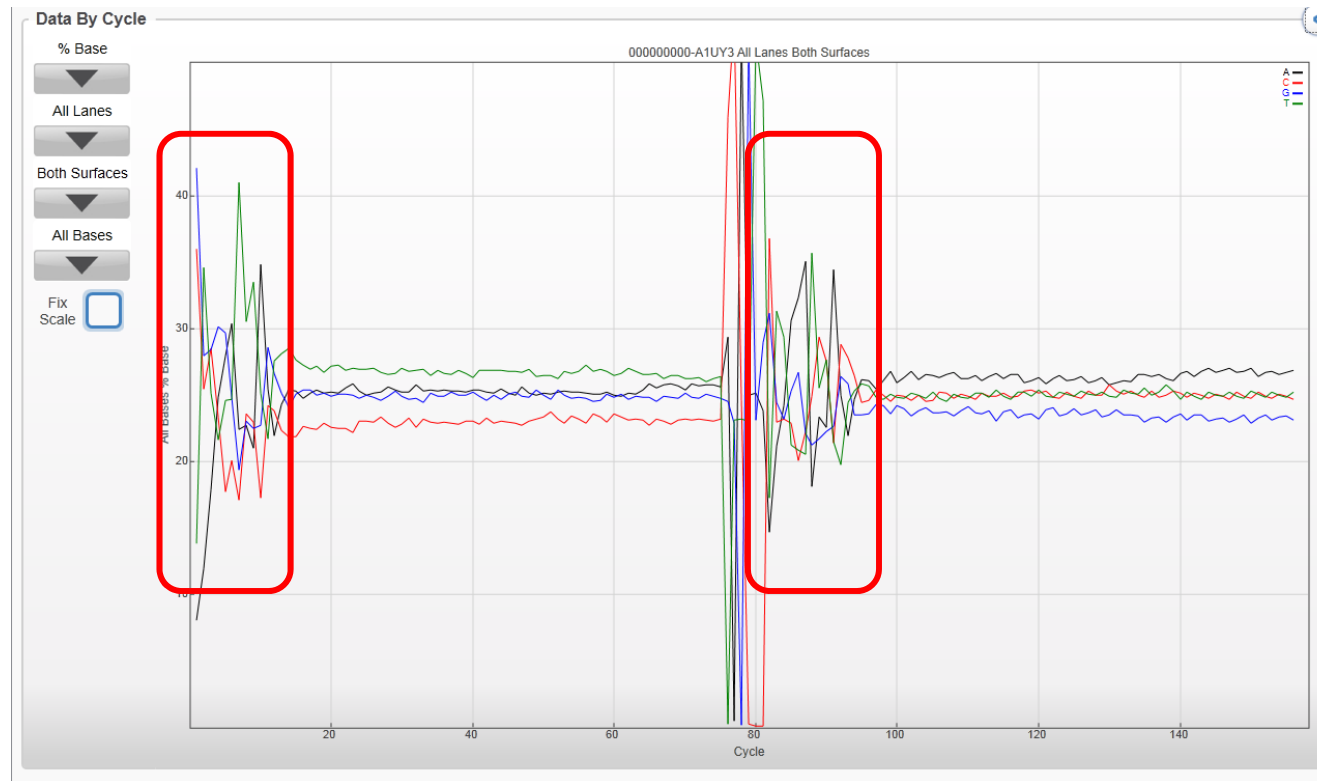
2) 少ないIndex数の場合は、ユーザーガイドのガイドライン（下記）に従うようにする。

**Table 5** Libraries Pooled: 6 or fewer; Sequencing Workflow: Single Index

Plex	Index 1 (i7) Selection	Index 2 (i5) Selection
1-plex (no pooling)	Any Index 1 adapter	Any Index 2 adapter
2-plex	<ul style="list-style-type: none"><li>• [option 1] N702 and N701</li><li>• [option 2] N702 and N704</li></ul>	
3-plex	<ul style="list-style-type: none"><li>• [option 1] N701, N702, and N704</li><li>• [option 2] N703, N705, and N706</li></ul>	
4- or 5-plex	<ul style="list-style-type: none"><li>• [option 1] N701, N702, N704, and any other Index 1 adapter</li><li>• [option 2] N703, N705, N706, and any other Index 1 adapter</li></ul>	
6-plex	N701, N702, N703, N704, N705, and N706	

# RNAサンプルは読み出し時に若干塩基配列が偏る

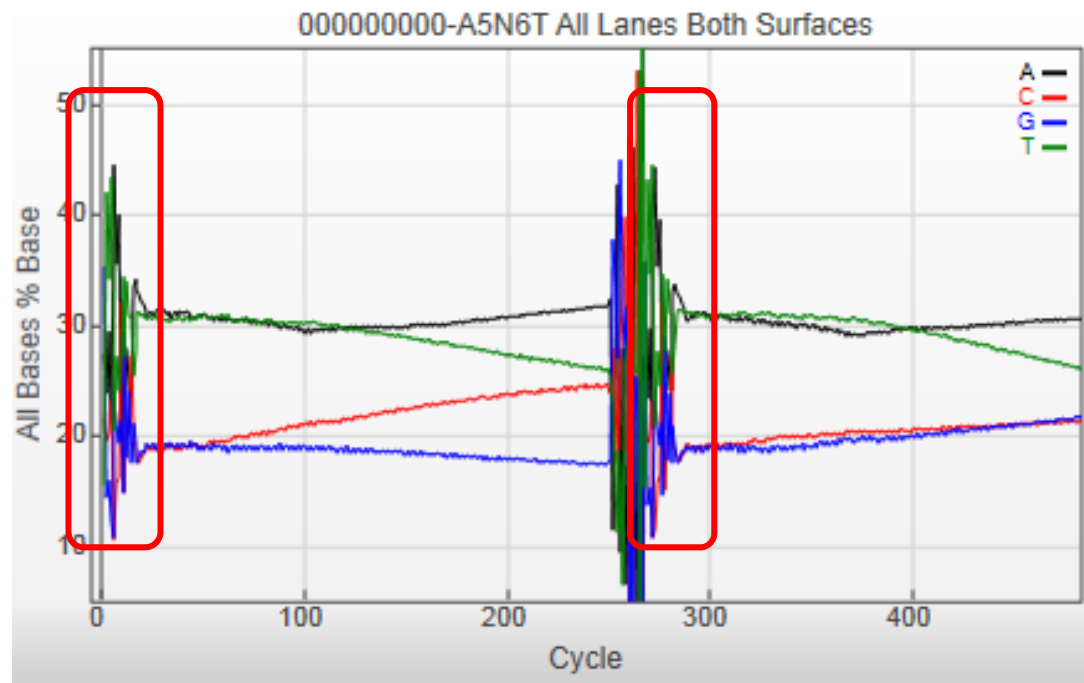
## ▶ ③RNAサンプル



- 主にランダムプライマーでの増幅時に生じるもの。
- 通常のクラスター密度であれば大きくデータに影響することはない。
- クラスター密度が非常に高い場合には通常のサンプルよりもQ scoreが落ちることがある

# Nexteraサンプルは読み出し時に若干塩基配列が偏る

## ▶ ④Nexteraサンプル



- 主にトランスポゾームでの切断時に生じる偏り。データに与える偏りは小さい。
- 通常のクラスター密度であれば大きくデータに影響することはない。
- クラスター密度が非常に高い場合には通常のサンプルよりもQ scoreが落ちることがある

Low diversityサンプルを解析するためのテクニックとは？



# Low diversityサンプルを解析すると どのような症状が予測されるか？

- ・ クラスタ密度が正しく計算できない→%Cluster PFや%>=Q30の低下
- ・ 塩基の種類により取得するシグナルの強弱が不均一→  
シグナルの取得(Registration)がうまくできない。  
前後のシグナルとの補正がうまくいかなくなる。  
→%>=Q30の低下、Phasing/Prephasingのスコアが大きくなる。

より専門的な情報は下記の以前行われましたサポートウェビナーを  
ご参考ください

2013/04/19

サポートウェビナーシリーズ 2013

「塩基に偏りのある配列をMiSeqでシーケンスするには」

イルミナ株式会社 シニア テクニカル アプリケーション サイエンティスト

酒井名朋子

# Low Diversityサンプルの見分け方は？

The screenshot displays the Illumina Sequencing Analysis Viewer 1.7.32 interface. The 'Analysis' tab is selected and highlighted with a red box. The 'Run Folder' field is empty, with 'Browse' and 'Refresh' buttons to its right. Below the 'Analysis' tab, the 'Status' section shows 'Extracted: 0', 'Called: 0', and 'Scored: 0'. The main interface is divided into four panels:

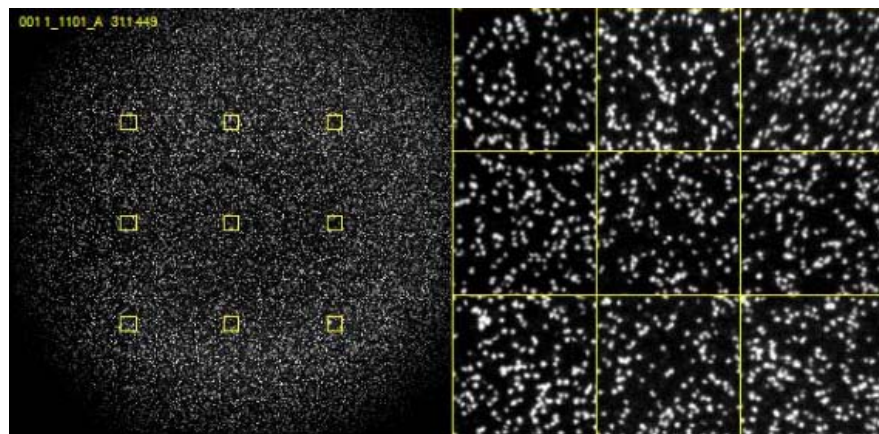
- Flowcell Chart:** A vertical bar chart showing intensity across different lanes. A color scale on the right ranges from 0.20 (blue) to 0.80 (yellow). Controls for 'Intensity', 'Top Surface', 'Cycle 1', 'Base A', and 'Fix Scale' are on the left.
- Data By Cycle:** A line graph showing intensity over cycles. A dropdown menu is open, with '% Base' selected and highlighted by a red box. A yellow box with the text '% Base' is overlaid on the graph. The x-axis ranges from 0 to 100, and the y-axis from 0 to 80.
- QScore Distribution:** A line graph showing the distribution of Q-scores. The x-axis ranges from 0 to 40, and the y-axis from 0 to 80. Controls for 'All Lanes', 'Both Surfaces', 'All Reads', and 'All Cycles' are on the left.
- Data By Lane:** A line graph showing data across lanes. The x-axis is labeled 1 through 8, and the y-axis from 0 to 80. Controls for 'Clusters' and 'Both Surfaces' are on the left.
- Qscore Heatmap:** A heatmap showing Q-scores across lanes. Controls for 'All Lanes' and 'Both Surfaces' are on the left.

# Low Diversityサンプルの見分け方は？

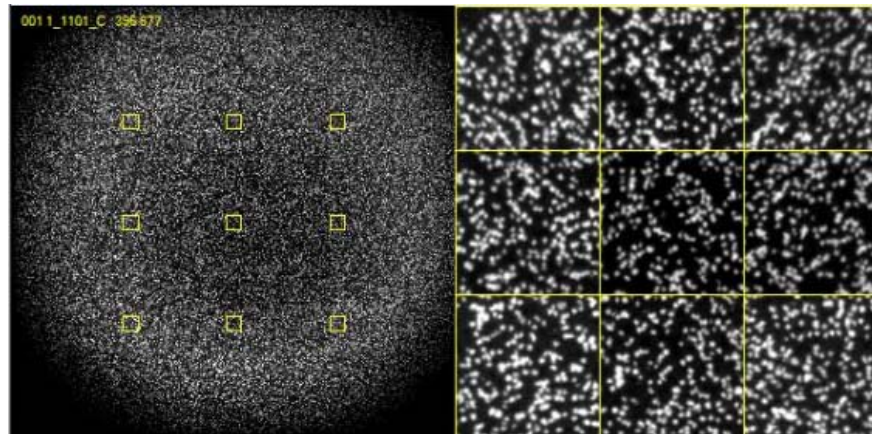
The image illustrates how to identify low diversity samples in sequencing data. It compares high diversity (stable base percentages) with low diversity (highly fluctuating base percentages) across different lanes and cycles. The central panel highlights the '% Base' metric in the 'Data By Cycle' view, which is used to detect these fluctuations.

# テクニック：簡単に考えると2通り

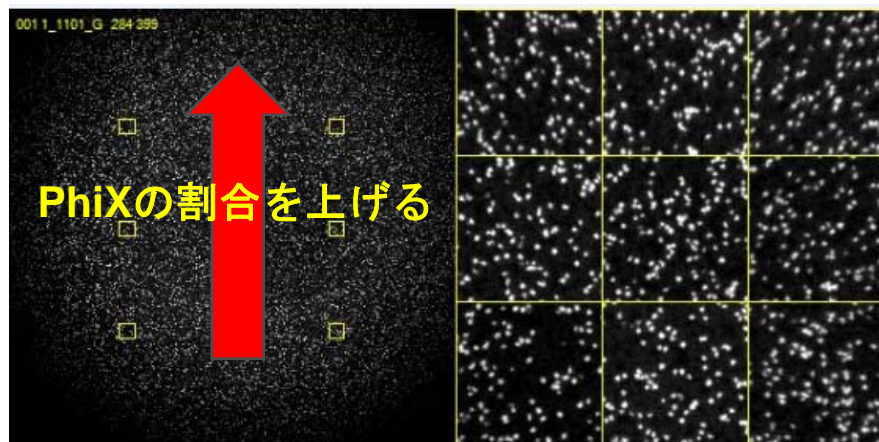
**A**



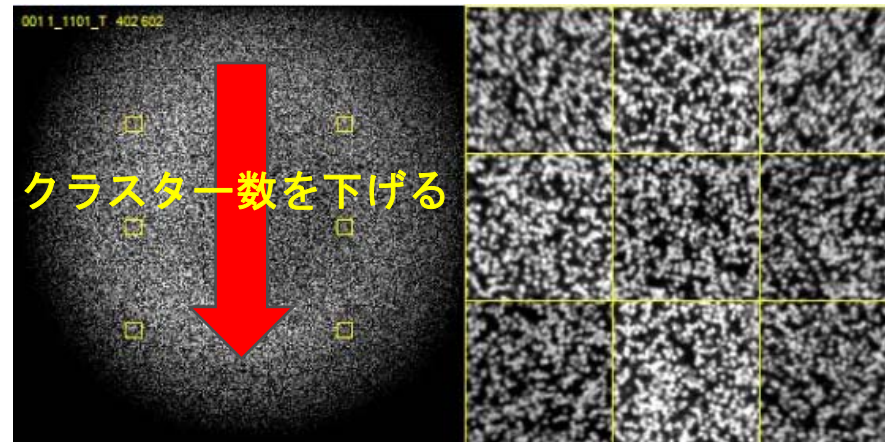
**C**



**G**: 若干少なめなクラスター密度



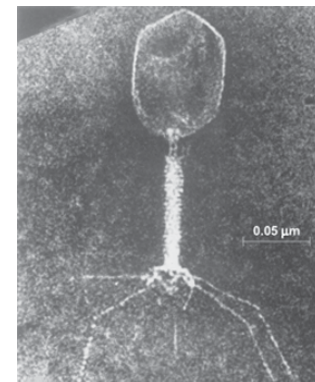
**T**: 明らかなオーバークラスター



# PhiXをspike-inして（加えて）解析する

PhiXは弊社で提供しているバクテリオファージ由来の  
とても塩基のばらつきの良いDNAライブラリーのコントロール  
別売りで提供、PhiX control v3 (FC-110-3001)

- 塩基の多様性を**全体的に**上げる
- サンプルのアウトプット量は減る  
（必要のないPhiX由来の配列が得られるため）
- PhiXはインデックスを持たないため、  
Index ReadのQ scoreの補正には役に立たない。



下記のサポートウェビナーを参照ください

2013/03/22

サポートウェビナーシリーズ 2013

「PhiXを使用してRun Qualityを改善する」

イルミナ株式会社 テクニカル アプリケーション サイエンティスト  
渡辺真子

# PhiXを何%加えればよいか？

## 16S Metagenomic Sequencing Library Preparation

*Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*



### NOTE

The recommended **PhiX control spike-in of  $\geq 5\%$  for low diversity libraries** is possible with RTA v1.17.28 or later, which is bundled with MCS v2.2. For optimal performance, update to v3 software (MCS 2.3). If you are using an older version of the MiSeq software or sequencing these libraries on the GA or HiSeq, Illumina recommends using  $\geq 25\%$  PhiX control spike-in.

### RTA v1.17.28 リリースノートより

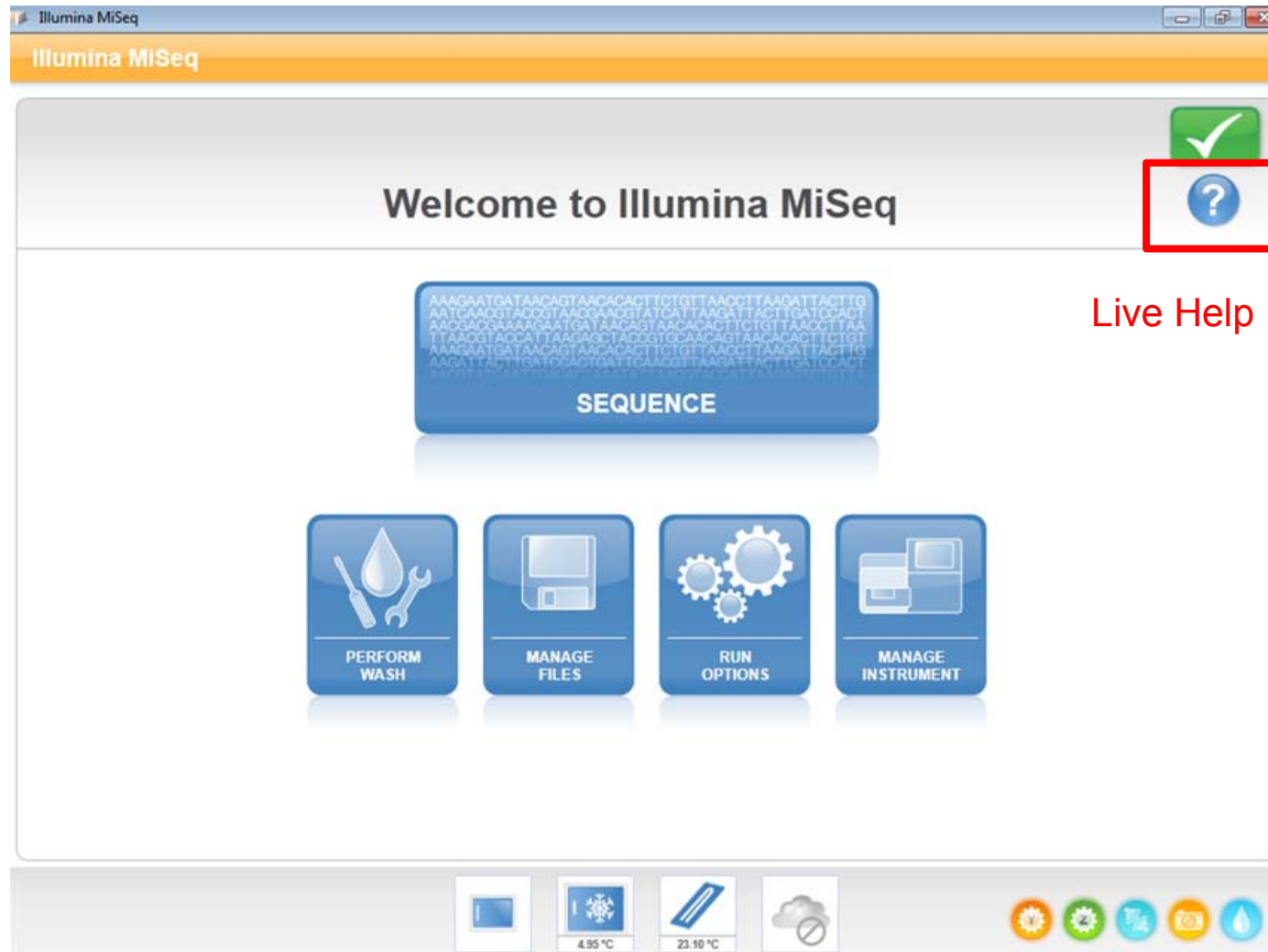
1. RTA 1.17.28 から一次解析のアルゴリズムがいくつか変更になりました。社内評価テストにおいては、少なくともこれ以前の RTA と同等あるいは若干の向上が見られる成果となっております。特に low diversity のサンプルにおいては、概ねデータクオリティならびに出力量 (Yield) がやや改善するケースが多くみられました。

MiSeqの場合は、Low Diversityサンプルに適した解析ソフト RTAv1.17.28 (MCS2.2) 以降にアップデートが必要  
⇒最低5%の添加で解析可能。

HiSeqの場合は、25%程度

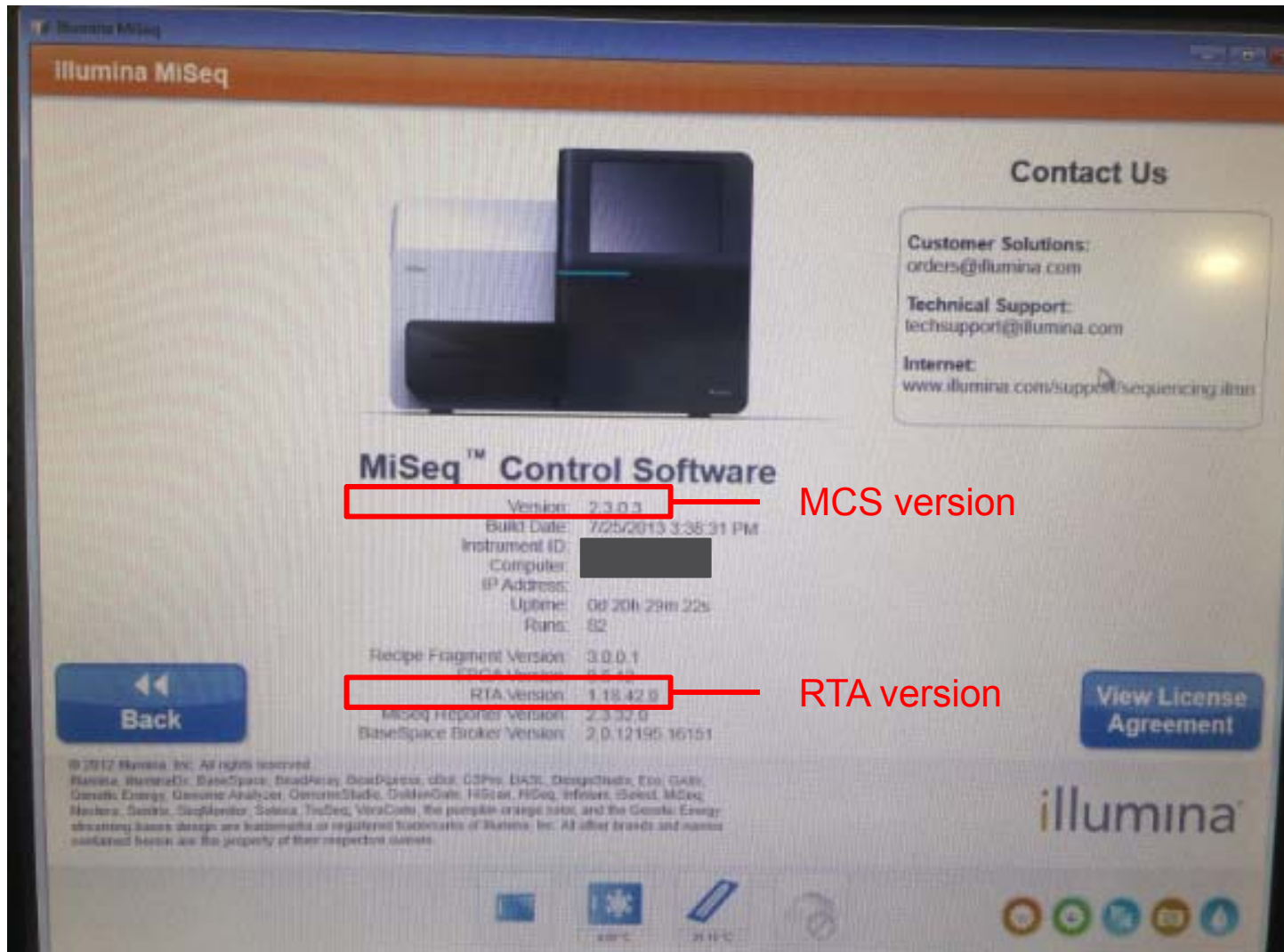
通常は25%程度のPhiX添加で結果が良好であれば下げていく方向で検討ください。

# MiSeqの場合・MCS/RTAのバージョンチェックの方法



Live Help

# MiSeqの場合・MCS/RTAのバージョンチェックの方法





## クラスター密度を減らす

- クラスター密度を減らして隣のクラスターとの距離を保つように調整する。  
⇒クラスターの認識の向上。（RTA1.17.28以上を使用を強くお勧め）  
目安としてクラスター密度を500K/mm<sup>2</sup>程度を目指してアプライする。

アウトプット量は減ることになるので、必要なカバレッジ数など注意する。



## まとめ

PCRアンプリコン、16S metagenomeサンプルなどは  
Low diversityサンプルである可能性が高いと思われるので、  
MiSeq) MCS2.2/RTA1.17.28以上のMiSeqで5~25% PhiXを加えて解析を行う  
HiSeq/GAllx) 25% PhiXを加えて解析を行う  
500K/mm<sup>2</sup>程度のクラスター密度を目指す（通常より少なめ）。

Nextera/RNAサンプル由来のライブラリーは上記のサンプルほどではないが  
クラスター密度の上限ぎりぎり（1000K/mm<sup>2</sup>以上）を狙わないように気をつける。  
PhiXを5%程度加えて解析する。

解析後はSAV上の%Base plotをモニターし、どの程度偏りがあるのかにより  
次ランのPhiX%やアプライ濃度を決めていく。  
SAV上でクラスター密度、%Cluster PF、Phasing/Prephasing、%>=Q30の数値も  
モニターし塩基の偏りが大きいことによりQ scoreが低くなっていないか調べる。