

PhiXを使用してRun Qualityを改善する

渡辺 真子

テクニカルアプリケーションサイエンティスト



© 2012 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPPro, DASL, DesignStudio, Eco, GALiX, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

illumina®

このウェビナーからは...

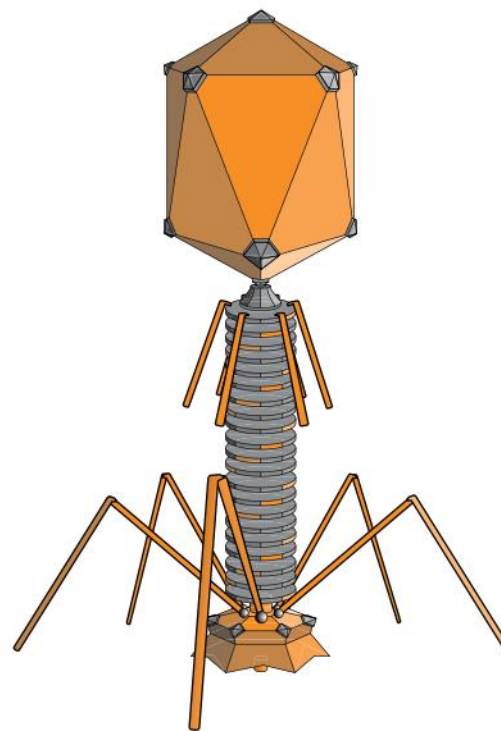
- ▶ PhiX とは何か、イルミナが提供するPhiX はどのようなものか。
- ▶ シーケンスの多様性とバランスを定義し、データのクオリティーとどのように関連しているか。
- ▶ 低い%と高い%でのPhiX添加 (spike-in) の違いとPhiX コントロールレーンの指定について。
- ▶ 塩基の多様性が低く、塩基のバランスがとれていないサンプルについて、ランのクオリティーを改善するために、PhiX をいかに利用できるか。



理解できるようになります。



PhiXとは?



PhiX DNA はランのコントロールとして使用

- ▶ PhiX はバクテリオファージ。
 - スモールゲノム → すぐにアライメントやエラーレートの評価ができる。
 - 多様性がありランダムな配列 → 約 45% GC と 55% ATを含む。
 - 明確な定義 → リファレンスゲノムがしっかりと注釈付けされている。
- ▶ イルミナのPhiX コントロールライブラリーはすぐに使用できる。
 - アダプターをライゲートして提供。
 - 濃度は10nM (目的の濃度に希釈して使用)。
- ▶ PhiX ライブラリーは以下を提供するもの。
 - Q Cに使用。
 - クラスター形成のポジティブコントロール。
 - シーケンスとアライメントのコントロール。
 - Cross-talk またはmatrix 算出、phasing・prephasing のコントロール。
(装置のキャリブレーション)

イリミナ PhiXライブラリー

- ▶ v3 PhiX (FC-110-3001)
 - インデックスなし (“undetermined” フォルダーにリードが保存)
 - 約420 bp (インサートは300 bp)





なぜバランスのあるランダムな
ゲノムが必要か？

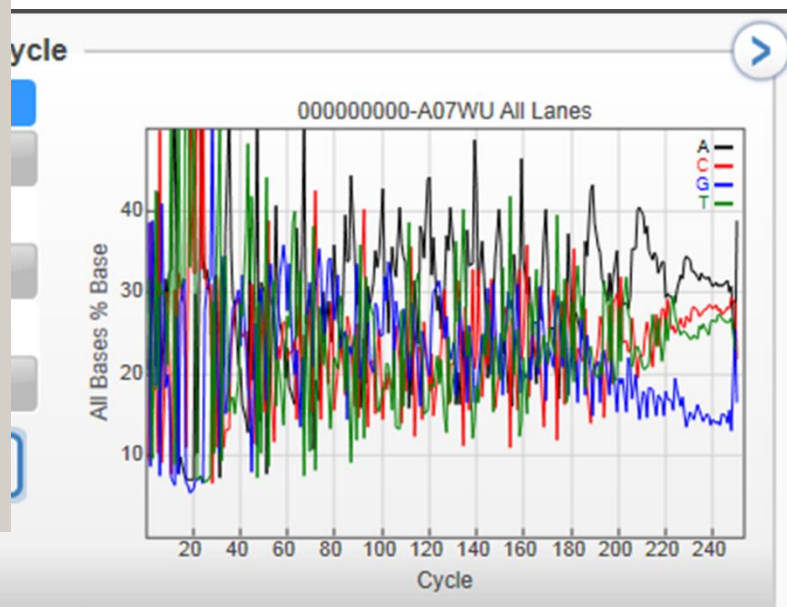
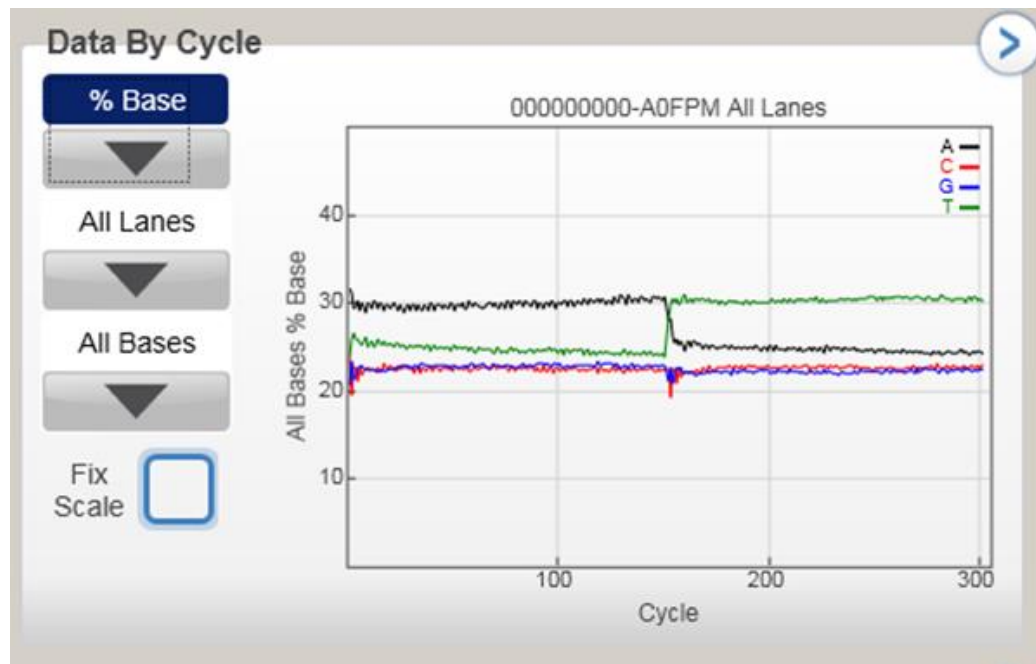
シーケンスの多様性の定義

- 塩基バランスのとれていないサンプル:

GC またはAT-含量の高い配列, Methyl-seq, ChIP-seq.

- 多様性の低いライブラリー: 同じ配列のライブラリー

(例えば、PCR amplicons, restriction enzyme libraries, initial cycle indexing).

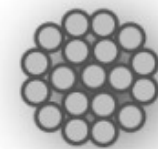
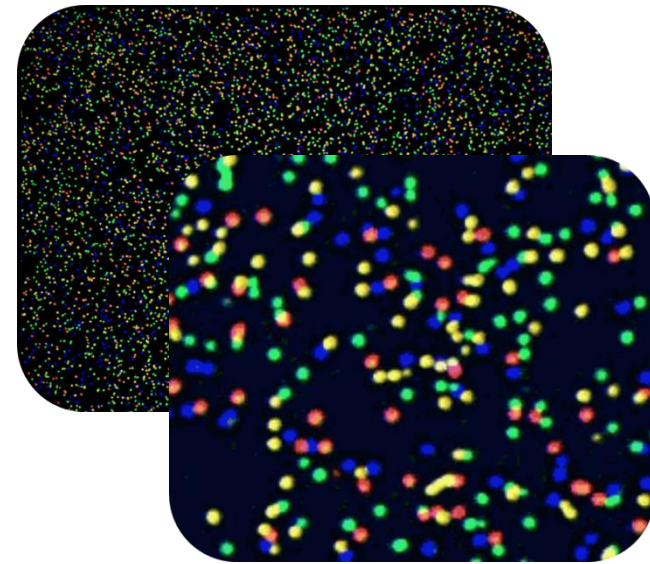


装置のキャリブレーション

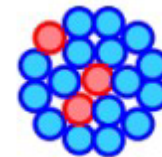
- ▶ **Template generation:** フローセル上のクラスターの位置のマッピング。
 - 2つの隣接したクラスターを区別するためにランダムな配列が必要。
- ▶ **Matrix calculations:** 各塩基からの蛍光強度の算出。マトリックスは残りのサイクルで補正した蛍光強度を生成するために使用される。
 - 4種類の塩基が必要
- ▶ **Phasing/prephasing:**
 - Phasing はクラスターにある分子が遅れて反応するときにかかる。
 - Prephasing はクラスターにある分子が進んで反応するときにかかる。

これらの算出のために、最初の12 サイクルで、バランスがありランダムな塩基組成が必要。

以上の値は、ラン全体のクオリティーに寄与する。



Cluster





ランのクオリティーを
改善するために
PhiXをどのように
利用できるか？



PhiXを使用する主な3つの方法*

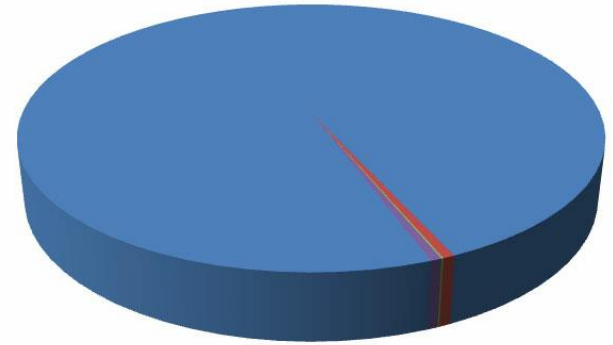
- ▶ 低い%の添加
- ▶ 高い%の添加
- ▶ コントロールレーンの指定



*De novo assembly またはPhiX の近縁種のシーケンスを行う場合は、添加に関して注意して下さい。

低いパーセンテージでの添加, PhiX 1% 添加

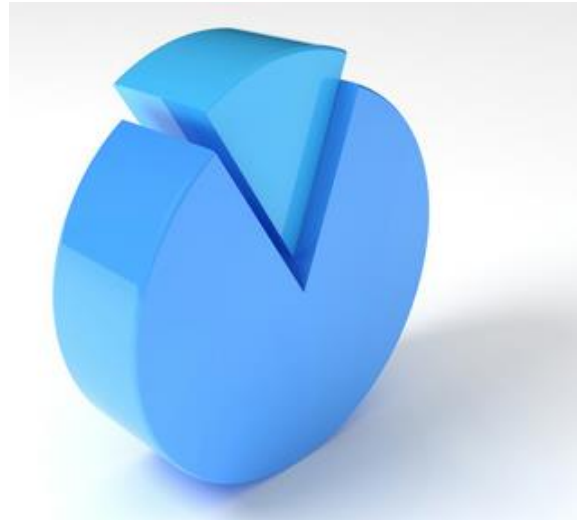
- ▶ 塩基バランスのとれたゲノムを使用し、レーン内にコントロールが必要な場合に利用。
- ▶ アライメントのコントロールとしての役割。
 - Real time analysis (RTA) ソフトウェアが、PhiX シーケンス25サイクル以降にリファレンスゲノムへのアライメントを開始し、エラーレートを算出する。
- ▶ ランのクオリティー、流路系等をモニターするのに役立つ。
- ▶ 数値化のコントロール
 - 1%PhiXを添加すると、1%のクラスターがPhiX として検出される。
- ▶ この濃度では、塩基の多様性が低く、バランスとれていないサンプルへの対応はできません。



高いパーセンテージでの添加 *

- ▶ 塩基組成に偏りがあるサンプルや多様性の低いサンプルのランのクオリティーの改善に役立つ。
 - e.g. GC-rich または PCR amplicons.
- ▶ 高い % のPhiX 添加 で多様性が増すことは、phasing/prephasing と matrix の補正に役立つ。
- ▶ 物理的にクラスターを分離するために、クラスター密度を下げ、PhiX の添加と組み合わせる。*
 - template generationに役立つ。

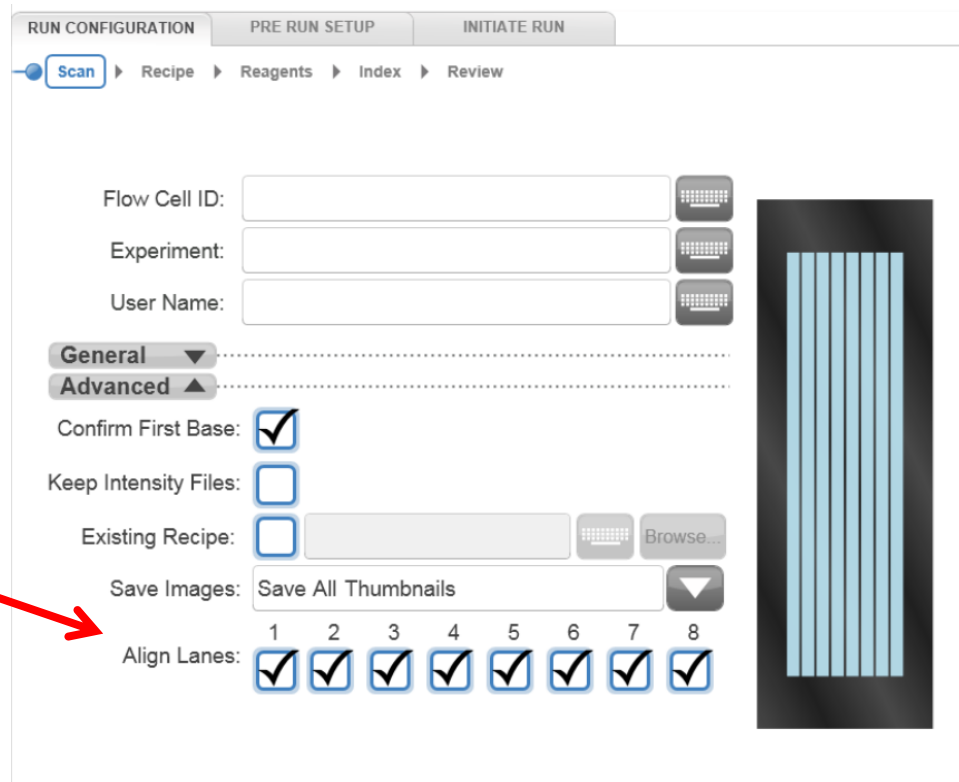
*添加%やクラスター密度を下げることは経験的に決定。



PhiX 添加によるエラーレートとアライメント

Align Lanes

デフォルトで、Real Time Analysis (RTA) でアライメントするよう全てのレーンが選択されている。サンプルにPhiX を添加していないレーンのチェックボックスをはずすと、RTA はアライメントやエラーレートの計算をそのレーンでしないことになる。



The screenshot displays the 'RUN CONFIGURATION' tab in the Illumina sequencing software. The 'Align Lanes' section is visible, showing a grid of 8 lanes, each with a checkbox. All checkboxes are checked, indicating that all lanes are selected for alignment. A red arrow points from the text on the left to the 'Align Lanes' section. To the right of the interface is a diagram of a flow cell with 8 vertical lanes.

Flow Cell ID:

Experiment:

User Name:

General ▼

Advanced ▲

Confirm First Base: ☒

Keep Intensity Files: ☐

Existing Recipe:

Save Images:

Align Lanes: ☒ 1 ☒ 2 ☒ 3 ☒ 4 ☒ 5 ☒ 6 ☒ 7 ☒ 8

PhiX 添加によるエラーレートとアライメント

Level	Yield Total (G)	Projected Total Yield (G)	Yield Perfect (G)	Yield <=3 errors (G)	Aligned (%)	% Perfect [Num Cycles]	% <=3 errors [Num Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	160.8	160.8	119.3	154.9	96.63	75.5 [100]	98.0 [100]	0.54	8271	76.2	93.3
Read 2	160.8	160.8	136.8	153.6	97.25	87.3 [100]	98.0 [100]	0.47	7254	76.4	91.4
Total	321.7	321.7	256.1	308.5	96.94	81.4	98.0	0.51	7762	76.3	92.3

Read 1

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	96	797 +/- 75	94.68 +/- 0.96	0.187 / 0.178	220.22	208.35	91.9	100	98.6 +/- 0.8	0.74 +/- 0.74	0.47 +/- 0.75	0.49 +/- 0.64	0.74 +/- 0.74	8253 +/- 315	76.4 +/- 0.9
2	96	788 +/- 73	94.58 +/- 1.24	0.191 / 0.173	217.76	205.80	93.1	100	98.5 +/- 1.0	0.63 +/- 0.76	0.02 +/- 2.04	0.64 +/- 0.97	0.63 +/- 0.76	8266 +/- 373	76.3 +/- 1.1
3	96	778 +/- 60	81.88 +/- 28.58	0.191 / 0.176	215.01	176.88	93.0	100	82.5 +/- 35.9	0.69 +/- 0.80	0.45 +/- 0.89	0.48 +/- 0.59	0.69 +/- 0.80	8265 +/- 403	75.6 +/- 2.5
4	96	780 +/- 77	94.99 +/- 0.79	0.183 / 0.172	215.70	204.74	92.1	100	98.8 +/- 0.3	0.80 +/- 1.13	0.25 +/- 0.36	0.84 +/- 1.44	0.80 +/- 1.13	8192 +/- 400	75.8 +/- 2.4
5	96	739 +/- 70	95.41 +/- 0.84	0.183 / 0.174	204.24	194.72	94.8	100	98.9 +/- 0.0	0.27 +/- 0.05	0.13 +/- 0.04	0.19 +/- 0.03	0.27 +/- 0.05	8291 +/- 396	76.1 +/- 1.2
6	96	784 +/- 62	94.93 +/- 1.18	0.185 / 0.172	216.82	205.68	93.5	100	98.8 +/- 0.3	0.52 +/- 0.45	0.22 +/- 0.23	0.48 +/- 0.52	0.52 +/- 0.45	8331 +/- 378	76.3 +/- 1.2
7	96	783 +/- 77	94.56 +/- 1.52	0.185 / 0.175	216.45	204.54	94.0	100	98.0 +/- 3.7	0.41 +/- 0.29	0.46 +/- 0.72	0.36 +/- 0.36	0.41 +/- 0.29	8341 +/- 338	76.4 +/- 0.9
8	96	792 +/- 80	94.94 +/- 0.92	0.188 / 0.173	218.86	207.60	94.3	100	98.9 +/- 0.1	0.29 +/- 0.05	0.14 +/- 0.04	0.20 +/- 0.04	0.29 +/- 0.05	8230 +/- 308	76.5 +/- 1.0

Read 2

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	96	797 +/- 75	94.68 +/- 0.96	0.188 / 0.170	220.22	208.35	91.2	100	97.4 +/- 0.2	0.48 +/- 0.16	0.18 +/- 0.13	0.33 +/- 0.09	0.48 +/- 0.16	7255 +/- 281	75.9 +/- 1.7
2	96	788 +/- 73	94.58 +/- 1.24	0.195 / 0.166	217.76	205.80	91.3	100	97.5 +/- 0.2	0.44 +/- 0.14	0.14 +/- 0.03	0.30 +/- 0.07	0.44 +/- 0.14	7350 +/- 317	75.6 +/- 2.1
3	96	778 +/- 60	81.88 +/- 28.58	0.209 / 0.164	215.01	176.88	90.5	100	95.8 +/- 10.2	0.71 +/- 0.62	0.26 +/- 0.37	0.52 +/- 0.54	0.71 +/- 0.62	7283 +/- 372	75.2 +/- 2.3
4	96	780 +/- 77	94.99 +/- 0.79	0.192 / 0.166	215.70	204.74	91.3	100	97.4 +/- 0.3	0.46 +/- 0.15	0.16 +/- 0.12	0.33 +/- 0.12	0.46 +/- 0.15	7284 +/- 329	75.9 +/- 1.9
5	96	739 +/- 70	95.41 +/- 0.84	0.187 / 0.168	204.24	194.72	92.2	100	97.5 +/- 0.2	0.41 +/- 0.09	0.13 +/- 0.02	0.28 +/- 0.05	0.41 +/- 0.09	7272 +/- 316	76.3 +/- 1.9
6	96	784 +/- 62	94.93 +/- 1.18	0.192 / 0.167	216.82	205.68	91.3	100	97.4 +/- 0.3	0.45 +/- 0.16	0.16 +/- 0.11	0.32 +/- 0.11	0.45 +/- 0.16	7265 +/- 367	77.1 +/- 1.4
7	96	783 +/- 77	94.56 +/- 1.52	0.183 / 0.167	216.45	204.54	91.6	100	97.4 +/- 0.2	0.41 +/- 0.09	0.14 +/- 0.04	0.29 +/- 0.06	0.41 +/- 0.09	7232 +/- 378	77.5 +/- 1.3
8	96	792 +/- 80	94.94 +/- 0.92	0.183 / 0.167	218.86	207.60	91.4	100	97.4 +/- 0.2	0.42 +/- 0.09	0.13 +/- 0.03	0.29 +/- 0.05	0.42 +/- 0.09	7086 +/- 384	77.8 +/- 1.1

高いパーセントでのPhiX添加によるランのクオリティーの改善

- ▶ イルミナでは、正確にベースコールされる確率を算出しており、データのクオリティー値として示しています。この値はQ scoreと呼ばれる。

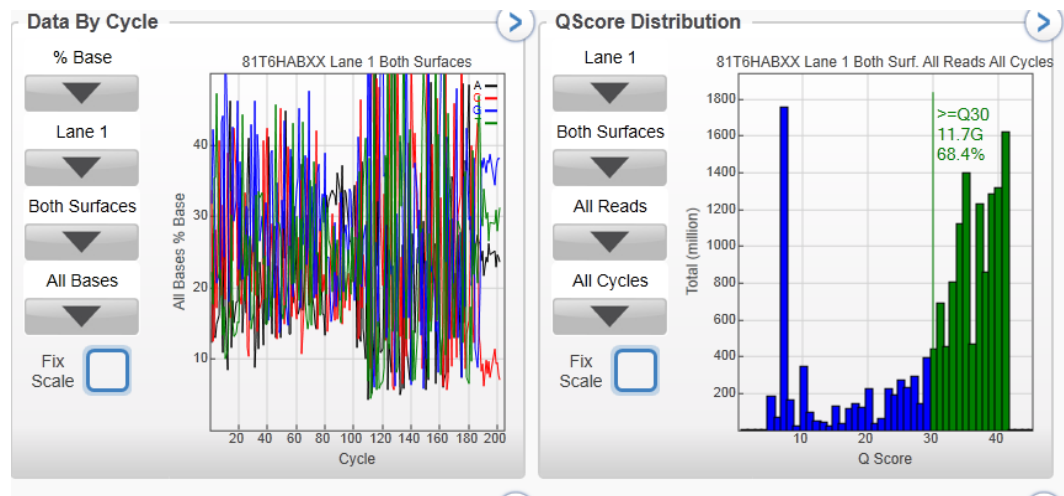


Percentage of Bases > Q30	> 85% (2 x 50 bp) > 80% (2 x 100 bp)	> 85% (2 x 50 bp) > 80% (2 x 100 bp)	> 85% (2 x 50 bp) > 80% (2 x 100 bp)	> 85% (2 x 50 bp) > 80% (2 x 100 bp)
---------------------------	---	---	---	---

- ▶ Q30 とは;
 - ベールコールのエラーの確率が1/1,000
 - ベースコールの正確性= 99.9%

PhiXを用いたランのクオリティーの改善

- 塩基の多様性の低いサンプル、低いQ scores



- 塩基の多様性が低いサンプルに、高い%のPhiX 添加し、クラスター密度を下げたもの。ランのクオリティーが改善。



PhiX コントロールレーンの指定 (HiSeq または GA のみ)

- ▶ 塩基の多様性が低く、偏りのあるライブラリーをロードしたフローセルに使用。
- ▶ 全レーンにわたり、指定したPhiXコントロールレーンからmatrixとphasingの計算を適用。
- ▶ 塩基の多様性が低いサンプルのQ scores を大きく改善。
- ▶ 指定したPhiXレーンは、オーバーロードしたレーンの補正はしません。
 - 推奨クラスター密度より高い。
 - パスフィルターを通った割合が低い。

HCS でのPhiX コントロールレーンの指定

Control Lane

設定する場合は、ドロップダウンリストからコントロールライブラリーを含むフローセルのレーンを選択。
設定しない場合はNone を選択。

RUN CONFIGURATION PRE-RUN SETUP INITIATE RUN

Scan Recipe Reagents Review

Please enter "Flowcell ID".

Flow Cell ID: Experiment: User Name: SBSUser

General

Flow Cell Type: Control Lane: None Output Folder: None

Advanced

1 2 3 4 5 6 7 8

BACK CANCEL HELP NEXT

Chiller 5.152°C FCA 25.471°C FCB 25.710°C

HCS でのPhiX コントロールレーンの指定

Sequencing Analysis Viewer

Run Folder:

Browse

Refresh

Analysis Imaging Summary Tile Status TruSeq Controls Indexing

Run Summary

Level	Yield Total (G)	Projected Total Yield (G)	Yield Perfect (G)	Yield <=3 errors (G)	Aligned (%)	% Perfect [Num Cycles]	% <=3 errors [Num Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	68.5	68.5	6.6	8.0	97.23	82.3 [50]	99.6 [50]	0.47	5698	91.9	92.9
Read 2 (I)	8.2	8.2	0.0	0.0	0.00	0.0 [6]	0.0 [6]	0.00	5303	0.0	78.1
Total	76.7	76.7	6.6	8.0	97.23	82.3	99.6	0.47	5500	91.9	91.1

Read 1

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	48	698 +/- 85	81.95 +/- 9.78	0.246 / 0.166	192.88	156.06	93.0	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6908 +/- 397	80.7 +/- 0.8
2	48	715 +/- 99	91.65 +/- 2.59	0.144 / 0.146	197.79	180.69	95.7	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6822 +/- 386	82.9 +/- 1.4
3	48	496 +/- 75	89.68 +/- 0.93	0.161 / 0.179	137.10	122.86	96.1	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	5966 +/- 376	94.8 +/- 2.0
4 (C)	48	671 +/- 57	89.71 +/- 4.76	0.154 / 0.148	185.41	165.87	94.5	50	97.2 +/- 3.8	0.47 +/- 0.54	0.48 +/- 0.74	0.00 +/- 0.00	0.00 +/- 0.00	5853 +/- 261	79.8 +/- 1.3
5	48	839 +/- 59	83.34 +/- 11.25	0.167 / 0.145	231.93	192.56	93.4	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	5769 +/- 407	91.6 +/- 1.9
6	48	849 +/- 62	82.03 +/- 9.45	0.212 / 0.154	234.85	192.16	91.4	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	4822 +/- 353	100.1 +/- 3.0
7	48	829 +/- 66	83.27 +/- 8.96	0.195 / 0.154	229.31	190.08	92.9	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	5305 +/- 376	97.8 +/- 2.2
8	48	968 +/- 55	63.90 +/- 15.66	0.237 / 0.123	267.74	170.15	87.4	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	4137 +/- 385	107.3 +/- 5.4

Read 2 (I)

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	48	698 +/- 85	81.95 +/- 9.78	0.000 / 0.000	192.88	156.06	92.0	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6935 +/- 378	0.0 +/- 0.0
2	48	715 +/- 99	91.65 +/- 2.59	0.000 / 0.000	197.79	180.69	95.5	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6516 +/- 408	0.0 +/- 0.0
3	48	496 +/- 75	89.68 +/- 0.93	0.000 / 0.000	137.10	122.86	87.9	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6853 +/- 365	0.0 +/- 0.0
4 (C)	48	671 +/- 57	89.71 +/- 4.76	0.000 / 0.000	185.41	165.87	0.0	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0 +/- 0	0.0 +/- 0.0
5	48	839 +/- 59	83.34 +/- 11.25	0.000 / 0.000	231.93	192.56	86.9	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6727 +/- 401	0.0 +/- 0.0
6	48	849 +/- 62	82.03 +/- 9.45	0.000 / 0.000	234.85	192.16	91.5	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	5149 +/- 465	0.0 +/- 0.0
7	48	829 +/- 66	83.27 +/- 8.96	0.000 / 0.000	229.31	190.08	92.9	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	6408 +/- 423	0.0 +/- 0.0
8	48	968 +/- 55	63.90 +/- 15.66	0.000 / 0.000	267.74	170.15	73.9	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	3837 +/- 886	0.0 +/- 0.0

Copy to Clipboard...

Generate IVC Plots...

PhiX: FAQs

- ▶ 多様性の低いサンプルに対して、PhiX を1%添加しランのクオリティーを改善できますか？
 - いいえ。1%添加はレーン内のコントロールとして使用しますが、ランのクオリティーを改善する役割はありません。高い%の添加により塩基の多様性を増やし、ランのクオリティーを改善します。
- ▶ PhiX レーンが良好な場合、何か他に悪いところがあるように見えるのは？
 - サンプル調製またはライブラリーの問題に起因する可能性があります。
- ▶ どのようにして、使用中の装置がランを通して適切に動作をしているか確認できますか？
 - もしレーン毎にPhiX を添加していれば、装置のパフォーマンスをレーン毎に評価することができます。
- ▶ 1% PhiXを添加したにもかかわらず、PhiXが全リードの多くを占めるのは何を意味するのでしょうか？
 - サンプルの定量に問題があります。
- ▶ PhiX の添加とPhiX コントロールレーンの指定を組み合わせた場合、何か利点がありますか？
 - もちろんあります。コントロールレーンの指定と添加は2つの異なる目的があります。実験のニーズに合わせて組み合わせられます。

まとめ

- ▶ PhiX は、ゲノムのサイズが小さく、多様性があり、良く定義されているため、良いコントロールとして機能します。
- ▶ コントロールレーンの指定;
 - 塩基バランスのとれたサンプルまたはランダムな配列をもつサンプルがないとき。
 - Phasing /pre-phasing の値をフローセル全体に提供するため。
- ▶ 高いパーセントでの添加;
 - 塩基に偏りがあるサンプル(例 GC-rich)や多様性が低いサンプル(例 PCR amplicons)で使用。
 - クラスターを分離するために、クラスター密度を下げる。
- ▶ 低いパーセントでの添加;
 - 塩基の多様性が低く、偏りのあるサンプルの補正には使用できません
 - アライメントのコントロールやレーン毎のサンプルのQCとして使用。

適切にPhiX を使用することで、利用できるデータとそうではないデータの違いを示すことができます。

▶ PhiX TechNote:

- http://www.illumina.com/documents/product/s/technotes/technote_phixcontrolv3.pdf

▶ Illumina homepage

- <http://www.illumina.com/>

▶ Sequencing Support

- <http://support.illumina.com/sequencing.ilmn>

Technical Note: Sequencing

illumina

Using a PhiX Control for HiSeq® Sequencing Runs

A low-concentration spike-in of Illumina PhiX Control v3 provides quality and calibration controls.

Highlights

- For most libraries, Illumina recommends using PhiX Control v3 (Catalog # FC-110-3001) in a low-concentration spike-in (1%) for improved sequencing quality control
- For samples with bias base composition, use a dedicated PhiX control lane to improve cross-talk and phasing calculations
- For samples with low diversity, use a dedicated control lane and a high-concentration spike-in (40% or higher) of PhiX to create a more diverse set of clusters

What are PhiX Control Libraries?

Control libraries generated from the PhiX virus serve as an effective control in sequencing runs. Characteristics of the PhiX genome provide several benefits:

- **Small**—PhiX is a small genome, which enables quick alignment and estimation of error rates.
- **Diverse**—The PhiX genome contains approximately 45% GC and 55% AT.
- **Well-Defined**—PhiX has a well-defined genome sequence.

Illumina cluster detection algorithms are optimized around a balanced representation of A, T, G, and C nucleotides. Illumina PhiX Control v3, with an insert size of 425–525 bp, is a balanced and diverse library that can help mitigate sequencing challenges in unbalanced and low diversity libraries.

Benefits of Using a PhiX Control

PhiX libraries provide a quality control for cluster generation, sequencing, and alignment, and a calibration control for cross-talk matrix generation, phasing, and prephasing.

Cluster Generation

Illumina PhiX Control v3 is shipped as a ready-to-use 10 nM library that can be used as a positive control in the clustering process. If a problem occurs in sample preparation, PhiX will still generate clusters. This helps you discern whether a lack of clusters is due to sample preparation failure or a failure in the cluster generation process.

Cross-Talk Matrix Generation

During an Illumina sequencing run, the cross-talk due to spectral overlap between the four fluorescently-labeled nucleotides is calculated during template generation (cluster identification) in cycles 1–4.

For proper cross-talk calculation, HiSeq Control Software (HCS) v1.3.8, or higher, requires approximately equal numbers and at least 50,000 clusters/mm² in each of the four bases. Therefore, it is imperative to have a balanced representation of bases at the beginning of each read, excluding the index read.

Phasing and Prephasing

During sequencing by synthesis, each DNA strand in a cluster extends by one base per cycle. A small proportion of strands may become out of phase with the current cycle, either falling a base behind (phasing) or jumping a base ahead (prephasing). The phasing and prephasing rates define the fraction of molecules that become phased or prephased per cycle.

Calculation of these rates requires a balanced and random base composition in cycles 2–12. Any library that does not comprise a balanced base composition (e.g., initial cycle indexing, restriction enzyme libraries) should use a control lane.

High GC samples typically show higher phasing rates. However, if the sample has good diversity (for example, whole-genome sequencing libraries with 40–60% GC), it does not require a control lane.

Alignment

Because PhiX has a small, well-defined genome sequence, it is an excellent alignment control. If a lane containing PhiX (control lane or low-concentration spike-in) is designated for your run, Real Time Analysis (RTA) software aligns complete sequences to the PhiX reference beginning after the 25th cycle is accumulated and calculates error rates, providing an indication of sequencing success during the run.

Using a PhiX Control for Unbalanced Samples

Most mammalian genomic or whole-transcriptome RNA samples have a balanced genomic composition (approximately equal proportions of A, T, G, and C). These samples do not require a dedicated control lane to generate accurate matrix and phasing estimations. Unbalanced samples contain genomes with high AT or GC content (less than 40% or greater than 60%). Some examples are Arabidopsis, Plasmodium, some bacteria, and bisulfite conversion studies.

Use a dedicated control lane with unbalanced samples to improve cross-talk and phasing calculations.

Using a PhiX Control for Low Diversity Libraries

Low diversity libraries are libraries where a significant number of the reads have the same sequence. This shifts the base composition because the reads are no longer random.

Low diversity can occur with some expression studies with > 25%



ご清聴ありがとうございました。

ご質問はtechsupport@illumina.comでも承ります。