

# NGSをはじめよう！ NextSeq500で何ができるのか？

April 10, 2015



崎川 真里  
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

# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめるアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめるアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# 次世代シーケンサーの用途（解析データ登録数）

## セルイノベーション公共SRAデータ統計情報より

[http://cell-innovation.nig.ac.jp/public/contents/sra\\_stat.html](http://cell-innovation.nig.ac.jp/public/contents/sra_stat.html)

集計日: 2015/04/08

**公共SRAデータ統計**

< Back

集計日: **2015/04/08**

データ件数(Run): **2,417,867 件** (live:1,248,795, suppressed:942,558, unpubl)

実験種別 (Library strategy)とは

シーケンサーの実験種別統計 (ランデータの件数)

No.	Sequencer Name	WXS	WGS	RNA-Seq	AMPLICON	ChIP-Seq
1	Illumina HiSeq 2000	96,100	311,586	109,032	13,314	17,706
2	454 GS FLX Titanium	7	12,866	1,952	76,940	3
3	Illumina Genome Analyzer II	8,938	40,469	15,742	1,784	8,614
4	Illumina MiSeq	712	32,638	2,431	37,836	458

解析の種類	HiSeq2000	MiSeq
エクソーム解析	96,100	712
全ゲノム解析*	311,586	32,638
RNA-Seq	109,032	2,431
アンプリコン	13,314	37,836
ChIP-Seq	17,706	458
Bisulfite-Seq	6,252	584
クローン配列確認	1,166	1,947

\*全ゲノム解析は MiSeqでは微生物、HiSeqではヒトなどの高等真核生物を対象とする場合が多い。

- 全ゲノム解析、エクソーム解析、RNA-Seqが人気

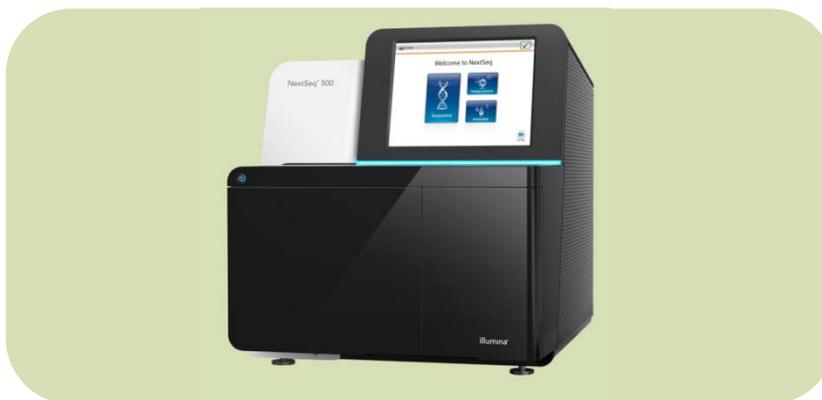
# イルミナ次世代シーケンサー・アプリケーション



	デスクトップ型	デスクトップ型	大型	大型	大型
	MiSeq	NextSeq 500	HiSeq 2500	HiSeq 4000	HiSeq X Ten
データ量	15 Gb	120 Gb	1,000 Gb	1,500 Gb	1,800 Gb
最大リード長	300x2	150x2	250x2	150x2	150x2
Gbあたりのコスト	約 1.7 万円	約 0.6 万円	HO 約 0.6 万円 RR 約 0.8 万円	約 0.4 万円	約 0.1 万円
アプリケーション					
遺伝子パネル	★★★	★★★★★	★★★	★★	非対応
ターゲット	★★★	★★★★★	★★★	★★★	非対応
RNA-Seq	★	★★★★★	★★★★★	★★★★★	非対応
エクソーム	★	★★★★★	★★★★★	★★★★★	非対応
全ゲノム	非対応	★★	★★★★★	★★★★★	★★★★★

# NextSeq500

全ゲノム解析にも対応のデスクトップ型シーケンサー



# NextSeq500のデータ出力量と試薬キット

キット	リード長	データ出力	ラン時間†	クオリティ‡
高出力フローセル (High Output Kit) ~400Mシングルリード ~800Mペアエンドリード	2 x 150 bp	100-120 Gb	29 時間	> 75% >Q30
	2 x 75 bp	50-60 Gb	18 時間	> 80% >Q30
	1 x 75 bp	25-30 Gb	11 時間	> 80% >Q30
中出力フローセル (Mid Output Kit) ~130Mシングルリード ~260Mペアエンドリード	2 x 150 bp	32-39 Gb	26 時間	> 75% >Q30
	2 x 75 bp	16-19 Gb	15 時間	> 80% >Q30

据付時の仕様は、イルミナPhiXコントロールライブラリーを使い、サポート範囲内でのクラスター密度（120-165 k/mm<sup>2</sup> パスフィルタークラスター）を基にしています。実際のパフォーマンス項目は、サンプルの種類、クオリティ、そしてパスフィルタークラスターに依存します。Q30以上の塩基の割合は、そのラン全体の平均値です。

†時間はNextSeq 500システム上で実施するクラスター形成、シーケンス、ベースコールを含みます（デュアルサーフェススキャン）。

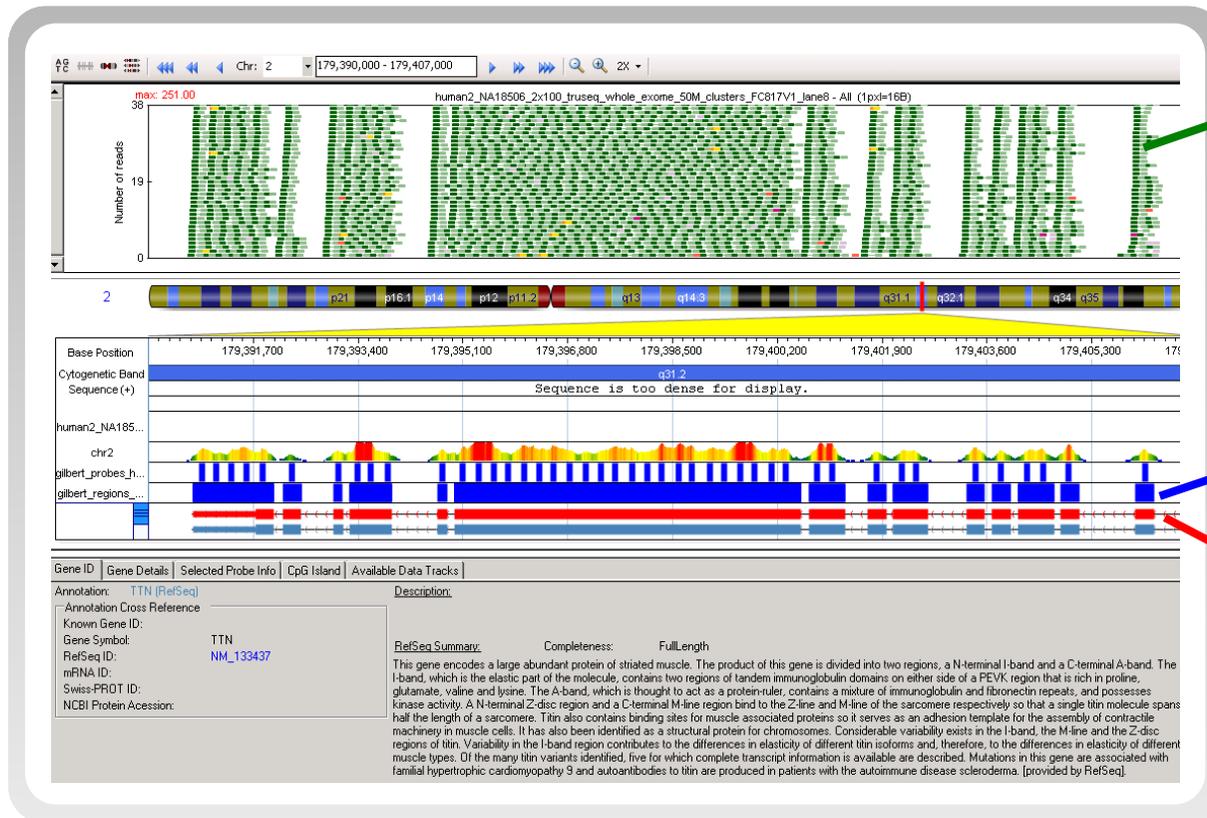
‡クオリティ値（Qスコア）は、ベースコールにおけるエラー確率の予測指標です。Q30以上の割合はラン全体の平均値を示します。リードごと、サイクルごとの値ではありません。

# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめするアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# エクソーム解析とは

gDNA断片からエクソン領域だけをキャプチャーし、濃縮を行い、シーケンス解析を行う。



シーケンスデータ  
(リード)

ターゲット領域

遺伝子  
(エクソン領域)

エクソン領域のみシーケンスを行うことで、シーケンスとデータ解析のコストを削減

# エクソーム解析のレビュー

## Nature Reviews Genetics

- ▶ Exome解析で原因遺伝子を示した29の文献を一覧表にまとめている
- ▶ ほとんどの解析が1～4家系、1～4症例で行われている
- ▶ 遺伝性疾患の解析に威力を発揮

Supplementary Table 1 | Diseases identified to date via exome sequencing

Disorder	PMID**	Mode of inheritance	N*	Strategy	Gene(s)
<b>Comparison of unrelated cases</b>					
Kabuki	20711175	AD	10	10 cases / 10 kindreds	MLL2
Schinzel-Giedion	20436468	AD	4	4 cases / 4 kindreds	SETBP1
Hadju-Cheney	21378985	AD	3	3 cases / 3 kindreds	NOTCH2
Fowler	20518025	AR	2	2 cases / 2 kindreds	FLVCR2
Sensenbrenner	20817137	AR	2	2 cases / 2 kindreds	WDR35
<b>Comparison of related cases</b>					
Miller	19915526	AR	4	4 cases / 3 kindreds	DHODH
hyperphosphatasia-MR	20802478	AR	3	3 cases / 1 kindred	PIGV
hypolipidemia	20942659	AR	2	2 cases / 1 kindred	ANGPTL3
retinitis pigmentosa	21295283	AR	3	3 cases / 1 kindred	DHDDS
novel skeletal dysplasia	21455487	AR	4	2 cases + parents / 1 kindred	POP1
spinocerebellar ataxia	21106500	AD	4	linkage + 4 cases / 1 kindred	TGM6
familial ALS	21145000	AD	2	linkage + 2 cases / 1 kindred	VCP
dilated cardiomyopathy	21353195	AD		linkage + 4 cases / 1 kindred	BAG3
hidradenitis suppurativa	21430701	AD	3	linkage + 2 cases <sup>3</sup> / 1 kindred	NCSTN
spinocerebellar ataxia	21106500	AD	4	linkage + 4 cases / 1 kindred	TGM6
primary failure tooth eruption	21404329	AD	4	linkage + 4 cases <sup>4</sup> / 1 kindred	PTH1R
TARP <sup>1</sup>	20451169	XLR	2	linkage + 2 cases / 2 kindreds	RBM10
X-linked leukoencephalopathy	21415082	XLR	2	linkage + 1 case <sup>3</sup> / 1 kindred	MCT8
<b>Homozygosity mapping</b>					
DFNB82 (hearing loss)	20602914	AR	1	1 case / 1 kindred	GPBM2
CNS malformations	20729831	AR	2	2 cases / 1 kindred	WDR62
Seckel	21131973	AR	1	1 case / 1 kindred	CEP152
NPHP-related ciliopathy <sup>2</sup>	20835237	AR	1	1 case / 1 kindred	SDCCAG8
autoimmune LPS	21109225	AR	1	1 case / 1 kindred	FADD
3MC	21035106	AR	1	1 case / 1 kindred	MASP1
complex I deficiency	21057504	AR	1	1 case / 1 kindred	ACAD9
non-syndromic MR	21212097	AR	2	2 obligate carrier parents	TECR
Ochoa	21450525	AR	1	1 case / 1 kindred	HPSE2
<b>Identification of <i>de novo</i> mutations</b>					
sporadic MR	21076407	complex	30	10 parent-child trios	multiple
autism		complex	60	20 parent-child trios	multiple

Bamshad MJ, et al., Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet.* 2011 Sep 27;12(11):745-55.

# 日本におけるエクソーム解析例

横浜市立大学 松本 直通 教授がエクソーム解析によって明らかにした遺伝病の原因遺伝子（一部）

年月	疾患名	原因遺伝子	掲載雑誌
2013年2月	SENDA（脳内鉄沈着神経変性症の1つ）	WDR45	Nature Genetics
2013年2月	ジュベール症候群と関連症候群	CEP290, TMEM67, INPP5E	Journal of Human Genetics
2012年8月	大田原症候群	KCNQ2	Annals of Neurology
2012年3月	コフィンーシリス症候群	SMARCB1、SMARCA4、SMARCE1、 ARID1A、ARID1B、SMARCA2	Nature Genetics
2012年2月	重症型もやもや病	RNF213	Neurology
2011年10月	HCAHC(びまん性大脳白質形成不全症)	POLR3A, POLR3B	American Journal of Human Genetics
2011年8月	ARCA(常染色体劣性遺伝性小脳変性症)	SYT14	American Journal of Human Genetics

※ 横浜市立大学ホームページより抜粋

## イルミナで提供しているエクソーム濃縮キットのご紹介

キット	Nextera Rapid Capture Exome	Nextera Rapid Capture Expanded Exome
ターゲットサイズ	37Mb	62Mb
コンテンツ	エクソン	エクソン、UTRs、miRNA
必要なシーケンス	4Gb以上	8Gb以上

カタログ番号	製品名
FC-140-1000	Nextera Rapid Capture Exome (8反応、1プレックス)
FC-140-1083	Nextera Rapid Capture Exome (8反応、3プレックス)
FC-140-1086	Nextera Rapid Capture Exome (8反応、6プレックス)
FC-140-1089	Nextera Rapid Capture Exome (8反応、9プレックス)
FC-140-1001	Nextera Rapid Capture Exome (2反応、12プレックス)
FC-140-1002	Nextera Rapid Capture Exome (4反応、12プレックス)
FC-140-1003	Nextera Rapid Capture Exome (8反応、12プレックス)

# NextSeq1ランに何サンプル流せるか？

## エキソーム解析

Illumina Sequencing System	Read Length	No. of Exomes/Run	% Bases at $\geq 20\times$
MiSeq System	2 × 150 bp	1	91%
<b>NextSeq 500 System</b>			
Mid Output Flow Cell (130 M)	2 × 150 bp	3	94%
High Output Flow Cell (400 M)	2 × 75 bp	6	90%
	2 × 150 bp	9	94%
<b>HiSeq 2500 System</b>			
Rapid-Run Mode	2 × 100 bp	24	90%
High-Output Mode	2 × 150 bp	96	85%

出典: NextSeq™ 500 System Exome Sequencing Solution

- ▶ NextSeq 500 Mid out putフローセル で推奨3サンプル
- ▶ NextSeq 500 High out putフローセル で推奨9サンプル

をインデックスを付けて一度に流すことができます。

# エクソーム解析に採用するカバレッジ

$$C = \frac{L \times N}{G}$$

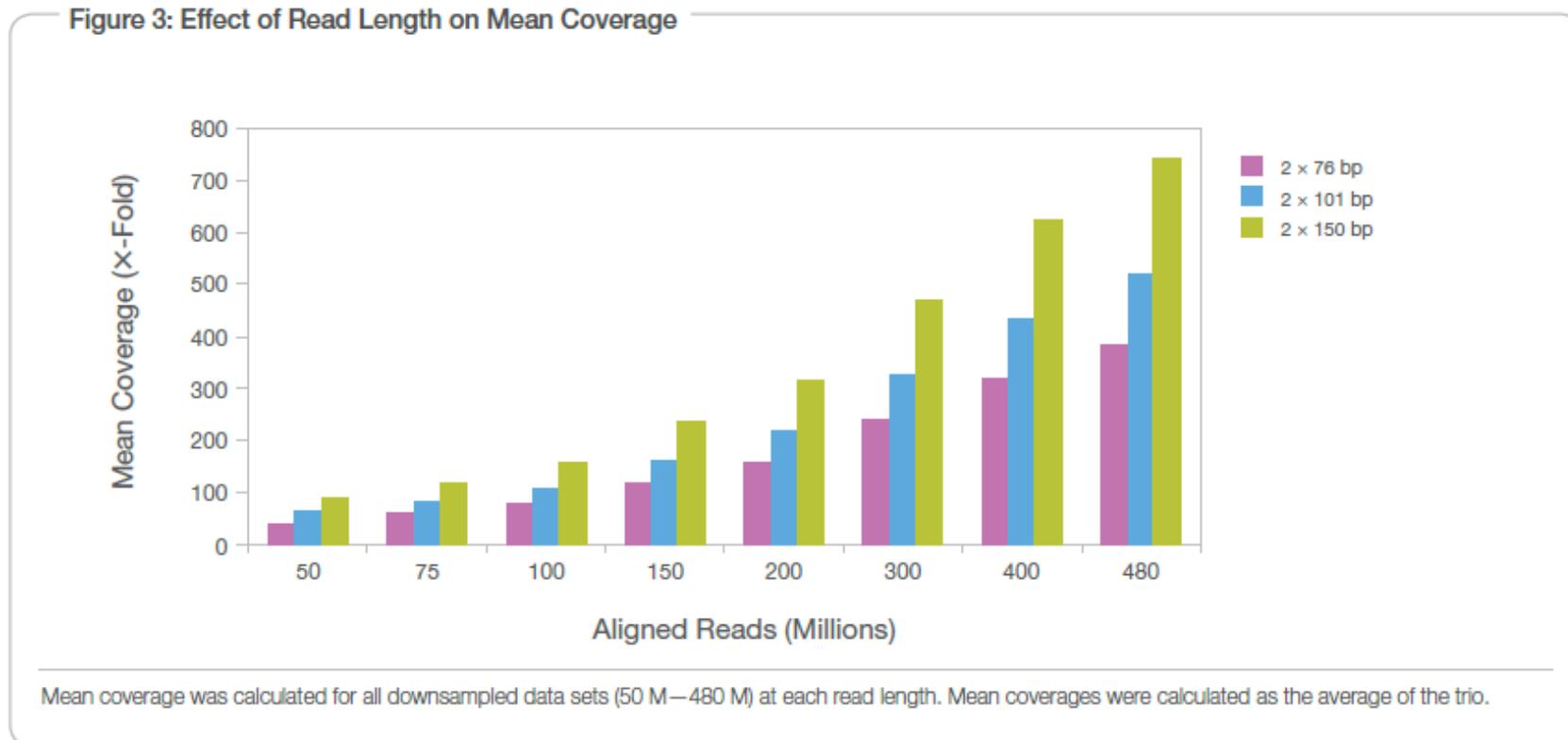
- ▶ C = x カバレッジ
- ▶ G = ターゲットサイズ(base)
- ▶ L = リード長(bp) = サイクル数
- ▶ N = リード数

キット	リード長	データ出力
高出力フローセル ~400M シングルリード ~800M ペアエンドリード	2 x 150 bp	100-120 Gb
	2 x 75 bp	50-60 Gb
	1 x 75 bp	25-30 Gb
中出力フローセル ~130M シングルリード ~260M ペアエンドリード	2 x 150 bp	32-39 Gb
	2 x 75 bp	16-19 Gb

例) Exome (ターゲットサイズ 37Mb) 3サンプルを等量で混合し、  
NextSeq500の中出力フローセルで、2x150bpのランを行った場合  
→ 1サンプルあたりのリード数は、260Mリード / 3 sample = 86.7 Mリード  
C = 150bp x 86.7 M / 37 Mb = およそ x351のカバレッジ

Coverage Calculation Tech Note  
[http://support.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote\\_coverage\\_calculation.pdf](http://support.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote_coverage_calculation.pdf)

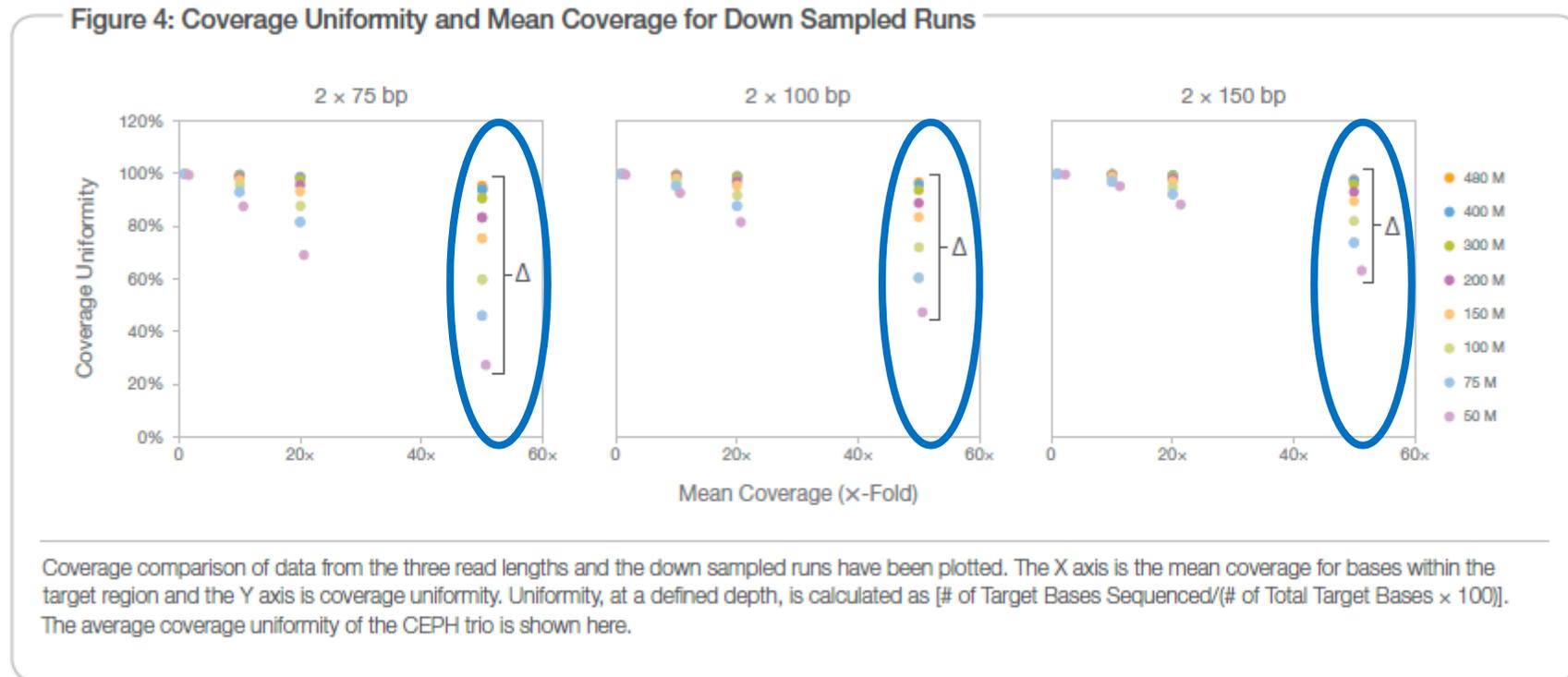
# リード数とカバレッジの関係



出典: Read Length and Nextera® Rapid Capture Exome Data

- ▶ Coverageを上げると、ターゲットにアラインされるリード数が増える
  - NextSeq 500 Mid out putフローセル で推奨3サンプル
  - NextSeq 500 High out putフローセル で推奨9サンプル

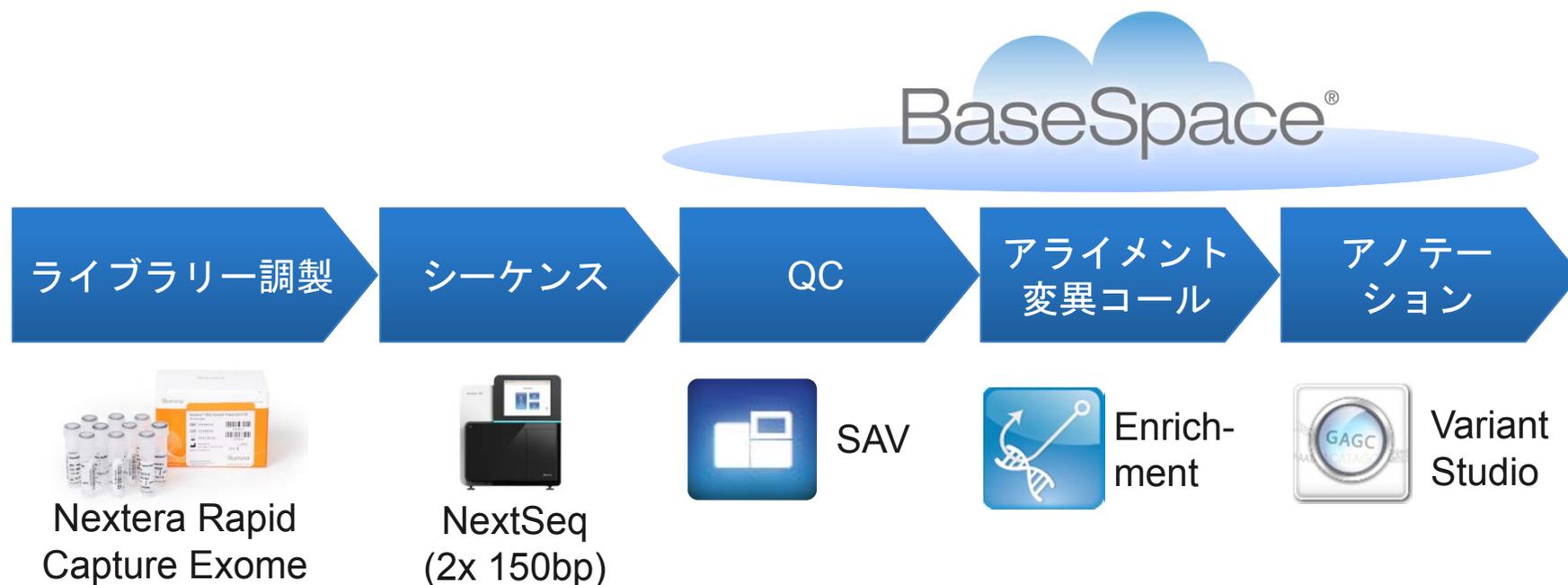
# リード長は長い方がいい



出典: Read Length and Nextera® Rapid Capture Exome Data

- ▶ リード長が長いほど、Coverage Uniformityにばらつきが少ない

# エクソーム解析 解析のワークフロー



2015/04/24 15:00-15:30

NGSをはじめよう！

Nextera Rapid Capture エクソームキットを用いたエクソームシーケンス？  
ウェット編-

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  
米田瑞穂

[+ 詳細](#)

[登録はこちら](#)

初心者向け  
サンプル調製

[http://www.illumina.co.jp/events/webinar\\_japan/support\\_webinar.ilmn](http://www.illumina.co.jp/events/webinar_japan/support_webinar.ilmn)

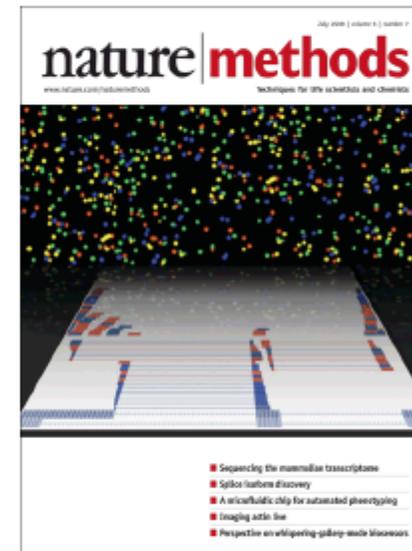
# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめするアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# RNA-Seqで何ができるか？

- ▶ 発現プロファイリング解析
  - 遺伝子の発現比較
  - エクソンの発現比較
  - **スプライスバリエーションの発現比較**
  - **アレルごとの発現比較**
  - **幅広いダイナミックレンジでの発現比較**
  - **アノテーションが充分で無い生物で発現比較**
  - 超微量からの発現解析
  - **ncRNAの発現プロファイリング解析**
- ▶ トランスクリプトーム解析
  - **mRNA領域の網羅的推定**
  - **新規スプライスバリエーションの発見**
  - **融合遺伝子の発見**
  - **ゲノム未知な生物において de novo アセンブルを用いて、mRNA配列の網羅的取得**
  - **新規 ncRNAの発見**

赤字 — マイクロアレイでは難しいこと



July 2008 - Vol 5 No 7  
Nature Methods 5 (2008)  
The beginning of the end for microarrays?

2014/04/18

NGSをはじめよう！

RNA-Seq入門(キットの選び方、実験デザイン)

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  
米田瑞穂



# RNA-Seq

## 癌において発現する偽遺伝子

Cell

Resource

### Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers

Shanker Kalyana-Sundaram,<sup>1,2,6,7</sup> Chandan Kumar-Sinha,<sup>1,2,7</sup> Sunita Shankar,<sup>1,2</sup> Dan R. Robinson,<sup>1,2</sup> Yi-Mi Wu,<sup>1,2</sup> Xuhong Cao,<sup>1,3</sup> Irfan A. Asangani,<sup>1,2</sup> Vishal Kothari,<sup>1</sup> John R. Prensner,<sup>1,2</sup> Robert J. Lonigro,<sup>1,2</sup> Matthew K. Iyer,<sup>1</sup> Terrence Barrette,<sup>1,2</sup> Achiraman Shanmugam,<sup>6</sup> Saravana M. Dhanasekaran,<sup>1,2</sup> Nallasivam Palanisamy,<sup>1,2</sup> and Arul M. Chinnaiyan<sup>1,2,3,4,5,\*</sup>

- 13種類の癌を対象に 293 サンプルを解析
- これまでの課題： 高い配列の類似性のために偽遺伝子と遺伝子の発現を分けて解析することは困難
- 本論文： RNA Seqにより、偽遺伝子を区別して発現を解析
  - 癌種特異的に発現している偽遺伝子の発現を観察
  - 偽遺伝子の発現は転写産物全体に重要な影響を与え、細胞間の違いや癌の進行に重要な役割を果たしていると考え

Kalyana-Sundaram S, et al., Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell*. 2012 Jun 22;149(7):1622-34.

# NextSeq 1ランに何サンプル流せるか？

## RNA-Seq

解析目的例	必要リード数	NextSeq サンプル数
遺伝子発現プロファイル (発現アレイと同等の解析)	10M (1x75 bp)	40サンプル/ラン
スプライスバリエント含む 転写産物発現プロファイル (mRNA-Seq)	25M (2x75 bp)	16サンプル/ラン
ncRNAを含む 転写産物発現プロファイル (Total RNA-Seq)	50M (2x75bp)	8サンプル/ラン

2014/10/17

RNA Seqをはじめよう！

BaseSpace で行う簡単NGSデータ解析

イリミナ株式会社 バイオインフォマティクス サポート サイエンティスト  
突生川絵里

[+ 詳細](#)



# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめするアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# 全ゲノム解析 1ランで何サンプル流せるのか？

- Sequencing Coverage Calculator

[http://support.illumina.com/downloads/sequencing\\_coverage\\_calculator.html](http://support.illumina.com/downloads/sequencing_coverage_calculator.html)

ゲノムサイズ、必要なカバレッジ、ランの条件から、1フローセルあたりに流せるサンプル数が計算される。

Sequencing Coverage Calculator

Genome or region size (in million bases):  Mb

Coverage:  x

Total number of cycles (e.g. 200 for 2x100):  cycles

Instruments (you can select multiple instruments):

- MiSeq
- NextSeq 500
- HiSeq 3000/4000
- HiSeq 1500/2500 Rapid Run
- HiSeq 1500/2500 High Output
- HiSeq 1000/2000
- Genome Analyzer IIx
- HiScanSQ

Sequencing Coverage Calculator

Thank you for using the Illumina coverage calculator. The results below can also be [downloaded](#) in a comma-separated values (CSV) report.

	NextSeq 500	NextSeq 500
Run type	High Output kit	Mid Output kit
Reads	400,000,000 per flow cell	130,000,000 per flow cell
Genome or region size (bases)	390,000,000	390,000,000
Coverage	30x	30x
Total number of cycles (e.g. 200 for 2x100)	300	300
Maximum number of cycles	2x150	2x150
Total output required (bases)	11,700,000,000	11,700,000,000
Output per unit	120,000,000,000 per flow cell	39,000,000,000 per flow cell
Number of units	0.10 flow cells	0.30 flow cells
Number of samples	10.26/flow cell	3.33/flow cell



# 本日のOutline

- ▶ イルミナのシーケンサーとNextSeqの位置づけ
- ▶ NextSeq500でおすすめるアプリケーションのご紹介
  - エクソーム解析
  - RNA-Seq
  - 全ゲノム解析
- ▶ NextSeq500を使う

# プッシュボタンで操作する簡易性



# NextSeq 500

## ライブラリーからデータ解析まで



# Library Preparation (ライブラリーの調製・準備)

## ライブラリーの変性と希釈

試薬カートリッジの準備

Flow Cellの準備

試薬カートリッジへの  
ライブラリーの分注

NCSインターフェース  
からランのセットアップ

- ▶ 調製済みライブラリーをNaOHで変性、HT1で希釈し、最適なクラスター密度になるようにする
- ▶ 推奨のライブラリー使用濃度は1.8 pM
  - qPCRでの定量をお勧め
- ▶ (オプション) PhiXコントロールを2 pMに希釈して、希釈済みライブラリーに1% (体積比で) スパイクイン

# NextSeq 500 試薬カートリッジの準備

ライブラリーの変性と希釈

**試薬カートリッジの準備**

Flow Cellの準備

試薬カートリッジへの  
ライブラリーの分注

NCSインターフェース  
からランのセットアップ

- ▶ 試薬カートリッジを室温の水 (蒸留水 or MilliQ水) に60 ~ 90 分間浸し、中身を溶かす
- ▶ 使用前にカートリッジを5回転倒攪拌する



# Flow Cellの準備

ライブラリーの変性と希釈

試薬カートリッジの準備

**Flow Cellの準備**

試薬カートリッジへの  
ライブラリーの分注

NCSインターフェース  
からランのセットアップ

- ▶ 使用前のFlow Cellをアルミ包装に入った状態で室温に30分間置き、室温に戻す
- ▶ ガラス部分をレンズ紙で拭いて乾かす



# 試薬カートリッジへのライブラリーの分注

ライブラリーの変性と希釈

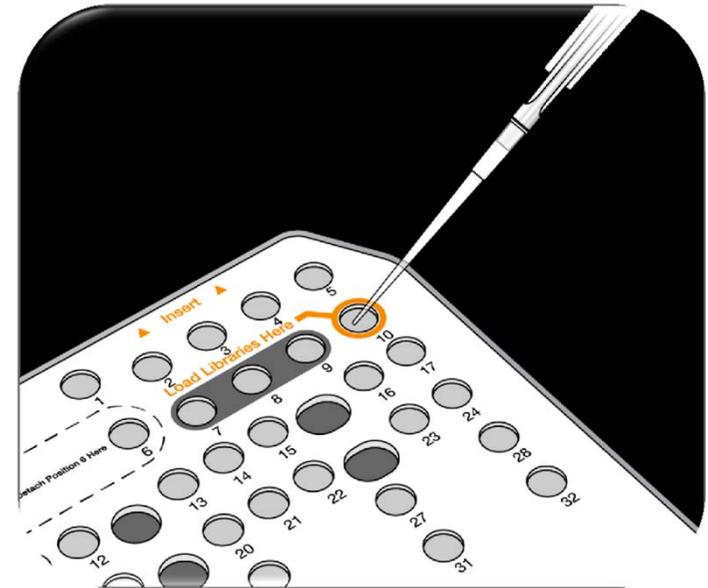
試薬カートリッジの準備

Flow Cellの準備

試薬カートリッジへの  
ライブラリーの分注

NCSインターフェース  
からランのセットアップ

- ▶ 希釈済みライブラリーを試薬カートリッジの**10番**ポートに分注する



# NCSインターフェイスからランのセットアップ

ライブラリーの変性と希釈

試薬カートリッジの準備

Flow Cellの準備

試薬カートリッジへの  
ライブラリーの分注

**NCSインターフェイス  
からランのセットアップ**

- ▶ ディスプレイに表示されるガイド画面をタッチパネル操作
- ▶ 大きなボタンで簡単な操作をナビゲート
- ▶ 数回タッチするだけでラン実行が可能

C52683 ✓

Load > **Run Setup** > Check > Sequence

Enter the run parameters. When complete, select Next.

<b>Run Name</b>	Run 123488			
<b>Library ID</b>	A1234			
<b>Recipe</b>	HiSeqFastQ			
<b>Read Type</b>	<input type="radio"/> Single Read		<input checked="" type="radio"/> Paired End	
<b>Read Length</b>	Read 1	Read 2	Index 1	Index 2
	151	151	0	0
<b>Custom Primers</b>	<input checked="" type="checkbox"/> Primer 1	<input type="checkbox"/> Primer 2	<input type="checkbox"/> Primer 3	

← Back    ✕ Exit    → Next

# NCSからのラン開始

ライブラリーの変性と希釈

試薬カートリッジの準備

Flow Cellの準備

試薬カートリッジへの  
ライブラリーの分注

**NCSインターフェース  
からランのセットアップ**

- ▶ プレランチェックで問題が無ければ **"Start"** をクリックしてラン開始

NS500309 cslabs@illumina.com

Log In > Load > Run Setup > **Check** > Sequence

The instrument is performing a self-check. When it is finished, select Start to begin sequencing.

✓ **System Checks** ^

- ✓ Doors Closed
- ✓ Consumables Loaded
- ✓ Required Software
- ✓ Instrument Disk Space
- ✓ Network Connection
- ✓ Network Disk Space

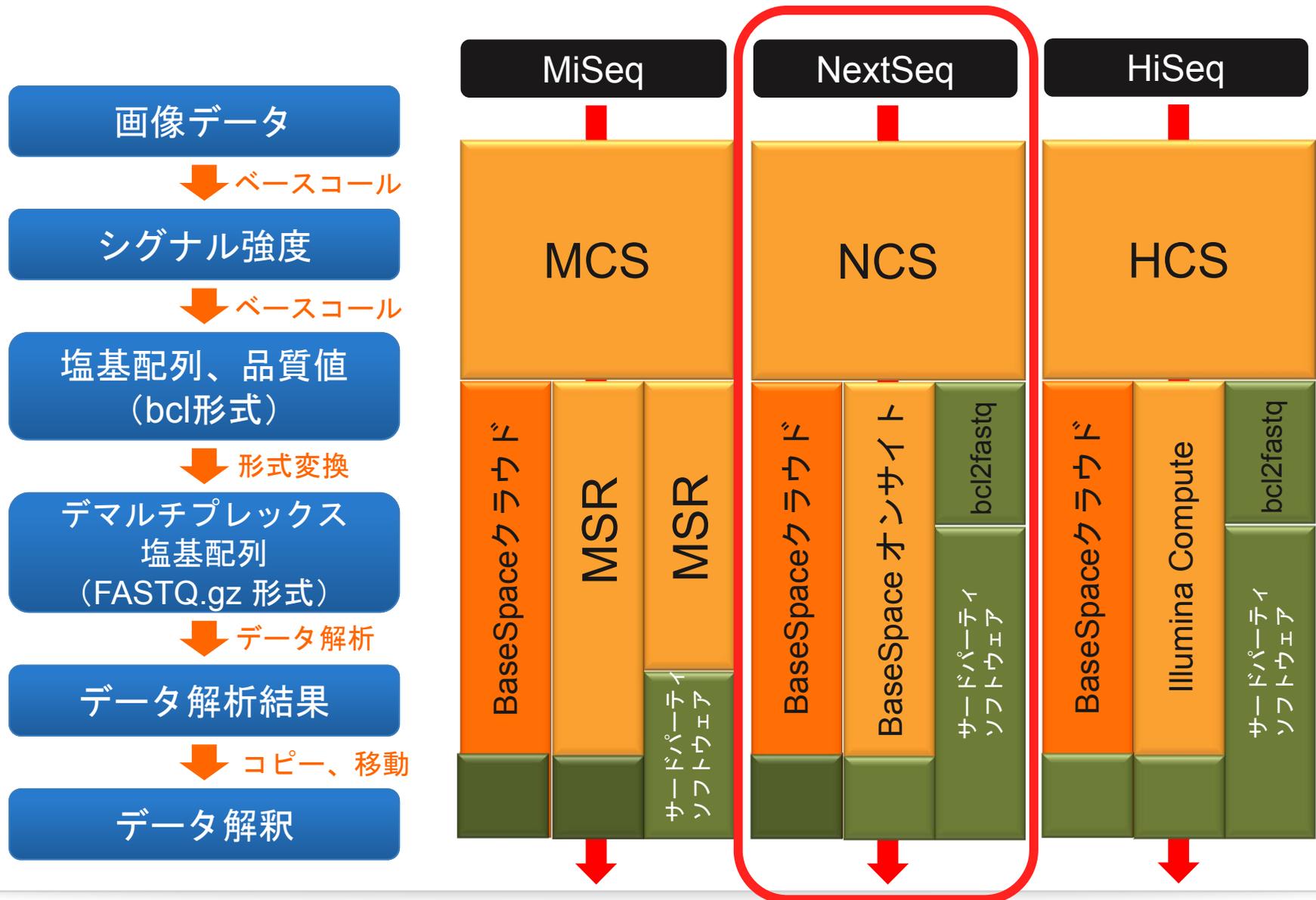
✓ **Temperature** v

⌚ **Imaging System** v

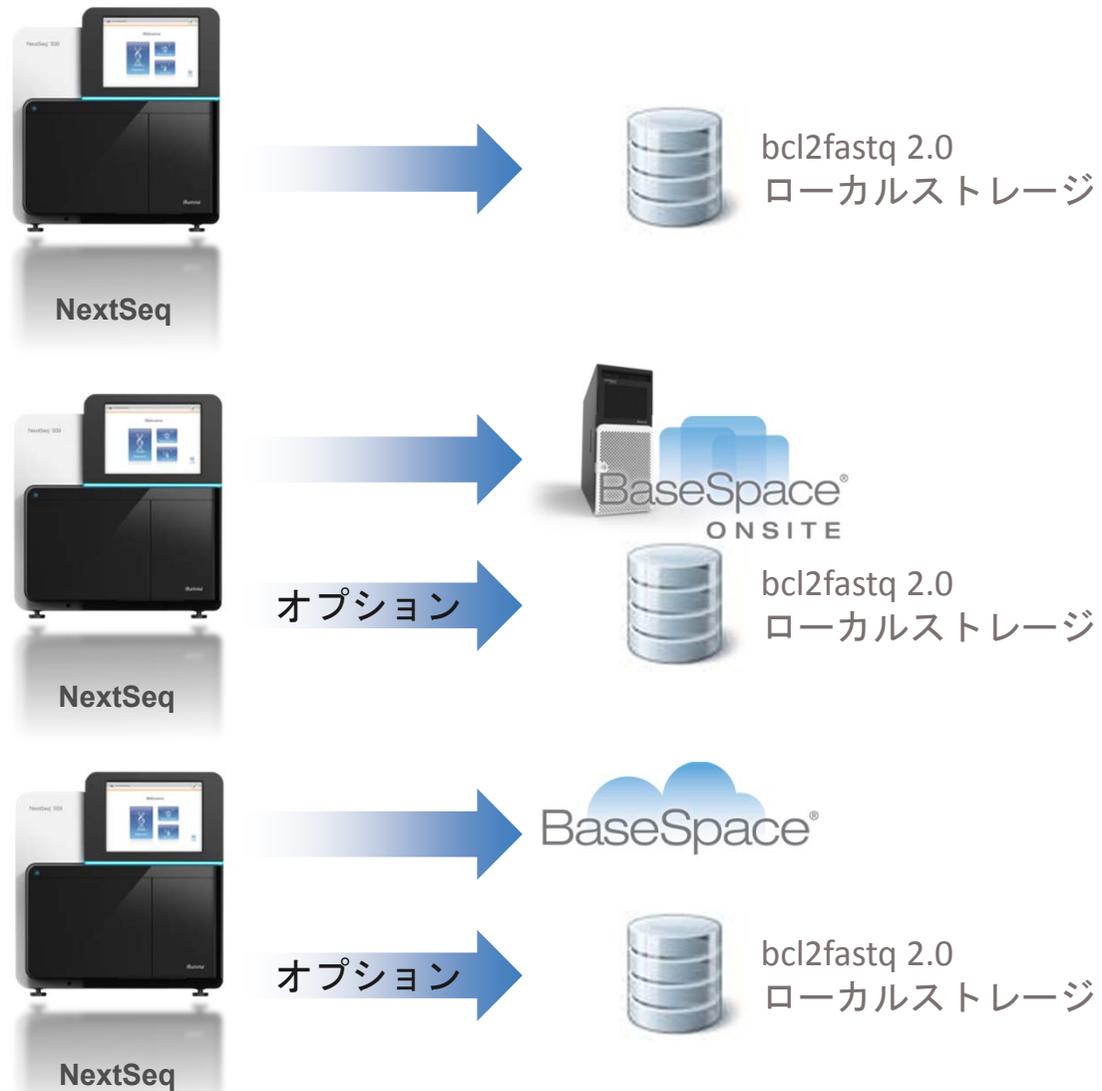
✓ **Reagent Delivery** v

Back Exit **Start**

# イルミナNGSデータ解析ワークフロー



# NextSeq データ解析 3つのオプション



## 参考/データベース

- ▶ NextSeq500の仕様等

<http://www.illumina.com/systems/nextseq-sequencer.html>

- ▶ 各アプリケーションのデータシート

<http://www.illumina.com/systems/nextseq-sequencer/literature.html>

- ▶ MiSeq/NextSeq500/HiSeqシリーズについての比較

<http://www.illumina.com/systems/sequencing-platform-comparison.html>

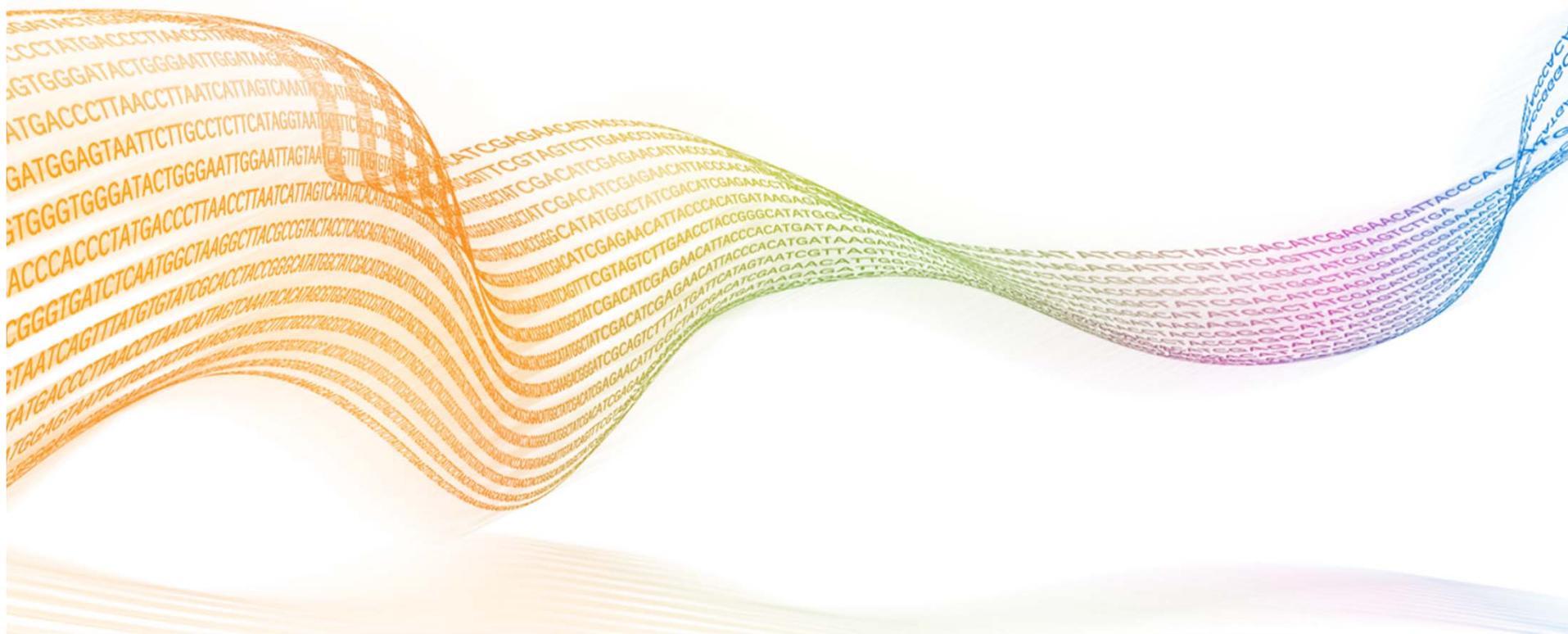
- ▶ デモデータ

BaseSpace上にアップされているデータをご覧いただけます。  
検索欄で”NextSeq”とご入力ください。

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

ご清聴ありがとうございました。

ご質問は[techsupport@illumina.com](mailto:techsupport@illumina.com)でも承ります。



# Questions?

