

Guide to Reviewing and Approving Custom Designs

Roche NimbleGen SeqCap Custom Designs




SeqCap 製品のChoiceあるいはDeveloper Library (Enrichment Kit) は、お客様の要望に従ってプローブデザインし製造するカスタム製品です。Roche NimbleGenデザイン担当者またはNimbleDesignソフトウェアから提案されたカスタムデザイン結果を確認し承認いただくプロセスが、全ての新規のご注文で必要です。ご承認いただくまでプローブの製造は開始されません。

このドキュメントではカスタムデザイン結果を確認・承認する方法について記載しています。ターゲットとして指定した領域が適切にプローブでカバーされ、実験の目的に合うデザインであるかをご確認ください。ご提案結果が適切ではない場合には、デザインの修正と確認のプロセスを繰り返します。




1. デザインの流れについて

1.1. デザインの流れと用語


お客様の指定された「ターゲット領域」は「**インプット領域**」として編集され、**primary_target.bed** が作成されます。

-  互いに重なり合う領域をターゲットとして指定した場合は、その重なり合いを無視します (コンソリデーション)。
-  SeqCap EZ と SeqCap Epi の場合、100bp以下のサイズのターゲットに対しては、キャプチャー効率を上げるために、その前後配列を含めた100bpに対してプローブの設計を行います (パディング)。SeqCap RNAの場合には、50bpが最少の長さとなるようパディングを行います。必要であればもう一度コンソリデーションを実施します。
-  条件によっては「ターゲット領域」と「インプット領域」は同一である場合があります。一方、最終的に全くプローブが設計できなかったフラグメントが除外されるため、インプット領域とprimary_target.bedは一致しない場合があります。

リファレンスシーケンスからプローブが設計され、そのうち「インプット領域」をカバーするプローブが選択されます。「**プローブカバー領域(capture_targets.bed)**」が作成されます。

-  効率的なキャプチャーのためにプローブはお互いに重なり合うように設計していますが、少なくとも1つのプローブでカバーされる領域を**プローブカバー領域**と定義します。
-  実験の目的により、「**プローブのユニーク性**」と、「**プローブカバレッジ**」のバランスを考慮する必要があります。このバランスは、以下のように大別され、それぞれ異なる特徴を持ちます。
 - **プローブのユニーク性を重視したデザイン**: 特異的なプローブのみを使用しています。プローブカバレッジは低くなる可能性があります、特異的なハイブリダイゼーションが期待できます。**unique_design**として提案される場合があります。
 - **プローブカバレッジを重視したデザイン**: プローブのユニーク性スレッシュホールドを下げたデザインです。プローブカバレッジが上がり、シーケンス目的領域全体に渡り取りこぼしなくデータを得ることができると期待できます。一方でデータ解析の際にはマッピングパラメーターの最適化が必要となるかもしれません。**relaxed_design**として提案される場合があります。
-  ターゲット領域以外の配列をキャプチャーする可能性を下げるために、低複雑性領域や高リピート領域のプローブの設計を避けるアルゴリズムを採用しています。

「プローブカバー領域」から「**シーケンシング予測領域**」が作成されます。

-  シーケンスキャプチャーでは、プローブとハイブリダイゼーションするゲノム断片の位置関係により、プローブカバー領域の外側もシーケンスされると予測されます。プローブカバー領域の外側が両端100bpずつシーケンスされると仮定してシーケンシング予測領域が算出されます。この仮定(両端100bpずつ)は、典型的なライブラリ長(約200bp)をIllumina社のペアエンドシーケンシングをする際には有効ですが、ライブラリ長が異なる場合やシングルエンドシーケンシングを実施する場合には妥当ではない可能性があります。

2. デザイン結果概要の確認

2.1. デザイン全体の数値的な確認

coverage_summary ファイル (TXT) を Microsoft Excel などを開き、ターゲット領域のカバー率や総サイズなどを確認します。

	A	B	C
1	Genome build	hg19	
2	Number of regions	4412	
3	Length of regions (bp)	1512482	
4			
5	Statistics	Probe_Coverage	Estimated_Coverage
6	Target Bases Covered	1430484	1484722
7	% Target Bases Covered	94.6	98.2
8	Targets with no coverage	54	51
9			
10	Target Bases Not Covered	81998	27760
11	Due to N's	0	0
12	Due to repeats	61498	27117
13	% Target Bases Not Covered	5.4	1.8
14	Due to N's	0	0
15	Due to repeats	4.1	1.8
16			
17	Total capture targets	5091	
18	Total capture space (bp)	1638969	

特にこれらの値を確認します

- **Genome build:** リファレンスシーケンス(データベース)
- **Number of regions:** インput領域数
- **Length of regions:** インput総塩基数

- **Probe_Coverage 列:** インput領域内のプローブカバー領域
- **Estimated_Coverage 列:** インput領域内のシーケンシング予測領域
 - ・ **Target bases covered:** インput領域内のカバーされる総塩基数 (表の F 列または H 列の総和)
 - ・ **Percent target bases covered:** インput領域内のカバーされる塩基の割合 (%)
 - ・ **Targets with no coverage:** インput領域内のカバーされない領域(フラグメント)数 (表の G 列または I 列が 0 であるフラグメントの数)

 - ・ **Target Bases Not Covered:** インput領域内のカバーされない塩基数 (表の J 列または M 列の総和)
 - ・ **Percent Target Bases Not Covered:** インput領域内のカバーされない塩基の割合 (%)
 - **Due to N's :** リファレンスシーケンスのターゲット領域内の “N” などの不明瞭な塩基による
 - **Due to repeats :** リファレンスシーケンスのターゲット領域内の低複雑性あるいは高リピート配列による

- **Total capture targets:** プローブでカバーされる総フラグメント数 (capture_target.bed ファイルの行数)
- **Total capture space (bp):** プローブでカバーされる総塩基数 (capture_target.bed ファイルから計算した総塩基数)。この数値はインput領域 (primary_target.bed) から大きく離れた値となる場合がありますので、検体あたりどのくらいのシーケンシングデータが必要となるかを試算するためにはこの値を利用します。

coverage ファイル (TXT または Zipped TXT) を Microsoft Excel などを開き、ターゲット領域のフラグメントごとのカバー率を確認します。


A	B	C	D	E	F	G	H	I
REGION_NAME	CHROMOSOME	START	STOP	LENGTH	BASES_PROBE_COVERAGE	FRAC_PROBE_COVERAGE	BASES_ESTIMATED_COVERAGE	FRAC_ESTIMATED_COVERAGE
chr1:45794977-45795109	chr1	45794977	45795109	132	132	1	132	1
chr1:45796187-45796229	chr1	45796158	45796258	100	100	1	100	1
chr1:45796853-45797006	chr1	45796853	45797006	153	153	1	153	1
chr1:45797091-45797228	chr1	45797091	45797228	137	137	1	137	1
chr1:45797332-45797521	chr1	45797332	45797521	189	189	1	189	1
chr7:6042083-6042267	chr7	6042083	6042267	184	68	0.37	184	1
chr7:6043320-6043423	chr7	6043320	6043423	103	70	0.68	103	1
chr7:6043602-6043689	chr7	6043596	6043696	100	73	0.73	100	1
chr7:6045522-6045662	chr7	6045522	6045662	140	76	0.543	140	1

必要であれば
フィルタリングや並べ替え
などを行います


J	K	L	M	N	O
BASES_W_NO_PROBE_COV	BASES_W_NO_PROBE_COV_DUE_TO_N	BASES_W_NO_PROBE_COV_DUE_TO_REPEATS	BASES_W_NO_EST_COV	BASES_W_NO_EST_COV_DUE_TO_N	BASES_W_NO_EST_COV_DUE_TO_REPEATS
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
116	0	0	0	0	0
33	0	0	0	0	0
27	0	0	0	0	0
64	0	0	0	0	0



- A 列 **REGION_NAME**: ターゲット領域のフラグメント名
- B 列 **CHROMOSOME**: 染色体番号 (リファレンスとする配列によりコンティグ名などで表記される場合があります)
- C 列 **START**: 開始位置
- D 列 **STOP**: 終了位置
- E 列 **LENGTH**: そのフラグメントの長さ
- F 列 **BASES_PROBE_COVERAGE**: そのフラグメント内のプローブでカバーされる塩基数
- G 列 **FRAC_PROBE_COVERAGE**: そのフラグメント内でのプローブでカバーされる塩基の割合
- H 列 **BASES_ESTIMATE_COVERAGE**: そのフラグメント内の、シーケンスされると予測される塩基数
- I 列 **FRAC_ESTIMATED_COVERAGE**: そのフラグメント内での、シーケンスされると予測される塩基の割合

- J 列 **BASES_W_NO_PROBE_COV**: そのフラグメントのプロープでカバーされない塩基数
- K 列 **BASES_W_NO_PROBE_COV_DUE_TO_N**: リファレンスシーケンスのターゲット領域内の
“N” などの不明瞭な塩基の存在によりプロープでカバーされない塩基数
- L 列 **BASES_W_NO_PROBE_COV_DUE_TO_REPEATS**: リファレンスシーケンスのターゲット領域内の
低複雑性あるいは高リピート配列によりプロープでカバーされない塩基数
- M 列 **BASES_W_NO_EST_COV**: そのフラグメントの、シーケンスされるとは予測されない塩基数
- N 列 **BASES_W_NO_EST_COV_DUE_TO_N**: リファレンスシーケンスのターゲット領域内の
“N” などの不明瞭な塩基の存在によりシーケンスされるとは予測されない塩基数
- O 列 **BASES_W_NO_EST_COV_DUE_TO_REPEATS**: リファレンスシーケンスのターゲット領域内の
低複雑性あるいは高リピート配列によりシーケンスされるとは予測されない塩基数

 条件によりJ列からO列のデータは作成されない場合があります。

(オプション) Design Summary File (PDF) または Probe Set Report (PDF) を PDF viewer で開き、プローブ設計条件や結果概要を確認します。

 条件によりこのファイルは提供されない場合がありますが、coverage_summaryファイル (TXT)で同等の情報を得ることができます。

Probe Set Report: NewEZDesign_P1

User Information
Name: _____ Email: _____

Design Information
Design Name: NewEZDesign
Organism: Homo sapiens
Final Design Name: 140328_H019_NewEZDesign_EZ_HXG
Genome Build: HG19 NCBI Build 37.1/GRCh37
Application: SeqCap EZ Choice
Description: Optional Design Description

Format
Format: SeqCap EZ Choice

Consolidation Report - Overlapping Regions
Initial Regions Count: 4,412
Final Regions Count: 4,412
Initial Sequence Total: 1,512,451
Final Sequence Total: 1,512,482

Probe Set
Probe Set Name: NewEZDesign_P1
Description: Optional ProbeSet Description

Preferred Close Matches: 3
Maximum Close Matches: 20

Statistics	Probe Coverage	Estimated Coverage
Target Bases Covered:	1,430,484	1,488,722
% Target Bases Covered:	94.6	98.2
Target Bases Not Covered:	81,998	27,760
Due to Ns	0	0
Due to Repeats	61,498	27,117
% Target Bases Not Covered:	5.4	1.8
Due to Ns	0	0
Due to Repeats	4.1	1.8

Related Deliverables
Coverage Summary: NewEZDesign_P1_coverage_summary.txt
Region by Region Coverage: NewEZDesign_P1_coverage.zip
Primary Targets BED: NewEZDesign_P1_primary_targets.zip
Capture Targets BED: NewEZDesign_P1_capture_targets.zip
Predicted Uncovered Targets BED: NewEZDesign_P1_predicted_uncovered_targets.zip

Note: Blank fields indicate the software does not have a value to report; refer to www.nimblegen.com/guides for information on reviewing and approving your design.

For life science research only. Not for use in diagnostic procedures. NIMBLEGEN, NIMBLEDESIGN and SEQCAP are trademarks of Roche. All other product names and trademarks are the property of their respective owners. © 2014 Roche NimbleGen, Inc. All rights reserved.

2.2. 個々の領域についての視覚的な確認方法

2.2.1. BED ファイルについて

BED (.bed) ファイルは、UCSC ゲノムブラウザや Broad Institute の IGV (Integrative Genomics Viewer) などの様々なゲノムブラウザや、Roche NimbleGen の提供する SignalMap ソフトウェアで開くことができます。これらのソフトウェアで開くことで、ターゲット領域やプローブ設計領域を視覚的に確認することができます。このドキュメントでは、2.2.2 章で UCSC ゲノムブラウザ、2.2.3 章で SignalMap ソフトウェアの使い方を概説しています。

BED ファイルはタブ区切り形式のテキストファイルであり、位置情報を BED フォーマットで(ヘッダなしの 3 列または 4 列で)保存しています。BED フォーマットについては <http://genome.ucsc.edu/FAQ/FAQformat.html#format1> をご参照ください。

- **capture_targets.bed:**

プローブカバー領域。各領域は、少なくとも 1 つ以上のプローブでカバーされています。

- **primary_targets.bed:**

インプット領域(お客様が指定したターゲット領域に、パディングおよびコンソリデーションの処理を施した領域)から、少なくとも 1 つ以上のプローブが存在する領域のみに編集されています。

- **predicted_no_coverage_regions.bed (predicted_uncovered_targets.bed):**

シーケンスされないと予測される領域。プローブの前後 100bp に入らない primary_targets.bed 上の塩基位置を示しています。



条件により、1つのファイルに複数のトラック情報が保存されたファイルが提供される場合があります。また、predicted_no_coverage_regions.bed ファイルは提供されない場合があります。

2.2.2. UCSC ゲノムブラウザを利用した視覚的な確認方法

<http://genome.ucsc.edu> を開き、**Genome Browser** リンクをクリックします。ゲノムブラウザゲートウェイページが開きますので、*group*、*genome* と *assembly* リストボックスから、デザインをご依頼いただいた生物種とそのデータベースのバージョンを選択し、**add custom tracks** ボタンをクリックします。



ゲートウェイページで **manage custom tracks** ボタンが表示されている場合には、**Click here to reset** をクリックすると **add custom tracks** ボタンが表示されます。

カスタムトラックの追加設定ページで、**Browse** ボタンから BED ファイルを選択して **Submit** ボタンをクリックし、BED ファイルをアップロードします。カスタムトラックの設定ページで、**go to genome browser** ボタンをクリックすると下記の様な画面が表示されます。

- 🔍 ヘッダが残ったBEDファイルはアップロードエラーとなる場合があります。
- 🔍 複数のBEDファイルをアップロードする場合、トラック名「User Track」を適宜編集してください。

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr5:131,921,849-131,951,081 29,233 bp enter position, gene symbol or search terms go

chr5 (q31.1) hg19

Scale chr5: 131,938,000 131,948,000 131,958,000 10 kb

primary target User Supplied Track

capture target User Supplied Track

RefSeq Genes

Common SNPs (142)

track search default tracks default order hide all manage custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

Custom Tracks refresh

Mapping and Sequencing refresh

Genes and Gene Predictions refresh

アノテーション情報で検索し、特定の領域を確認します

2つのBEDファイルをアップロードしています

UCSCゲノムブラウザの機能を利用してデザインの確認を行います。ポジティブコントロール領域や必要な領域がカバーでされていない場合には、パラメーターを変更して再デザインを行う必要があります。

- **特定のゲノム位置を表示:**

上部テキストボックスにゲノム位置やキーワード等を入力して**go**ボタンをクリックします。

- **その他のアノテーション情報の表示・非表示:**

ページ下部に表示されるそれぞれのアノテーション情報のリストボックスでそれぞれの項目の表示・非表示を選択して、**refresh**ボタンをクリックします。

- **拡大縮小:**

zoom in と**zoom out** ボタンで1.5-, 3-, 10-倍に拡大・縮小表示できます。

- **表示位置の移動:**

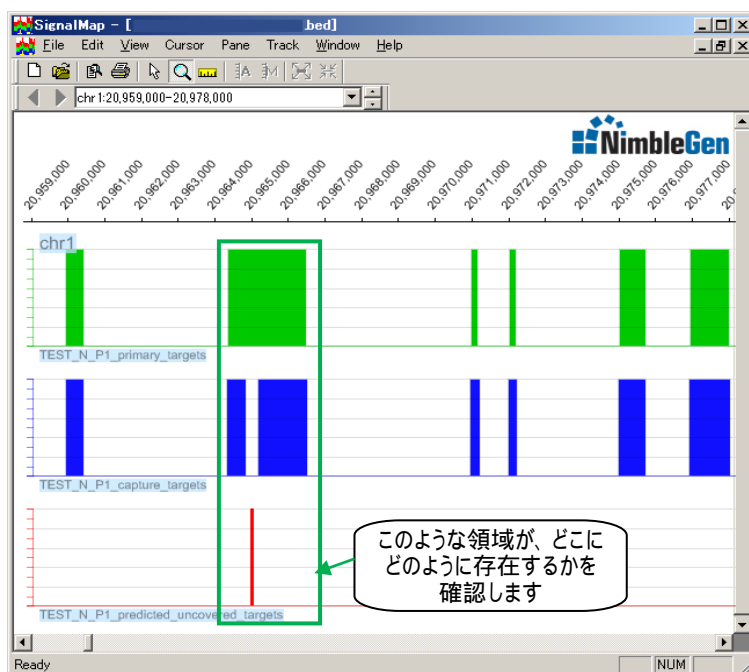
move ボタンをクリックして左右に画面をスクロールします。

- primary_targetトラックでは一続きのボックスとして表示されるにも関わらずcapture_targetトラックでは間が空いて見える領域(ギャップ)および、predicted_no_coverage_regionsにボックスとして表示される領域に注目してください。

2.2.3. SignalMap ソフトウェアを利用した視覚的な確認方法

SignalMap ソフトウェアを起動し、**File -> New** からいずれかの BED ファイルを開きます。複数の BED ファイルを一つの画面で表示したい場合には **File -> Import** から残りの BED ファイルを開きます。

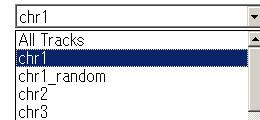
- SignalMapソフトウェアは <http://www.nimblegen.com/signalmap> から無料でダウンロードできます。



ソフトウェアの下記のような機能を使用してデザインを確認します。primary_target トラックでは一続きのボックスとして表示されるにも関わらず capture_target トラックでは間が空いて見える領域(ギャップ)および、predicted_no_coverage_regions にボックスとして表示される領域に注目してください。ポジティブコントロール領域や必要な領域がカバーでされていない場合には、パラメーターを変更して再デザインを行う必要があります。ソフトウェアの使用法の詳細は SignalMap User's Guide をご参照ください。

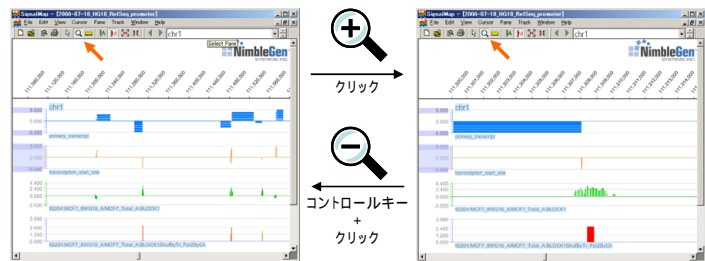
■ 染色体(領域・区画)表示の切り替え:

SignalMapの横軸(x軸)は染色体(領域・区画)単位での位置情報を表示しています。ツールバーのpane selector fieldから特定の染色体(領域)表示に切り替えてください。



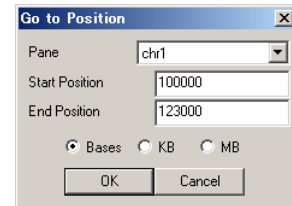
■ ズームイン、ズームアウト:

ツールバーの虫眼鏡ボタンをクリックしてズームモードにし、ウインドウ内の任意の位置でクリックするか、拡大したいゲノム領域をドラッグ選択します。ズームアウトするには、ズームモード時にCtrlキーを押して任意の位置をクリックしてください(ポイントの虫眼鏡が+表示 から - 表示 に切り替わります)。



■ 特定のゲノム位置の表示:

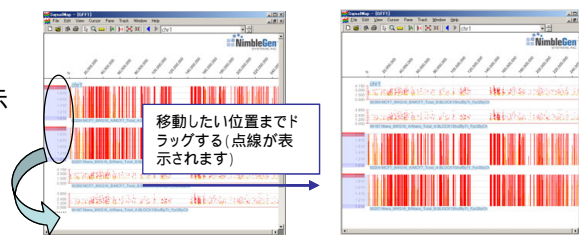
View -> Go to position を選択して、染色体番号、開始位置、終了位置を入力してください。



🔍 アノテーションファイル (BED形式またはGFF形式) を開いておくと、Edit -> Searchで、アクセッション番号などの任意のキーワードで遺伝子等を検索することができます。

■ トラック表示位置アレンジ:

移動させたいトラックを選択し、ドラッグしたままカーソルを移動させたい位置に移動させます。カーソル表示が灰色の点線表示になった位置でドラッグを解消するとトラックが移動します。







■ カーソルの固定:

フラグメント単位で表示を切り替えたい場合にはこの機能を利用すると便利です。基本となるトラックを選択し、右クリックから**Cursor -> Attach Cursor**を設定します。ゲノム上の次の領域へはキーボードの矢印キー[]を押すことで移動できます。設定を解消するには、**Cursor -> Detach Cursor**を設定します。

3. デザイン結果の承認


「自分の狙った領域がどのくらいプローブでカバーされているか」や「絶対に読みたいという(ポジコンなどの)一部の特定の領域がある場合、その領域がきちんとプローブでカバーされているか」などをチェックした結果、満足の得られるデザインではなかった場合には、下記のように対応してください。

- ◇ NimbleDesign を利用の場合： パラメーターを再設定してもう一度デザインを作成してください。
- ◇ NimbleGen デザイン担当者によるデザインの場合： その旨を tokyo.seqsol@roche.com までご連絡ください。

-  ご承認いただけるまで、デザインの修正と確認のプロセスを繰り返します。
-  パラメーターを変更してもデザイン結果に差が出ない場合がございます。
-  ご要望内容によってはNimbleDesignでは対応できない場合がございますので、NimbleGenデザイン担当者によるデザインに変更してください。
-  通常のRoche NimbleGenのデザインでは、ゲノム中に2～5回出現するような低コピーリピート領域に対してもプローブは設計されません。そのような領域もキャプチャー&シーケンスする必要がある場合には、その旨をお知らせいただくと、厳密性を下げた基準で再設計を行います。しかし、そのような場合には、プローブのカバー率は上がりますが、ターゲット以外のリードが多数出現してしまうためキャプチャー効率は低下すると予想されます。

提案されたデザインで問題が無い(提案された複数のデザインから1つのデザインを選択する)場合には、下記のように対応してください。

- ◇ NimbleDesign を利用の場合： デザインを Approve し、Transmit をクリックしてください。同時に、代理店様へご注文ください。
- ◇ NimbleGen デザイン担当者によるデザインの場合： その旨を tokyo.seqsol@roche.com までご連絡ください。

-  この最終のご連絡が無い場合には、実際のプローブ製造プロセスには進めませんのでご注意ください。



ロシュ・ダイアグノスティクス株式会社

〒105-0014 東京都港区芝 2-6-1

TEL: 03-5443-5287