



NimbleGen シークエンスキャプチャー

新規カスタムデザイン依頼用 ターゲット領域抽出ガイド

Roche NimbleGenでは、2種類のキャプチャー法、液相反応試薬であるSeqCap EZ Libraryと、アレイベースのSequence Capture Arrayを提供しています。どちらの方法を用いる場合でも、お客様ご自身の研究対象のゲノム領域(カスタムターゲット領域)を抽出する製品を製造するようご依頼いただくことが可能です。

カスタムターゲット領域のシーケンスキャプチャーを行う場合には、Roche NimbleGenの専任の担当者が確実に均一的なキャプチャーを実現するように最適なアルゴリズムを用いて最大210万種類のプローブを利用したカスタムデザインをご提案しています。

カスタムターゲット領域の特定方法例:

★ 遺伝子リストからターゲット領域を特定する

- ・ 特定の遺伝子のエクソン
- ・ 特定の遺伝子(エクソン+イントロン)
- ・ 特定の遺伝子のプロモーター領域

★ ゲノム位置情報からターゲットを特定する

- ・ 他の実験により特定されたゲノム位置の周辺領域
- ・ 既知のSNP位置の周辺領域
- ・ 疾患関連領域など比較的長いゲノム領域

★ 個々の特定ができない

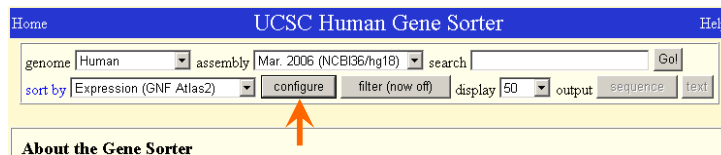
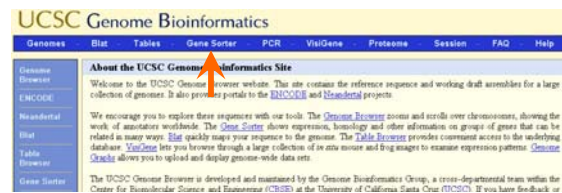
- ・ 既知の配列全て(exome や whole EST)
- ・ 特定の染色体全体

デザイン 1. 遺伝子リストからターゲット領域を特定する

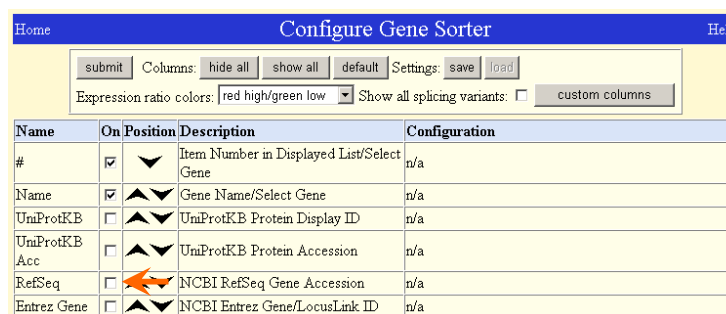
① 遺伝子リストをRefSeqのアクセッション番号で作成する

遺伝子リストは文献情報や他の実験によってスクリーニングされたものなど様々な情報元から作成することができますが、RefSeqのアクセッション番号が特定されている必要があります。

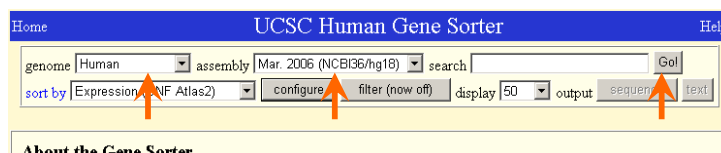
RefSeqのアクセッション番号を検索するには <http://genome.ucsc.edu> にアクセスし、UCSC Genome Browserページを開いてください。まず、このページの上にあるメニューリスト(または画面左のフレームメニュー)から **Gene Sorter** をクリックしてください。開いたGene Sorter ページでは、検索結果に RefSeq を表示するよう設定するため、初めに **configure** ボタンをクリックします。



configure ボタンをクリックして表示される画面で #, Name, RefSeq, Description の”On” チェックボックスにチェックした状態で **submit** ボタンをクリックしてください。クリックするとGene Sorter ページに戻ります。



例えば ヒト BDNF 遺伝子を検索してみます。生物種とデータベースBuildをプルダウンメニューからそれぞれ選択し、*search* フィールドに“BDNF”とタイプして **Go!** ボタンをクリックしてください(検索ワードは遺伝子名以外に、アクセッション番号やその他のタームが利用できます。*sort by* フィールドは“Protein Homology – BLASTP”に設定すると検索結果をより絞り込むことができます)。



複数の遺伝子が検索された場合には、対象の遺伝子のリンクをクリックしてください。

Simple Search Results

Known Gene Names

[BDNF](#) - brain-derived neurotrophic factor isoform a
[BDNFOS](#) - Homo sapiens non-coding transcript.BT2C (BDNF) mRNA, complete sequence; alternatively spliced.
[BDNFOS](#) - Homo sapiens trapped 3' terminal exon, clone B2F11.

Known Gene Descriptions

[BDNFOS](#) - Homo sapiens non-coding transcript.BT2C (BDNF) mRNA, complete sequence; alternatively spliced.
[NTRK2](#) - neurotrophic tyrosine kinase, receptor, type 2

RefSeq番号をコピーして、エクセルファイルなどにペーストし、タブ区切りテキスト形式で保存してください。RefSeq番号のみのリストとして保存しておく、ゲノム位置の検索の際にインプット情報をコピー&ペーストではなくファイルのアップロードで実施することが可能になります。遺伝子名などの他の情報をメモしておいた場合にはファイルのアップロードではなくRefSeq番号のセルをコピーして検索を行うことになります。

#	Name	RefSeq	Genome Position	Description
1	BDNF	NM_170734	chr11 27,655,387	brain-derived neurotrophic factor isoform c
2	NTF3	NM_01102654	chr12 5,443,133	neurotrophin 3 isoform 1 preproprotein
3	NTF4	NM_006179	chr19 54,257,572	neurotrophin 5 preproprotein
4	NGF	NM_002506	chr1 115,656,220	nerve growth factor, beta polypeptide precursor

同様の操作で全てのターゲット遺伝子のRefSeq番号リストを検索してください。

② RefSeqのアクセッション番号リストからゲノム位置を抽出する

UCSC Genome Browser の上部メニューバーの **Tables** または左側のフレームメニューの **Table Browser** をクリックするとTable Browser 画面が開きます。UCSC browser はユーザー設定を記憶するよう設計されていますので、まずTable Browser 下部の [click here](#) リンクをクリックして全てのユーザー設定をデフォルト設定に戻してください (1)。

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade: genome: assembly:

group: track:

table:

region: genome ENCODE Pilot regions position:

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to [Galaxy](#) [GREAT](#)

output file: (leave blank to keep output in browser)

file type returned: plain text gzip compressed

To reset all user cart settings (including custom tracks), [click here](#).

(2) *genome* フィールドに “Human” を、*assembly* フィールドに “Mar. 2006” を選択し、**output format** フィールドのドロップダウンメニューから “BED - browser extensible data” を選択してください。

(3) **identifiers (names/accessions)** フィールドの **paste list** ボタンで遺伝子リストをコピー & ペーストするか、**upload list** ボタンで遺伝子リストをアップロードするかで遺伝子リストを読み込ませてください。

(4) Table Browser 画面に戻りますので、**get output** ボタンをクリックしてください。その他のパラメータを変更する必要はありません。Output knownGene as BED ページが開きます。

目的に応じて *Create one BED record per* 項のオプションを変更して **get BED** ボタンをクリックしてください。

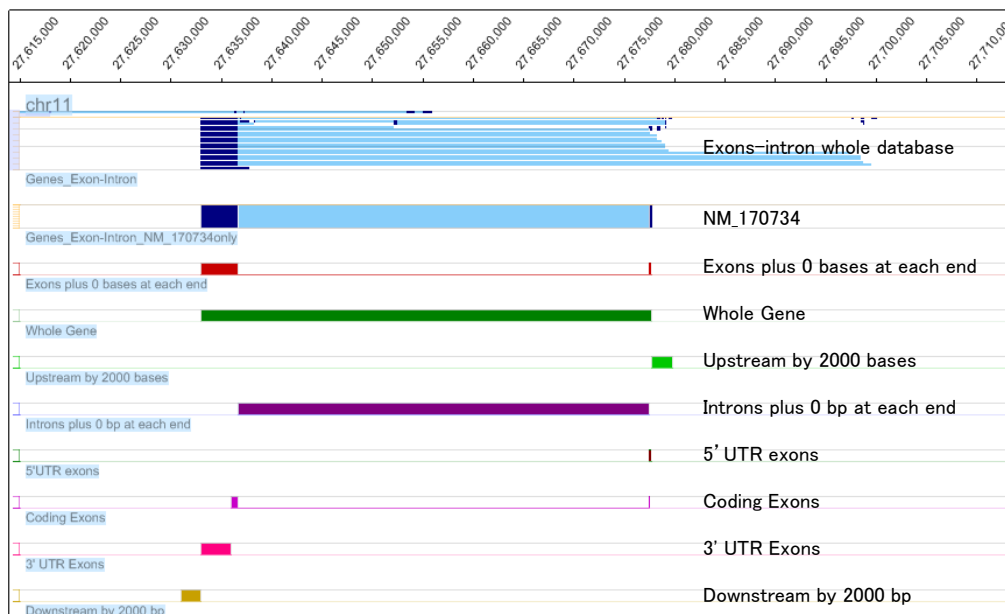
1. 特定の遺伝子エクソン
 - = “Exons plus 0 bases at each end” (オープンボックスに「0」を入力してください)
 - = “Coding Exons”
 - など
2. 特定の遺伝子(エクソン+イントロン)
 - = “Whole Gene”
3. 特定の遺伝子のプロモーター領域
 - = “Upstream by XXX bases” (オープンボックスには適当な数値を入力してください)

🔍 例えば BDNF (NM_170734) を検索した場合には下記のゲノム領域がそれぞれ検索されま

す。

“Exons plus 0 bases at each end”	“Coding Exons”
chr11 27633017 27636708	chr11 27635943 27636708
chr11 27677504 27677756	chr11 27677504 27677528
“Whole Gene”	“3' UTR Exons”
chr11 27633017 27677756	chr11 27633017 27635943
“Upstream by 2000 bases”	“Downstream by 2000 bp”
chr11 27677756 27679756	chr11 27631017 27633017
“Introns plus 0 bp at each end”	
chr11 27636708 27677504	
“5' UTR exons”	
chr11 27677528 27677756	

それぞれの位置関係は下図のとおりです。Exons-intron whole databaseトラックでは、紺色がエクソンを、水色がイントロンを示しています。また、BDNF 遺伝子はリバース鎖にコードされていますので、図の右側がこの遺伝子の上流となります。



検索結果を Microsoft エクセルなどにコピーし、染色体番号、開始位置、終了位置のみをリスト化してタブ区切りテキスト形式で保存してください。

Chromosome	START	END
chr11	27,633,017	27,636,708
chr11	27,677,504	27,677,756

終了位置から開始位置を引いた各領域の長さを積算し、総ターゲットサイズをご確認ください。

Chromosome	START	END	=END position - START position
chr11	27,633,017	27,636,708	3,691
chr11	27,677,504	27,677,756	252
			3,943 = SUM(total length)

カスタムデザイン依頼ファイルとして Roche NimbleGen にご送付ください。

デザイン 2. ゲノム位置情報からターゲットを特定する

ゲノム位置情報の基本は染色体番号、開始位置、終了位置ですが、様々なフォーマットで記述されます。このうちGFF (General Feature Format) ファイルはゲノムデータを記載する際の標準的なファイルで、Roche NimbleGenで取り扱っているSignalMapソフトウェアやDEVA Softwareで開くことができるファイルです。NimbleGen マイクロアレイ解析では、CGH解析ではsegMNTファイル、ChIPアレイ解析やメチレーションアレイ解析ではピークファイルなどがGFF形式で作成されます。こうした解析結果の位置情報からシーケンスをしたいターゲット領域を抽出するには、ファイルの構成を理解する必要があります。一方で、BEDファイルはUCSCゲノムブラウザで位置情報を確認することができるファイル形式で、A列、染色体番号、B列、開始位置、C列、終了位置 がそれぞれ記載されています。

🕒 NimbleGenのwebページからダウンロードしたヒト hg18、hg19 を元に作成されたアノテーションファイルの位置情報を利用することも可能です。

- Genes
- Genes_Exon-Intron
- Transcription_Start_Sites
- Structural_Variants
- 42M_CNV_Regions
- NimbleGen_CNV_Regions
- Segmental_Duplications
- Cytogenetic_Ideogram
- miRNA

GFFファイルはタブ区切り形式のテキストファイルですのでMicrosoft エクセルなどで開くことができます(ソフトウェアのバージョンによって全行開かない場合があります)。各列には下記の情報が記載されていますので、A列、D列、E列がゲノム位置情報となります。

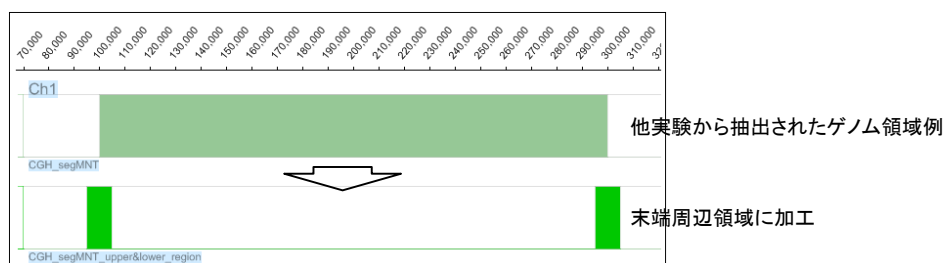
列	内容
A	<seqname> 染色体番号(領域・区画)名
B	<source> データ情報元
C	<feature> フィーチャー名。SignalMap Softwareではこの列の名前がトラック名となります。
D	<start> このデータポイント値のゲノム上のスタート位置
E	<end> このデータポイント値の、ゲノム上のエンド位置
F	<score> この行のデータポイント値
G	<strand> ストランド情報。指定しない場合には「.」を記入します。
H	<frame> フレーム情報。指定しない場合には「.」を記入します。
I	[attributes] [comment] SignalMapでの表示の色調指定や、アノテーション情報等

例えばこうした位置情報をそのままシーケンスキャプチャー用のターゲット領域として指定すると領域サイズが広すぎる場合があります。その場合にはD列、E列を加工することで特定位置の周辺領域のみに限定することができます。

例) 100,000から300,000までのセグメントの両末端周辺のみをターゲットとしたい場合には、そのセグメン

トの開始位置と終了位置に周辺領域サイズを加算・減算します(この例では両末端から5kbずつの10kbずつをそれぞれ抽出しています)。

	A	B	C	D	E	F	G	H	I
1	Ch1	NGS	CGH_segMNT	100000	300000	1	.	.	
2									
3									
4				=START-5000	=START+5000				
5	Ch1	NGS	CGH_segMNT_upper	95000	105000	1	.	.	
6									
7				=END-5000	=END+5000				
8	Ch1	NGS	CGH_segMNT_lower	295000	305000	1	.	.	



切り出したり加工したりした後の位置情報をまとめ、染色体番号、開始位置、終了位置のリストをタブ区切りテキスト形式で保存してください。

chromosome	START	END
Chr1	95,000	105,000
Chr1	295,000	305,000

必要に応じて終了位置から開始位置を引いた各領域の長さを積算し、総ターゲットサイズをご確認ください。

Chromosome	START	END	=END_position - START_position
chr11	27,633,017	27,636,708	3,691
chr11	27,677,504	27,677,756	252
			3,943 = SUM(total length)

カスタムデザイン依頼ファイルとして Roche NimbleGen にご送付ください。

デザイン 3. 個々の特定ができない

既知の配列全て(exome や whole EST)や特定の染色体全体をシーケンスしたい場合には、データベース(Build)と、概要情報をお知らせください。また、ヒトエクソームをシーケンスする場合には設計済み製品である NimbleGen SeqCap EZ Human Exome Library v2.0 をご利用いただくことができます。

お手持ちのシーケンス情報からシーケンスキャプチャー用プローブを設計したい場合には、FASTA 形式でシーケンス情報をお送りください。

カスタムターゲット領域抽出用シーケンスキャプチャーアレイ、SeqCap EZ Library 製品ラインナップ

製品番号	製品名	生物種、ターゲットサイズ
6266282	SeqCap EZ Choice Library, 12 Reaction	ヒト、7Mbまで
6266304	SeqCap EZ Choice Library, 24 Reaction	ヒト、7Mbまで
6266312	SeqCap EZ Choice Library, 48 Reaction	ヒト、7Mbまで
6266363	SeqCap EZ Choice XL Library, 12 Reaction	ヒト、50Mbまで
6266371	SeqCap EZ Choice XL Library, 24 Reaction	ヒト、50Mbまで
6266380	SeqCap EZ Choice XL Library, 48 Reaction	ヒト、50Mbまで
6471684	SeqCap EZ Developer Library, 12 Reaction	ヒト以外、50Mbまで
6471706	SeqCap EZ Developer Library, 24 Reaction	ヒト以外、50Mbまで
6471714	SeqCap EZ Developer Library, 48 Reaction	ヒト以外、50Mbまで
5394538	Seq Cap Custom 385K Arr Del	ヒト、5Mbまで
5478731	Titanium Opt Seq Cap 385K Arr Del	ヒト、5Mbまで
5329841	Seq Cap Custom 2.1M Arr Del	ヒト、30Mbまで
5394589	Seq Cap Developer 385K Arr Del	ヒト以外、5Mbまで
5451965	Seq Cap Developer 2.1M Arr Del	ヒト以外、30Mbまで



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
AS 事業部
TEL: 03-5443-5287 FAX: 03-5443-7098
<http://www.nimblegen.com>
E-mail: tokyo.arraygroup@roche.com