

# DesignStudioを用いたプローブデザインの方法と最適化のヒント

January 22, 2016



米田 瑞穂  
イリミナ株式会社  
テクニカルアプリケーションサイエンティスト

© 2015 Illumina, Inc. All rights reserved.  
Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPPro, CytoChip, DesignStudio, Epicentre, ForenSeq, Genetic Energy, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiSeq, MiSeqDX, MiSeq FGx, NeoPrep, NextBio, Nextera, NextSeq, Powered by Illumina, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the US and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

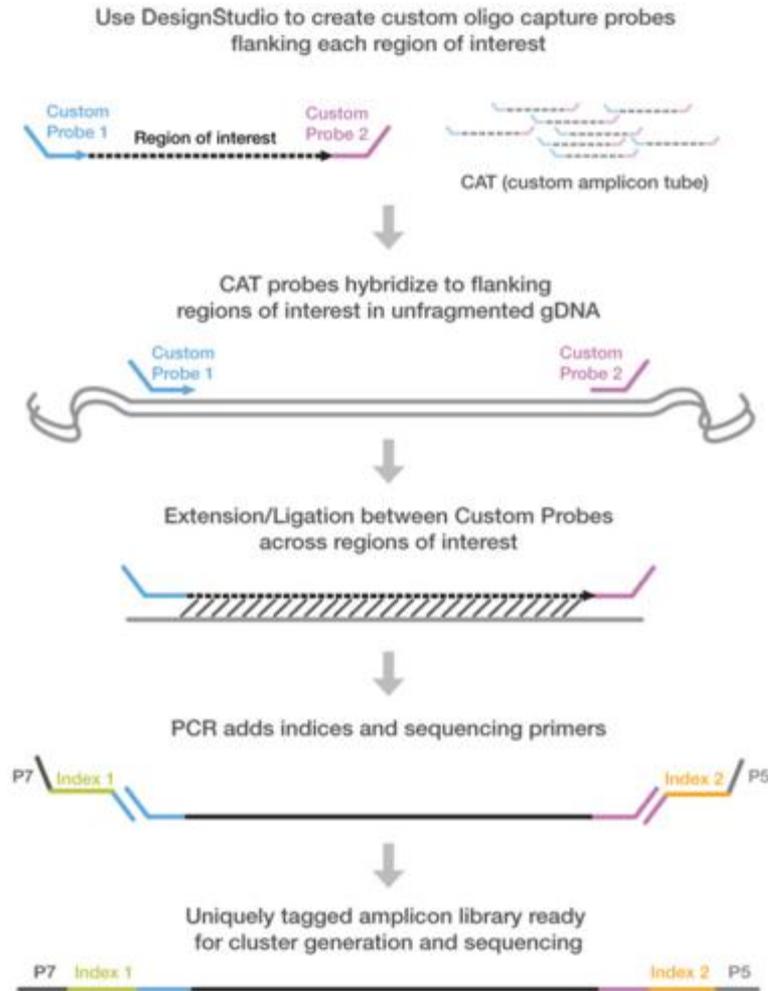
illumina®

# 本日のOutline

- ▶ TruSeq Custom Amplicon (TSCA)とDesignStudioの概要
- ▶ デザインのアルゴリズム
- ▶ デザイン改善へのヒント



# TruSeq Custom Amplicon Workflow



# TruSeq Custom Amplicon (TSCA) v1.5およびlow input

| テクノロジー             | TRUSEQ CUSTOM AMPLICON LOW INPUT | TRUSEQ CUSTOM AMPLICON V1.5 |
|--------------------|----------------------------------|-----------------------------|
| アンプリコンプレックス数(1プール) | ~1,536                           | ~10,000(コンシェルジュサービス利用時)     |
| 必要なインプット量          | わずか 10ng                         | 50ng                        |
| FFPEサンプル           | 対応 - 10 ng~(FFPE DNAの品質に依存)      | 対応 - 250 ng                 |
| アンプリコンサイズ          | 150, 175, 250 bp                 | 150, 175, 250, 425 bp       |
| 装置                 | MiSeq/NextSeq                    | MiSeq/NextSeq               |
| リード長               | ~150 bp x 2                      | ~300 bp x 2                 |
| 時間                 | 6.5 時間                           | 10 時間                       |
| キットサイズ(サンプル数)      | 16, 96~                          | 96~                         |

<http://jp.illumina.com/products/truseq-custom-amplicon-low-input.html>

# TruSeq Custom Amplicon Design Workflow



Assay Type      Configure Design      Add Targets      Submit Design      Review Design

1 ..... 2 ..... 3 ..... 4 ..... 5

### Configure Design

Assay Version ?

Comparison Chart

Variant source ?

Amplicon size ?

**TruSeq Custom Amplicon Low Input:**  
Used for low input and/or FFPE applications. Not to be used with applications requiring 425bp.

Design dual pools ?

**Population:**

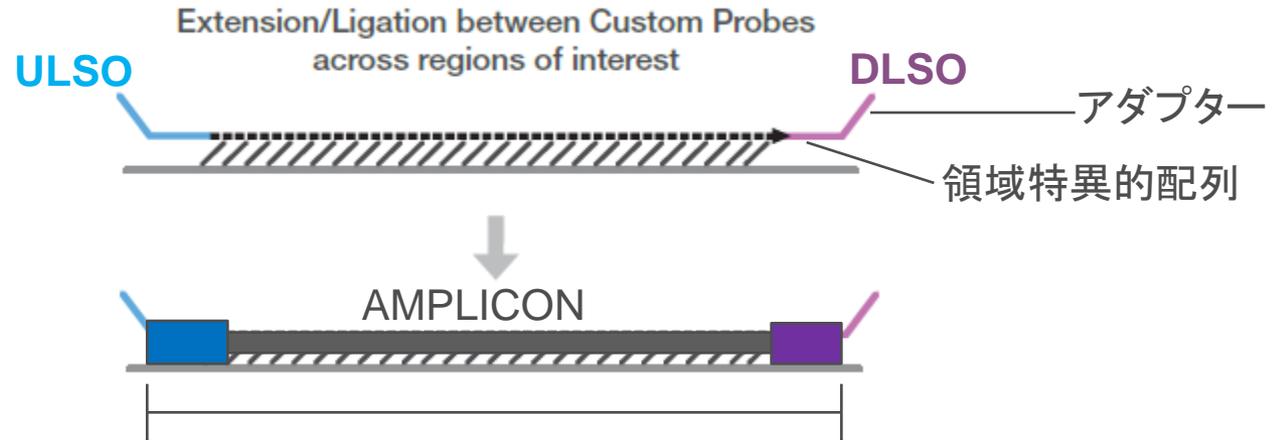
AFR  AMR  ASN  EUR

1000 Genomes: common variants detected in at least 1% of the population(s) studied.

Next



# DesignStudio Definition of Amplicon



Q1: ULISO と DLISO もシーケンスされる？

A: Yes.

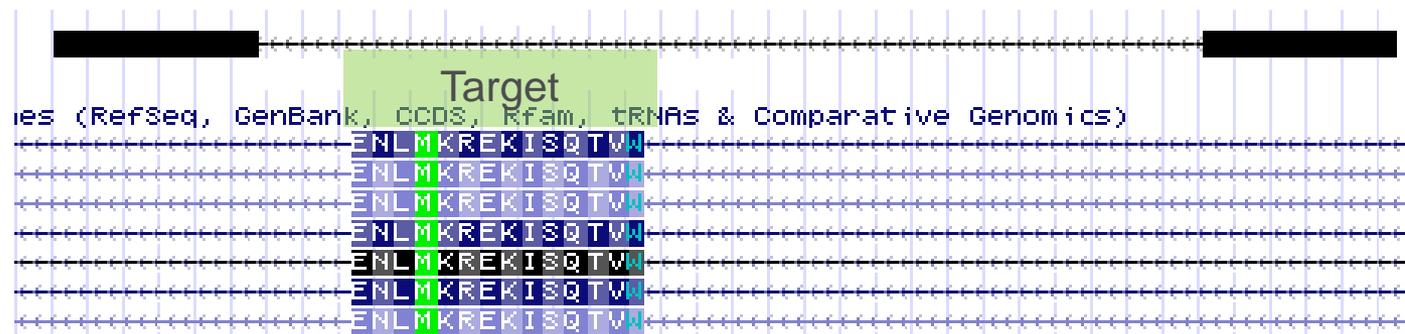
Q2: ULISOとDLISOからもバリエーションコールされる？

A: No. アライメント後に除去され、解析には用いられない

# Amplicon design

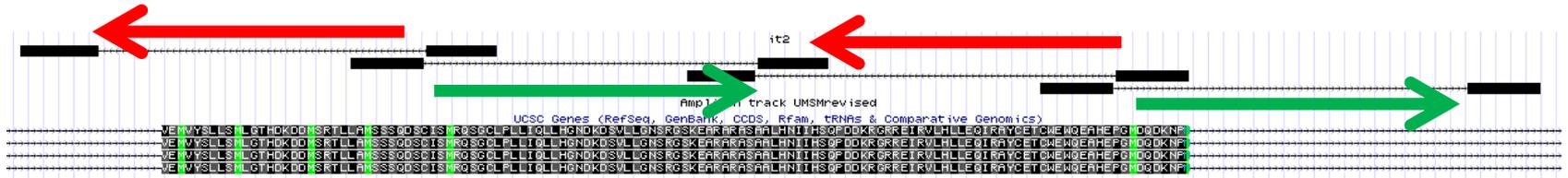
最も単純なデザイン:

1ターゲットに対し、1 amplicon。デザイン簡単



# Tiling

ターゲットに対し、1 ampliconではカバーできない場合  
ターゲットの領域が長くなると、複数のampliconを重複させてカバー



← Binds to forward strand

→ Binds to reverse strand

# TruSeq Custom Amplicon Design Workflow

Mylluminaにログイン

生物種を選択とプロジェクト概要の設定

ターゲットの選定；  
CDS/exons/full regions

デザイン確認、評価、追加/削除

最終デザインの確定、  
概算費用の確認

発注

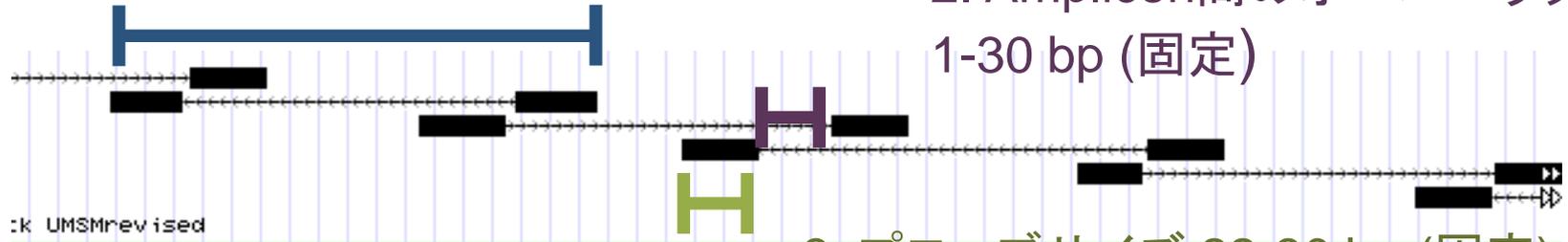
The screenshot displays the 'Configure Design' step (step 2) of the workflow. At the top, a progress bar shows five steps: 1. Assay Type, 2. Configure Design (highlighted), 3. Add Targets, 4. Submit Design, and 5. Review Design. The main content area is titled 'Configure Design' and includes several configuration options:

- Assay Version:** Two options are shown: '1.5' and 'Low Input'. The 'Low Input' option is highlighted in orange. To its right, text reads: 'TruSeq Custom Amplicon Low Input: Used for low input and/or FFPE applications. Not to be used with applications requiring 425bp.'
- Comparison Chart:** A checkbox is checked.
- Variant source:** Two options are shown: 'dbSNP 138' and '1000 Genomes'. The '1000 Genomes' option is highlighted in orange. To its right, text reads: 'Population:  AFR  AMR  ASN  EUR.   
1000 Genomes: common variants detected in at least 1% of the population(s) studied.'
- Amplicon size:** Three options are shown: '150 bp', '175 bp', and '250 bp'. The '175 bp' option is highlighted in orange.

A 'Next' button is located at the bottom right of the interface.

# 設計アルゴリズム: 調節できること、できないこと

## 1. Amplicon サイズ(調節可能)



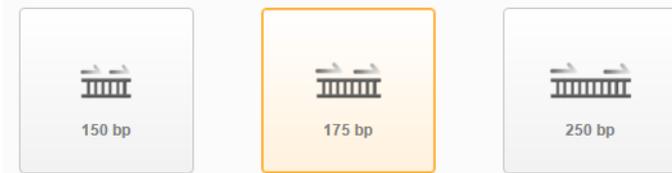
## 2. Amplicon間のオーバーラップ長 1-30 bp (固定)

## 3. プローブサイズ 22-30 bp (固定)

### TruSeq Custom Amplicon v1.5



### TruSeq Custom Amplicon low input

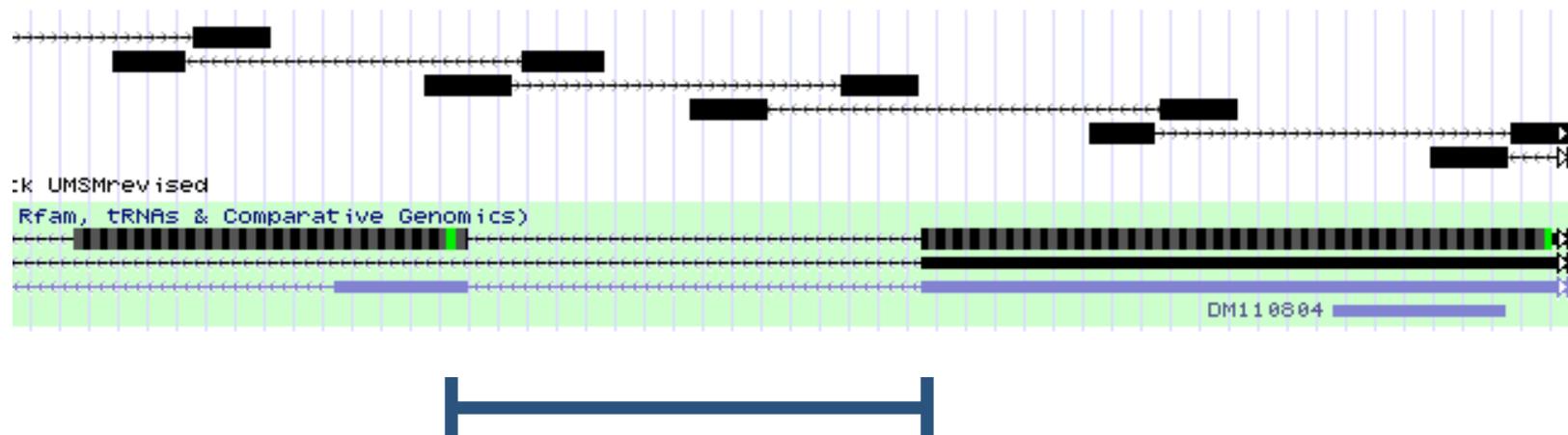


### Amplicon サイズの設定値

| Setting | Min | Max |
|---------|-----|-----|
| 150     | 125 | 190 |
| 175     | 170 | 190 |
| 250     | 225 | 275 |
| 425     | 400 | 450 |

# ターゲットのマージ

二つのターゲットが隣接していて、かつ、ターゲット間の距離が1 ampliconの最大サイズ設定値よりも短い時に強制実行



? Q: 250-bpのampliconサイズを設定している際、ターゲット間の距離が260-bp位置にプローブが設計された。この場合、ターゲットはマージされるか？

A: Yes. Ampliconサイズの最大値は275-bpで、ターゲット間距離よりも大きいため。

# Avoiding SNPs (調節可能, デフォルトは YES)

“Avoid SNP” にチェックを入れると、、、

| Regions                  | Targets       | Filter | Filtered (0) |            |             |                                     |
|--------------------------|---------------|--------|--------------|------------|-------------|-------------------------------------|
| <input type="checkbox"/> | Target Region | Chr    | Start        | Stop       | Targets     | Avoid SNP                           |
| <input type="checkbox"/> | FAS           | 10     | 90,750,288   | 90,775,542 | 10 / 10 CDS | <input checked="" type="checkbox"/> |

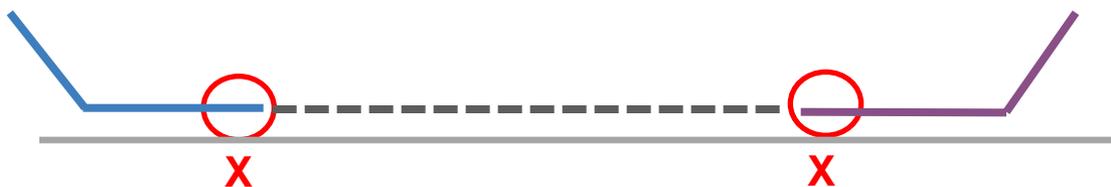
以下の場合にプローブが設計されなくなる：

プローブの設計位置に**3箇所以上** SNPが存在する



OR

末端1/4の箇所に**1箇所でも** SNPが存在する



# プローブデザインのアプローチ

- ▶ Tm値がアッセイ上の温度範囲内にあること
- ▶ G/C 含量が15-80%の範囲
- ▶ クロスハイブリする可能性のあるプローブは除外
- ▶ ヘアピン構造を取る可能性のある範囲は除外
- ▶ > 30箇所以上 off-target が存在するターゲットは除外



Q: クロスハイブリ、ヘアピン構造形成、極端なGC含量、といった特性のあるプローブを避けたい場合、何をすれば良い？

A: 何もしなくて良い。設計アルゴリズムが自動的に排除してくれる

# 本日のOutline

- ▶ TSCAとDesignStudioの概要
- ▶ デザインのアルゴリズム
- ▶ デザイン改善へのヒント



# TruSeq Custom Amplicon Design Workflow



Assay Type      Configure Design      Add Targets      Submit Design      Review Design

1      2      3      4      5

### Configure Design

Assay Version ?

Comparison Chart

Variant source ?

Amplicon size ?

TruSeq Custom Amplicon Low Input:  
Used for low input and/or FFPE applications. Not to be used with applications requiring 425bp.

Design dual pools ?

Population:

AFR  AMR  ASN  EUR

1000 Genomes: common variants detected in at least 1% of the population(s) studied.

Next

# SNP settings

Variant source 



dbSNP 138



1000 Genomes

Population:

AFR  AMR  ASN  EUR

 1000 Genomes: common variants detected in at least 1% of the population(s) studied.

データベースの選定を考慮する

# SNP settings



Q1: SNPはデザイン時に考慮した方が良い？

A: **Yes!** SNPが存在する箇所ではプローブのハイブリ効率に影響が出る。カバレッジの低下や偽陽性のバリエーション検出などの危険もあり。

Q2: 1000 Genomes or dbSNP\_138。違いは？

## 1000 Genomes:

- 1KGP由来のSNP
- MAF **1%** 以上のSNPのみを対象に設定
- 集団ごとのSNP情報を適用可能

## dbSNP138:

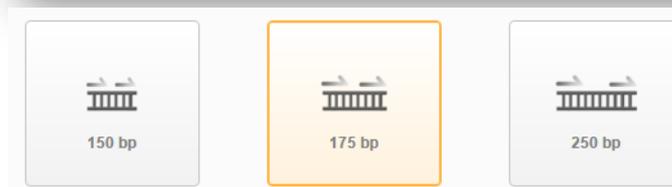
- 全集団のSNP情報を考慮
- MAFの上限/下限は設定なし。全てのSNPを考慮
- 複数のSNPが報告された領域では、デザインが難しくなる場合も

# Ampliconサイズの設定

TSCA v1.5



TSCA low input



Q1: 短いampliconを設計する利点は？

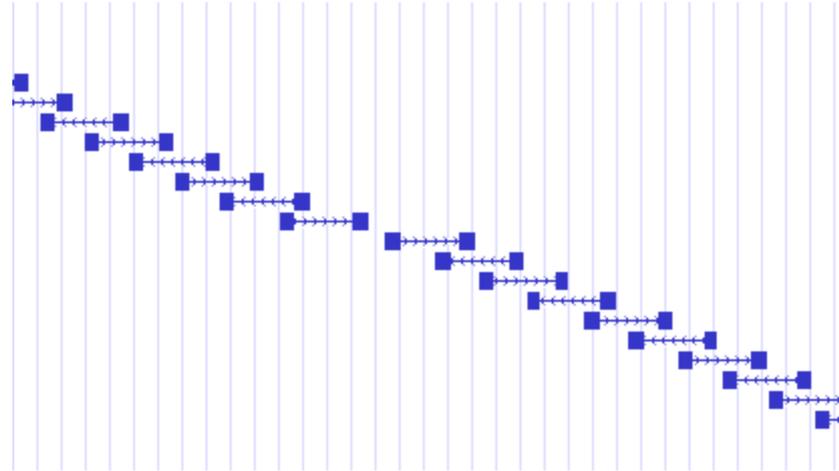
- FFPEや分解したDNAに強い
- シーケンス長が短くて済む  
(シーケンスコストが低く抑えられる)

Q2: 短いampliconの欠点は？

- 特定サイズのターゲットを解析するのにより多くのAmpliconの設計が必要  
(ライブラリー調製コストが上がる)
- タイリングが難しく、ギャップを生じやすくなる場合も

# 1. FFPEサンプルに適したAmpliconサイズ

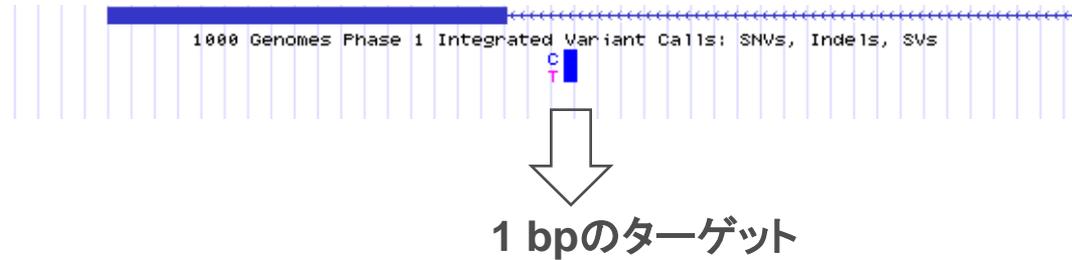
短いampliconサイズが良いが、最短サイズ (150 bp) にするとタイリングが難しくギャップを生じるリスクもある



→ FFPEには**175 bp**のサイズ設定適用  
がお勧め  
(最大amplicon長は150 bpの設定時と同じ)

| Setting | min | max |
|---------|-----|-----|
| 150     | 125 | 190 |
| 175     | 170 | 190 |
| 250     | 225 | 275 |
| 425     | 400 | 450 |

## 2. 短いターゲット / SNPの検出に適したサイズ設定



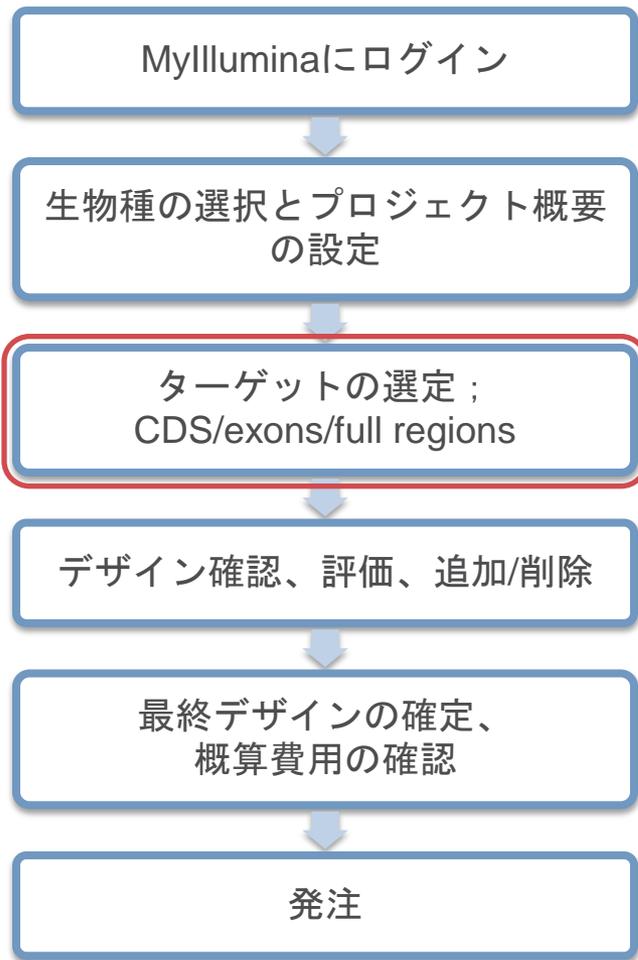
→タイリングの要らない短いターゲットには150 bpの設定がお勧め

### 3. その他の場合はどうするか？ 250 or 425-bp?

| 250-bp                      | 425-bp                             |
|-----------------------------|------------------------------------|
| より多くのアンプリコンが必要 (-)          | 少ないアンプリコンでターゲットをカバー (+)            |
| リード長が短くて済む (+)              | 長いリード長が必要 (-)                      |
| V2 300サイクルキットが使える (+)       | V2 500サイクル or v3 600サイクルキットが必要 (-) |
| アンプリコンの中央辺りでも高いQ30が得られる (+) | アンプリコンの中央辺りでQ30が落ちる (-)            |
| ホモロジーの高い領域や偽遺伝子の識別精度が劣る     | ホモロジーの高い領域や偽遺伝子の識別精度が高い            |

\*\*425bpは TSCA v1.5でのみ利用可能

# TruSeq Custom Amplicon Design Workflow



Assay Type      Configure Design      Add Targets      Submit Design      Review Design

1      2      3      4      5

### Configure Design

Assay Version ?

Comparison Chart

Variant source ?

Amplicon size ?

TruSeq Custom Amplicon Low Input:  
Used for low input and/or FFPE applications. Not to be used with applications requiring 425bp.

Design dual pools ?

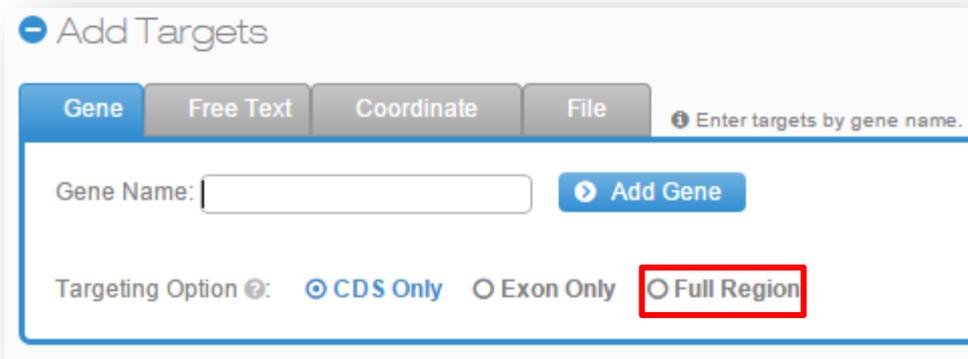
Population:

AFR  AMR  ASN  EUR

1000 Genomes: common variants detected in at least 1% of the population(s) studied.

Next

# ターゲットの選定方法: Full Region



The screenshot shows a web interface for adding targets. At the top, there are tabs for 'Gene', 'Free Text', 'Coordinate', and 'File'. Below the tabs is a text input field for 'Gene Name' and an 'Add Gene' button. Underneath, there are radio buttons for 'Targeting Option': 'CDS Only', 'Exon Only', and 'Full Region'. The 'Full Region' option is selected and highlighted with a red box.

**Full Region:** TSCAではあまり向かない:

-イントロンにはリピート配列が多く、TSCAではカバーが難しい

-サイズが大き過ぎ!!! TSCAが連続領域でデザインできる最長は 24 kb

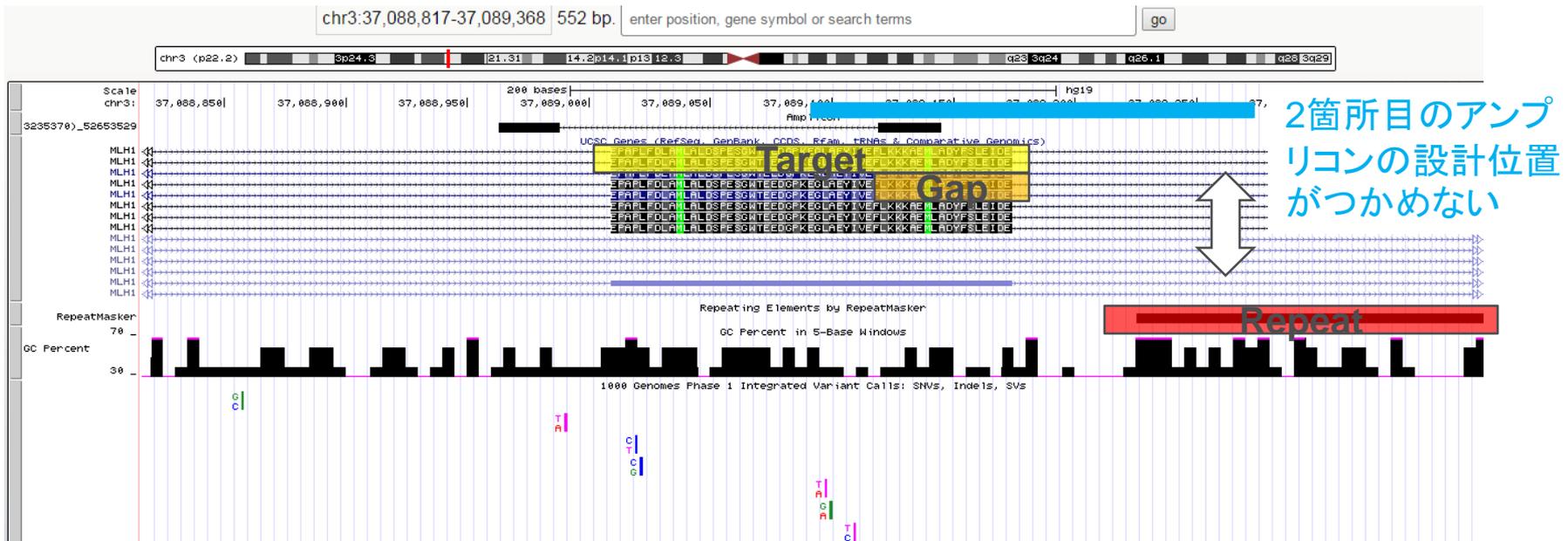
→ **打開策:** ターゲットを10-kb未満の小領域に分け、ターゲット間にアンプリコンサイズ範囲の最大値以上の間隔を空ける



# リピートに隣接する箇所はギャップになる

DesignStudioは設定した部位に最も近くなるように設計を実施する

ターゲットエクソン付近にリピート配列があると、アンプリコンの設計に失敗してギャップになる可能性が高い (サイズの小さいエクソンほど問題になりやすい)。



**打開策:** 設計位置の設定範囲をずらす、あるいは設計範囲を広げる



# まとめ

## デザイン改善へのヒント

- デザインプロジェクトの設定条件を変更
- ターゲット領域を個別に変更して、再デザイン
- デザインアルゴリズムと注意点を確認しながら、デザインを進めていく。

## 参考データベース

- TruSeq Custom Amplicon  
<http://jp.illumina.com/products/truseq-custom-amplicon-low-input.html>
- DesignStudio  
<http://designstudio.illumina.com/>

<http://designstudio.illumina.com/>

ご清聴ありがとうございました

本日セッション終了後のご質問は、  
[techsupport@illumina.com](mailto:techsupport@illumina.com)にお問い合わせください