



NGS現場の会 第三回研究会
2013年9月5日(木)

Kapaエンジニア酵素を利用した ヒトの主要組織適合遺伝子複合体(HLA) タイピング法の開発事例

細道一善

国立遺伝学研究所人類遺伝研究部門

ロングPCRとNextera Kitによるシーケンス

✖ イメージを表示できません。メモリ不足のためにイメージを開くことができないか、イメージが破損している可能性があります。コンピューターを再起動して再度



Nextera™ DNAサンプル調製キット



手軽にNGSを使う

- ・数十kb程度の領域(遺伝子)をシーケンスしたい
- ・96検体を1ランで解析したい
- ・PCR産物をクローニングしてシーケンスするような解析

DNAライブラリー調整法の改良

HLA遺伝子でのGC%の高い領域での改善例



Nexteraライブラリーのノーマライゼーション



トピックス

- ロングPCRとNextera Kitによる
HLA遺伝子の配列決定
- Kapaエンジニア酵素による
ライブラリー調整における改良事例
- AMPure XPによる
ライブラリーのノーマライゼーション

トピックス

- ロングPCRとNextera Kitによる
HLA遺伝子の配列決定
- Kapaエンジニア酵素による
ライブラリー調整における改良事例
- AMPure XPによる
ライブラリーのノーマライゼーション

HLAとは

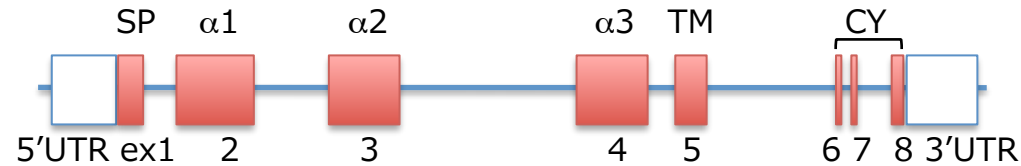
- HLA (Human Leukocyte Antigen
= ヒト白血球抗原)
 - 1954年、白血球の血液型として発見
- 組織適合性抗原
(MHC; Major Histocompatibility Complex)
- ヒトに関しては MHC = HLA

HLA分子

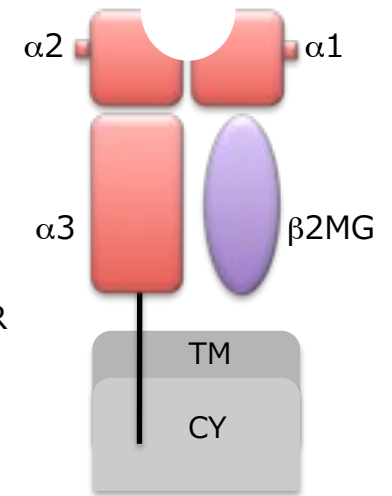
HLAクラスI

HLA-A, -C, -Bなど

内在抗原提示
細胞性免疫

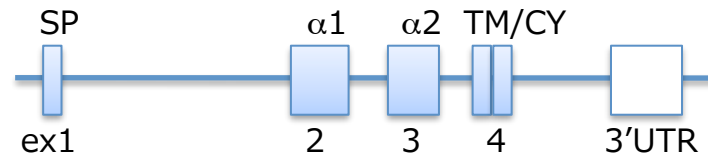


HLAクラスI分子

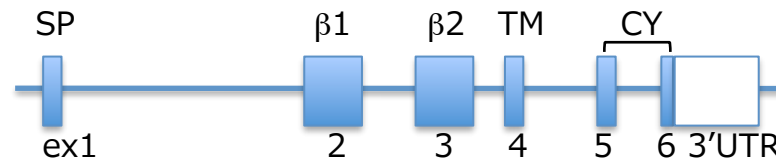


HLAクラスII

α鎖遺伝子
HLA-DRA1, -DQA1など

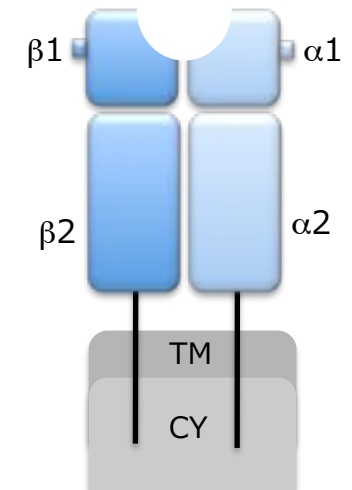


β鎖遺伝子
HLA-DRB1, -DQB1, -DPB1など

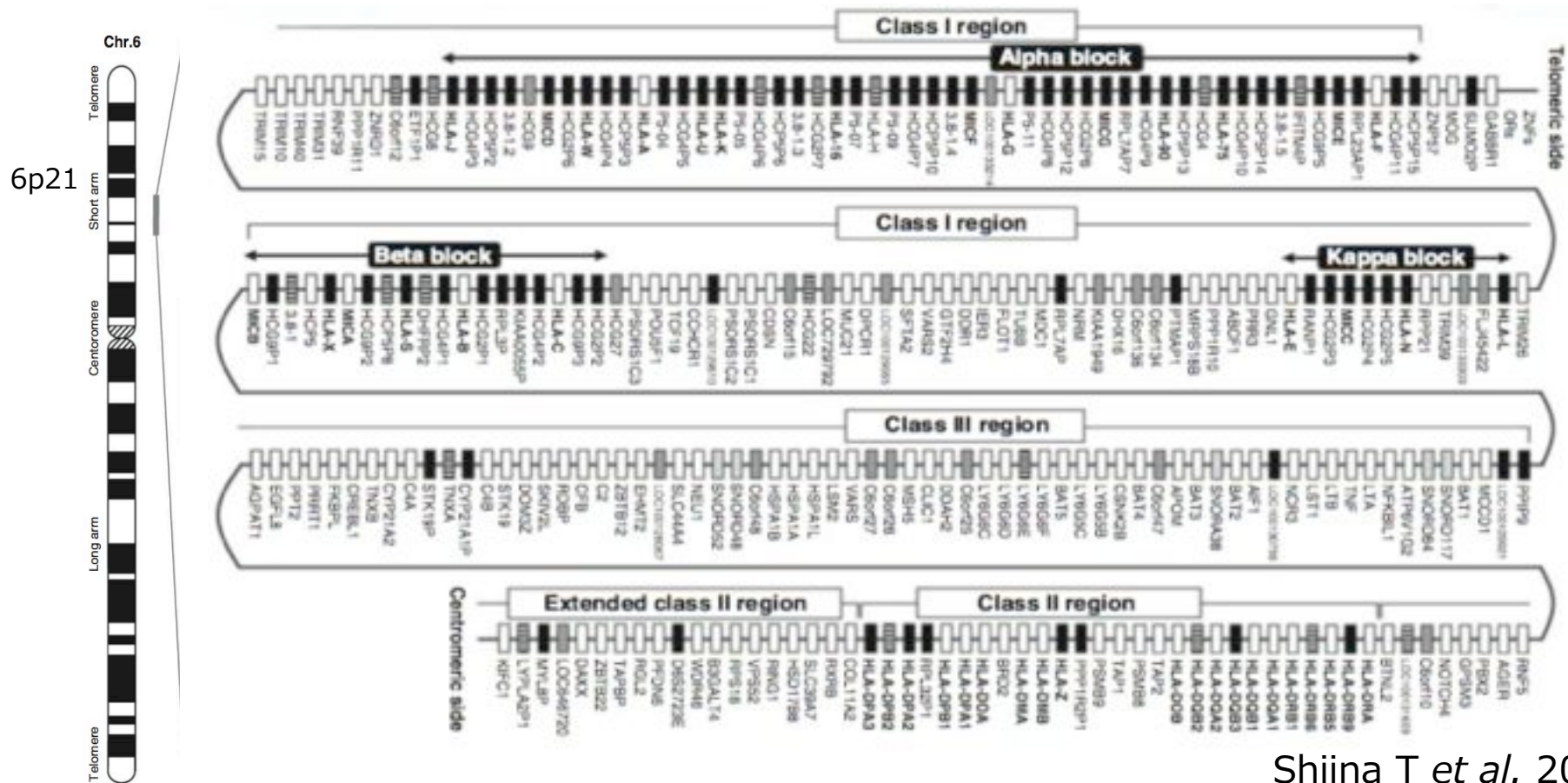


外来性抗原提示
液性免疫

HLAクラスII分子



HLA領域のゲノム配列



Shiina T et al. 2009

HLA領域の特徴

252の遺伝子、6つの古典的HLA遺伝子と少なくとも132のタンパク質をコードする遺伝子を含む

極めて高度な多型性を示す

100以上の疾患および薬剤副作用と関連する

HLA クラスI α 1および α 2ドメインの多様性

5 amino acids 3

```

B52:01 AGSHSMRYFYTAMSRPGRGEPFRFIAVGYVDDTQFVRFSDAASPRTEPRAPWIEQEGPEYWDRETQISKNTNTQTYRENLFIALRYYNQSEAG
B44:03 AGSHSMRYFYTAMSRPGRGEPFRFITVGYVDDTLFVRFSDATSPRKEPRAPWIEQEGPEYWDRETQISKNTNTQTYRENLFTALRYYNQSEAG
B07:02 AGSHSMRYFYTSVSRPGRGEPFRFISVGYVDDTQFVRFSDAASPREPRAPWIEQEGPEYWDRNTQIYKAQAQTDRESLFLNLRGYYNQSEAG
B54:01 AGSHSMRYFYTAMSRPGRGEPFRFIAVGYVDDTQFVRFSDAASPRGEPAPWVEQEGPEYWDRNTQIYKAQAQTDRESLFLNLRGYYNQSEAG
B46:01 AGSHSMRYFYTAMSRPGRGEPFRFIAVGYVDDTQFVRFSDAASPRMAPRAPWIEQEGPEYWDRETQKYKRQAQTDREVSLFLNLRGYYNQSEAG
*****..*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****

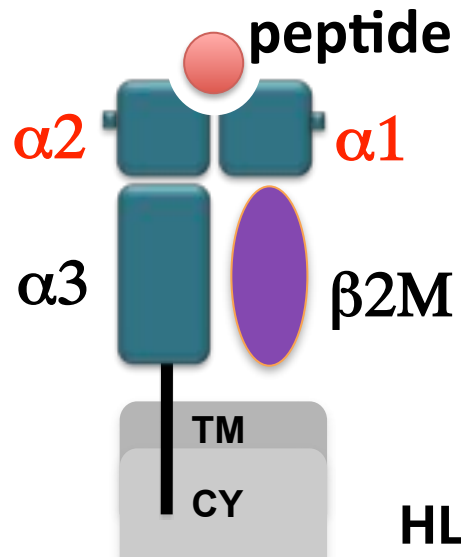
```

```

B52:01 SHTWQITMYGCDVGPDPGRLLRGHNQYAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVEWLRRLHLENGKETLQRAD
B44:03 SHIIQRMYGCDVGPDPGRLLRGYDQDAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGLCVESLRRYLENGKETLQRAD
B07:02 SHTLQSMYGCDVGPDPGRLLRGHDCYAYDGKDYIALNEDLRSWTAADTAAQITQRKWEAAREAEQLRAYLEGECVEWLRRLYLENGKDKLERAD
B54:01 SHTWQITMYGCDLGPDPGRLLRGHNQLAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAARVAEQLRAYLEGTCVEWLRRLYLENGKETLQRAD
B46:01 SHTLQRMYGCDVGPDPGRLLRGHDCSAYDGKDYIALNEDLSSWTAADTAAQITQRKWEAAREAEQLRAYLEGLCVEWLRRLYLENGKETLQRAD
**..*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****

```

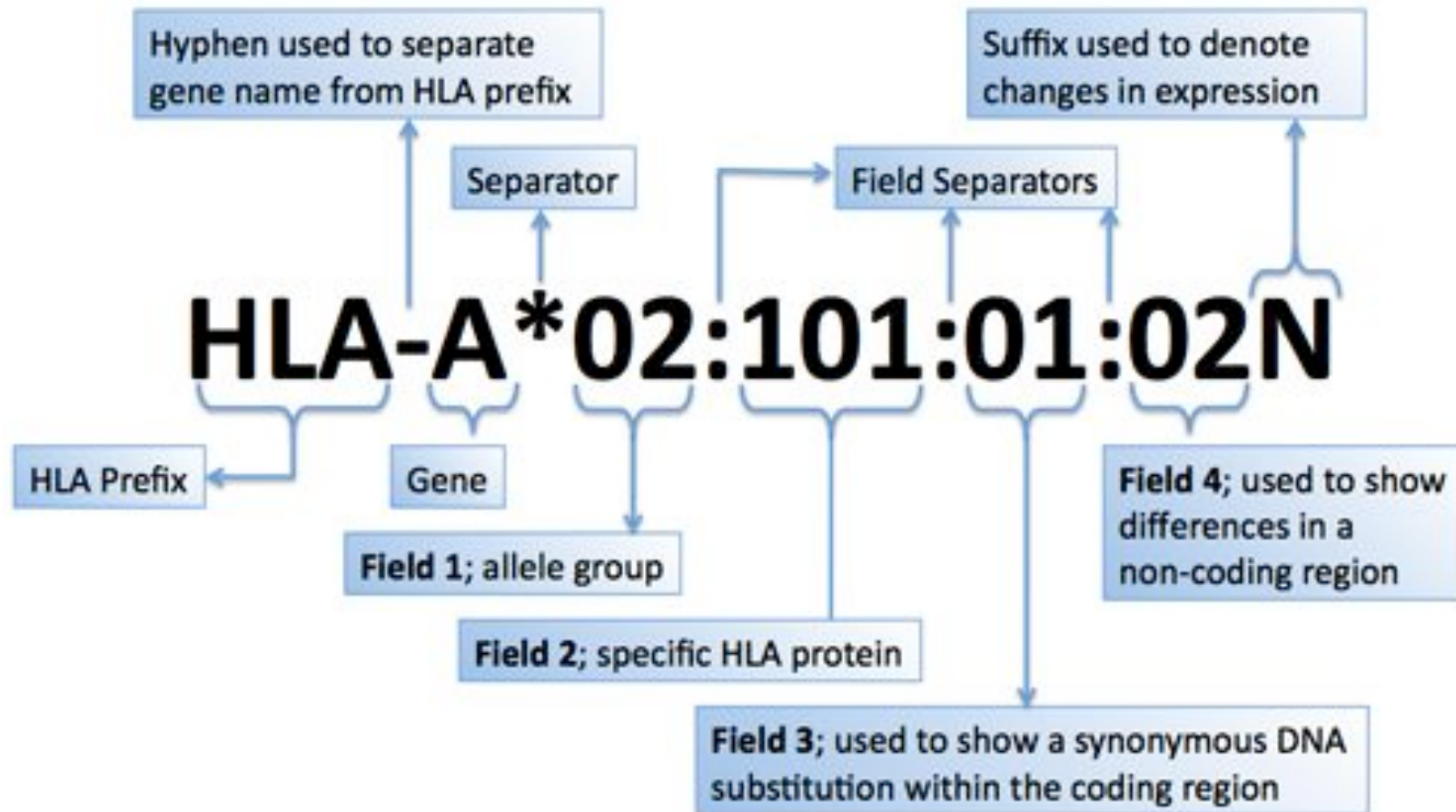
3 3 4 3 3



日本人で頻度の高いHLA-B アレル5種間

アミノ酸配列での類似性: **88.9%**
 塩基配列での類似性: **93.9%**

HLAアレルの命名法

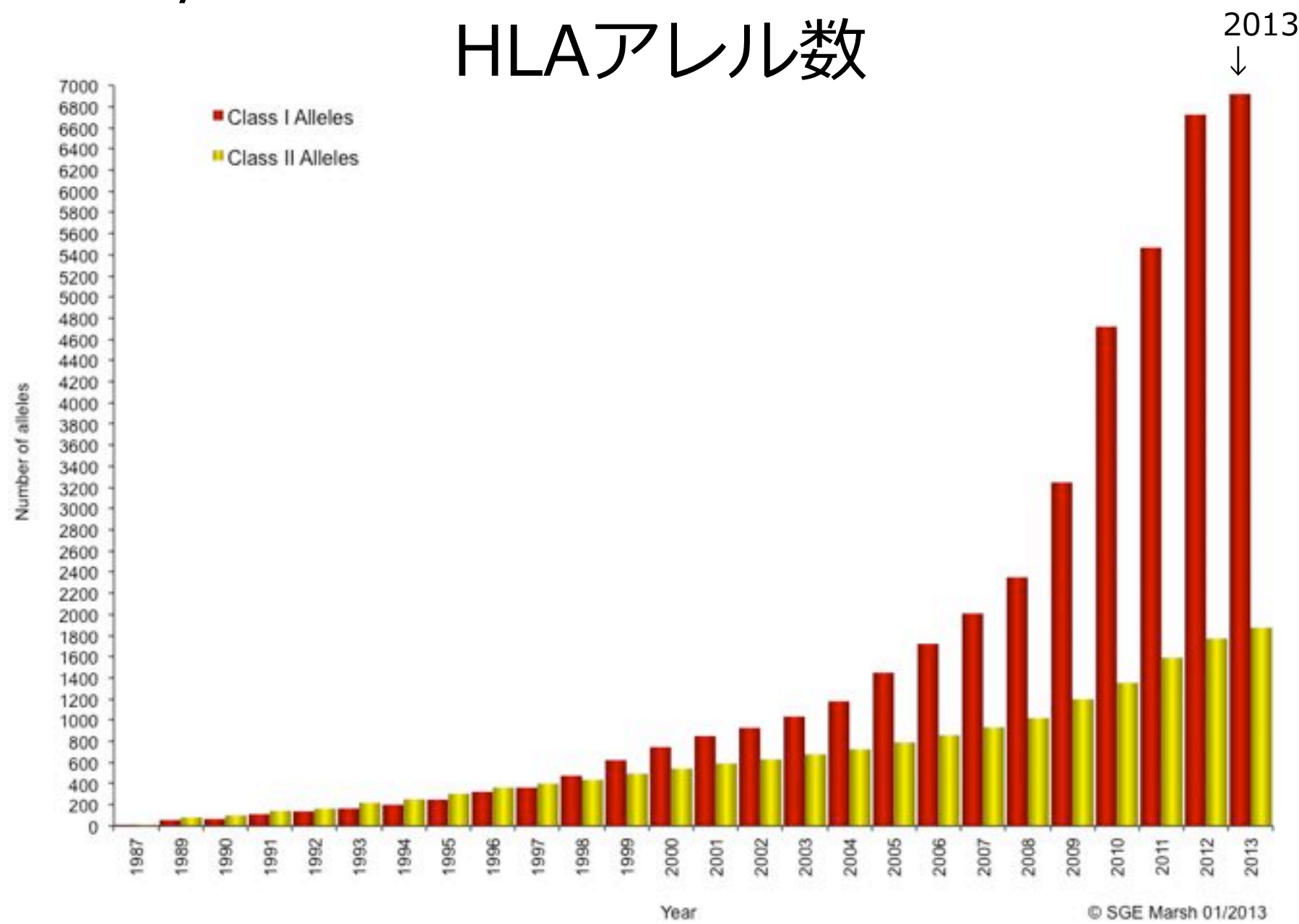


HLAアレルの数字が示すのはHLA遺伝子配列そのもの

IMGT/HLAデータベースに登録されている HLAアレル数

Numbers of HLA Alleles			
HLA Class I Alleles		7,089	
HLA Class II Alleles		2,065	
HLA Alleles		9,154	
HLA Class I			
Gene	A	B	C
Alleles	2,244	2,934	1,788
Proteins	1,612	2,211	1,280
Nulls	109	97	47
HLA Class II			
Gene	DRB	DQB1	DPB1
Alleles	1,418	323	185
Proteins	1,051	216	153
Nulls	32	7	6

IMGT/HLAデータベースに登録されている HLAアレル数

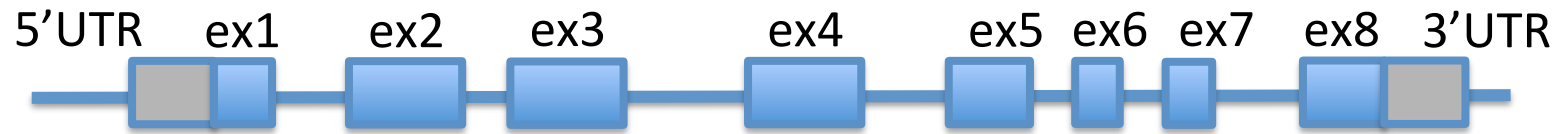


HLAの医学的興味

- 臓器移植の成否
- 造血幹細胞移植に伴う移植片対宿主病
- 生活習慣病、自己免疫疾患
- ウイルス感染症における防御と重症化
- 薬剤副作用
- 再生医療用iPS細胞ストック
- がんワクチン療法

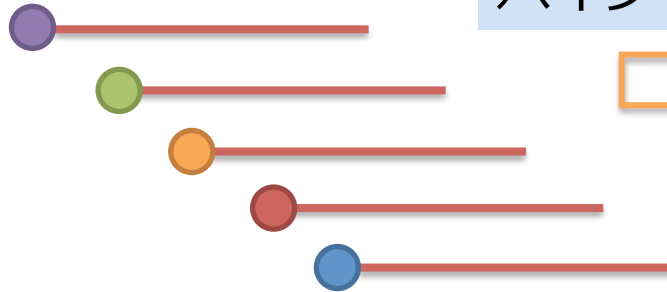
現行のHLAタイピング法

PCR-SSO (sequence specific oligonucleotide)



+

ハイブリダイズ

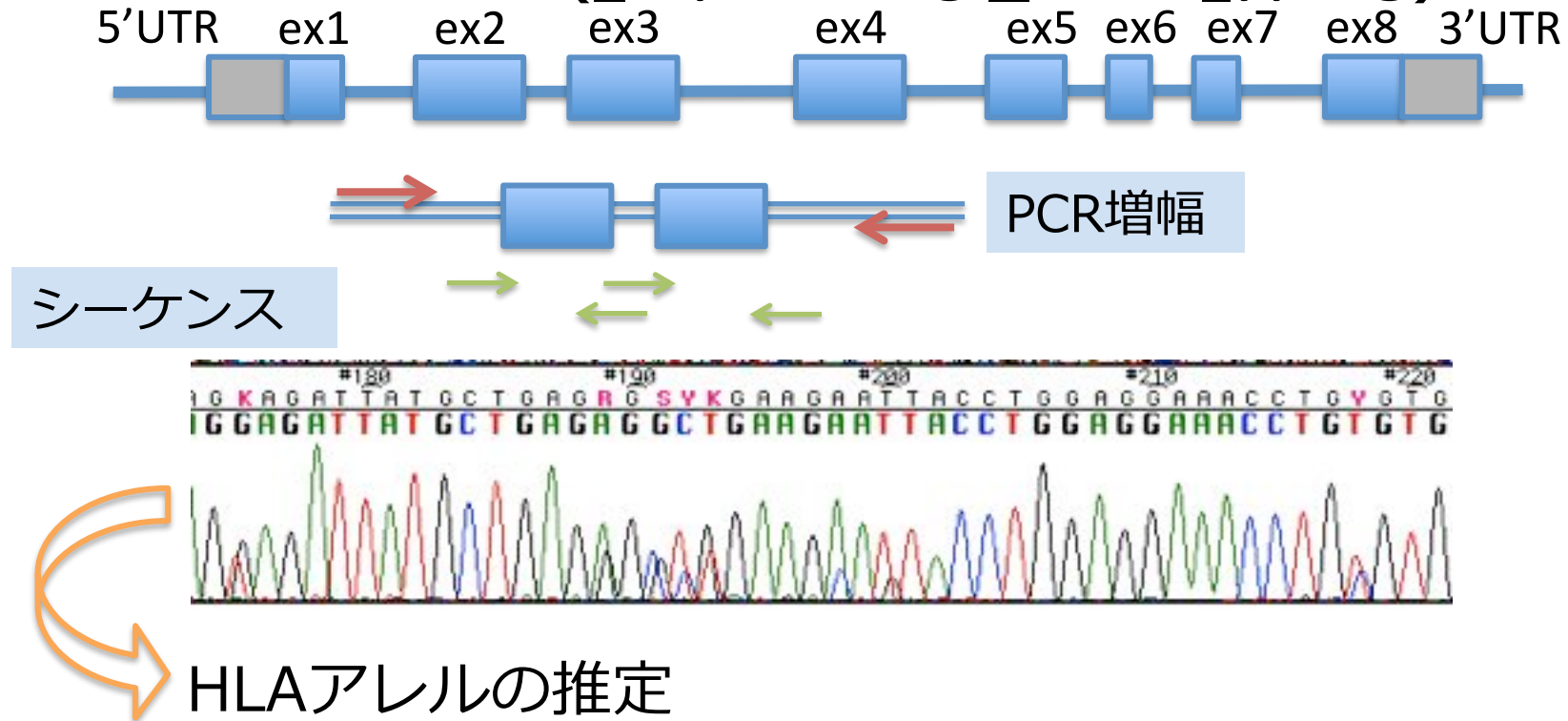


HLAアレルの決定

*B*51:02:01* *B*58:01:01*

現行のHLAタイピング法

PCR-SBT (sequencing based typing)



複数の組み合わせ候補

Combination 1 $B^*07:021 + B^*35:011$

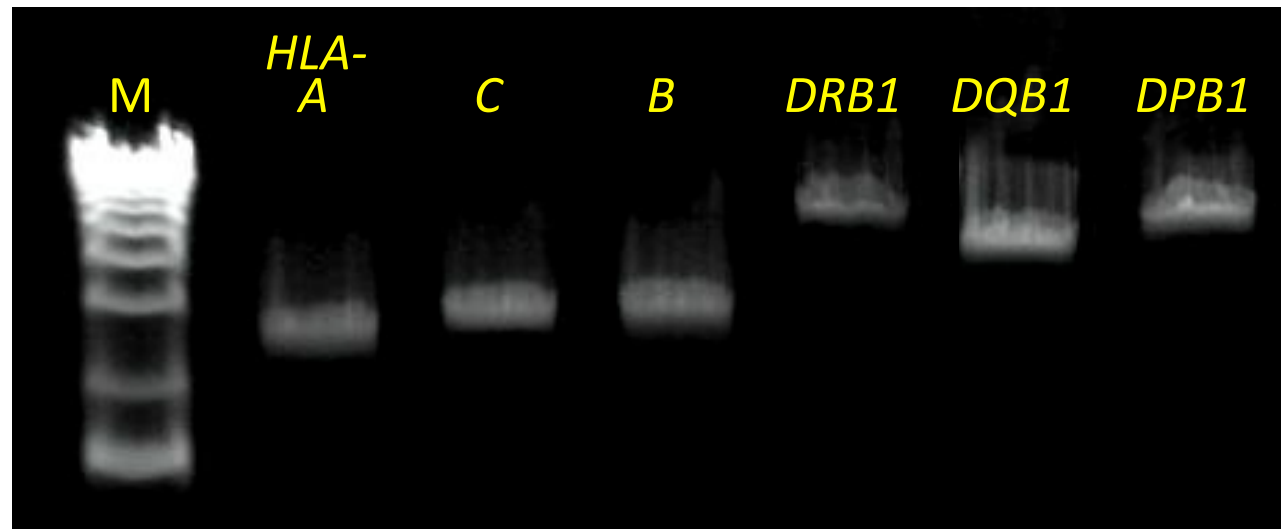
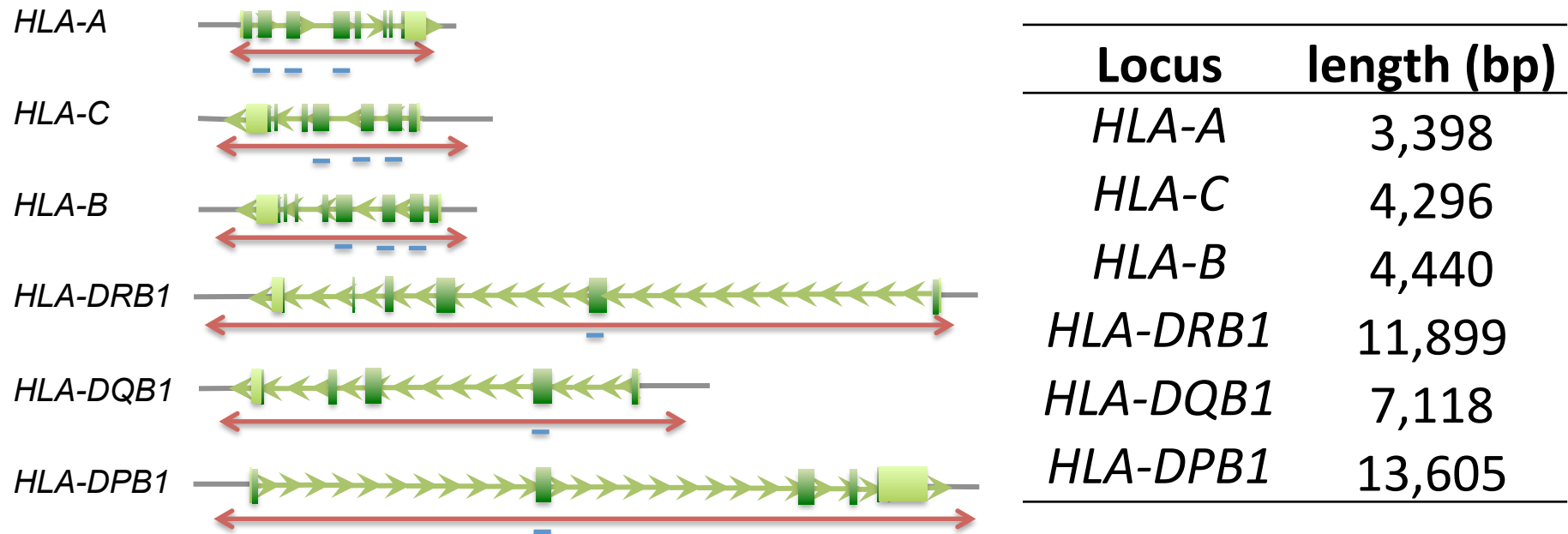
Combination 2 $B^*07:18 + B^*35:05$

Combination 3 $B^*07:09 + B^*35:34$

Combination 4 $B^*07:24 + B^*35:15$

Ambiguity

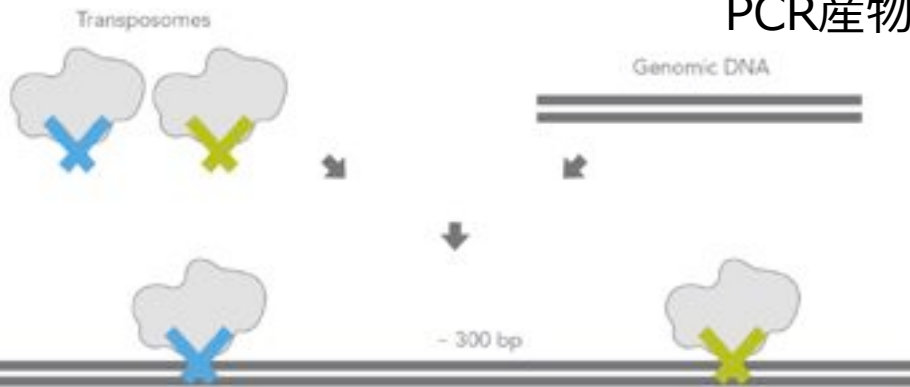
HLA遺伝子のロングPCR



Nextera DNA Sample Preparation Kit によるライブラリ調整

PCR産物20 ng

Nextera transposome と
アダプターのテンプレート
DNAへの結合



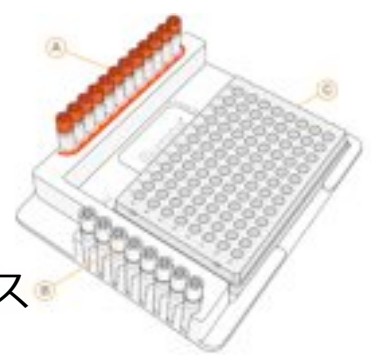
断片化とアダプターの結合
(Tagmentation)



Reduced-Cycle
PCR Amplification



最大96サンプル
のマルチプレックス



PCR によるIndexと
アダプター配列の追加



90分

Library length : 300bp ~ >1.2kb

NexteraとMiSeqによる HLA遺伝子のシーケンシング

PCR増幅

locus	Length (bp)
HLA-A	3,398
HLA-B	4,296
HLA-C	4,440
HLA-DRB1	11,899
HLA-DQB1	7,118
HLA-DPB1	13,605

ライブラリ調整

Nextera DNA Sample Preparation Kit
(illumina)

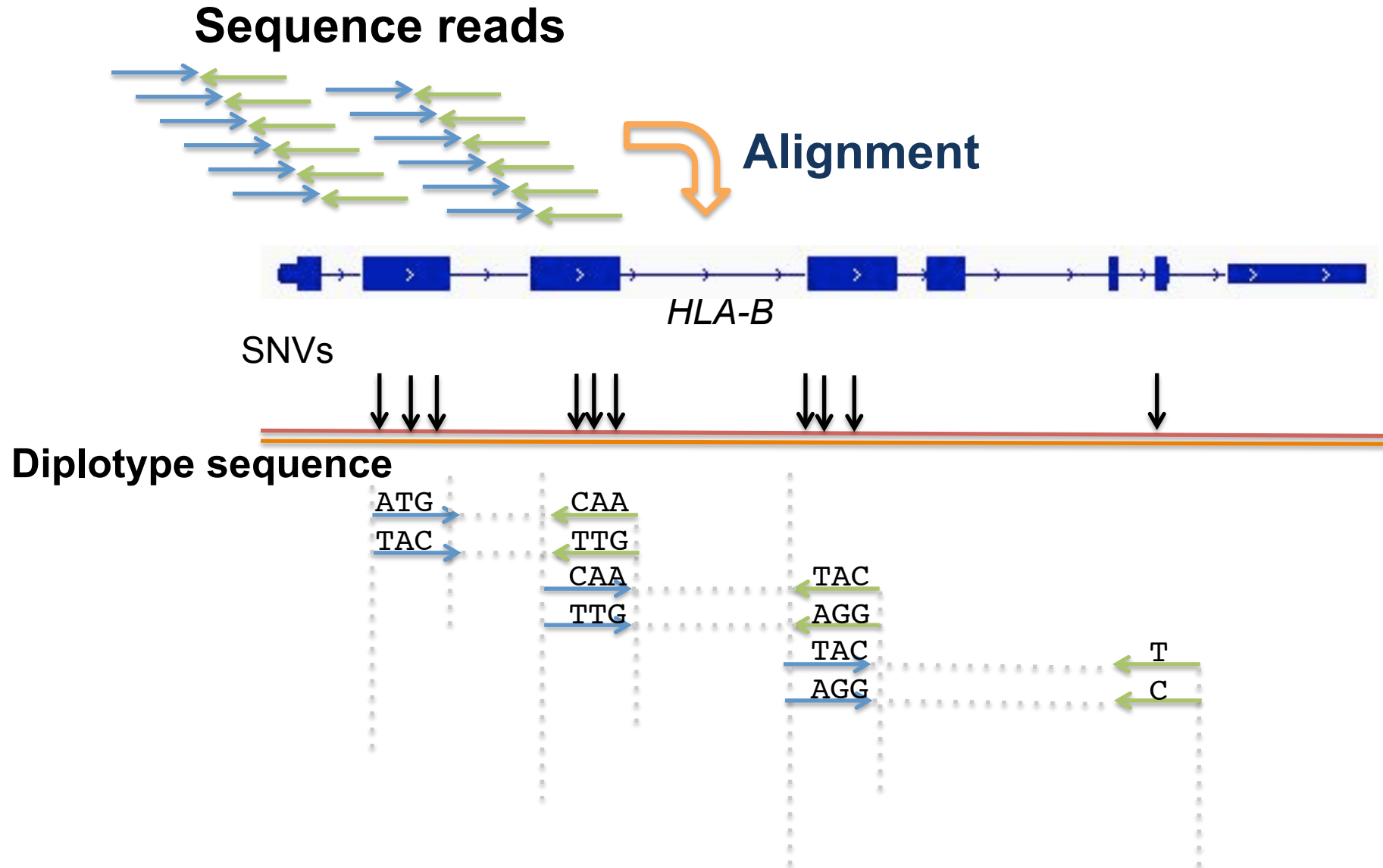


シーケンシング MiSeq v2 (illumina)

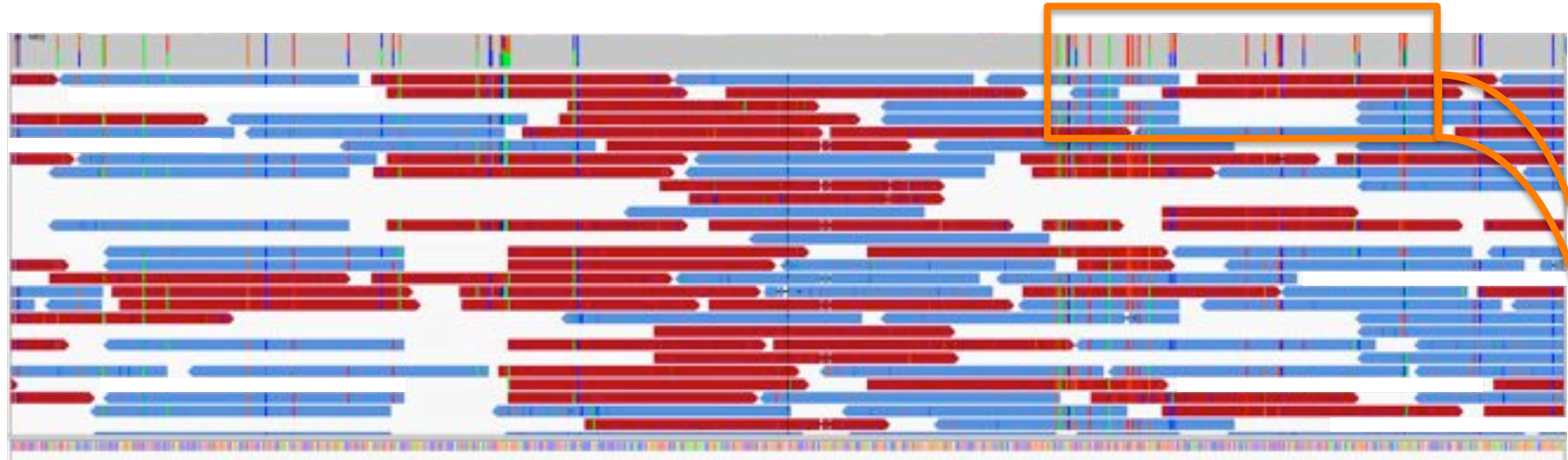
Read length 2 x 250 bp
Output 7.5-8.5 Gb
Total time 39 hr

データ解析

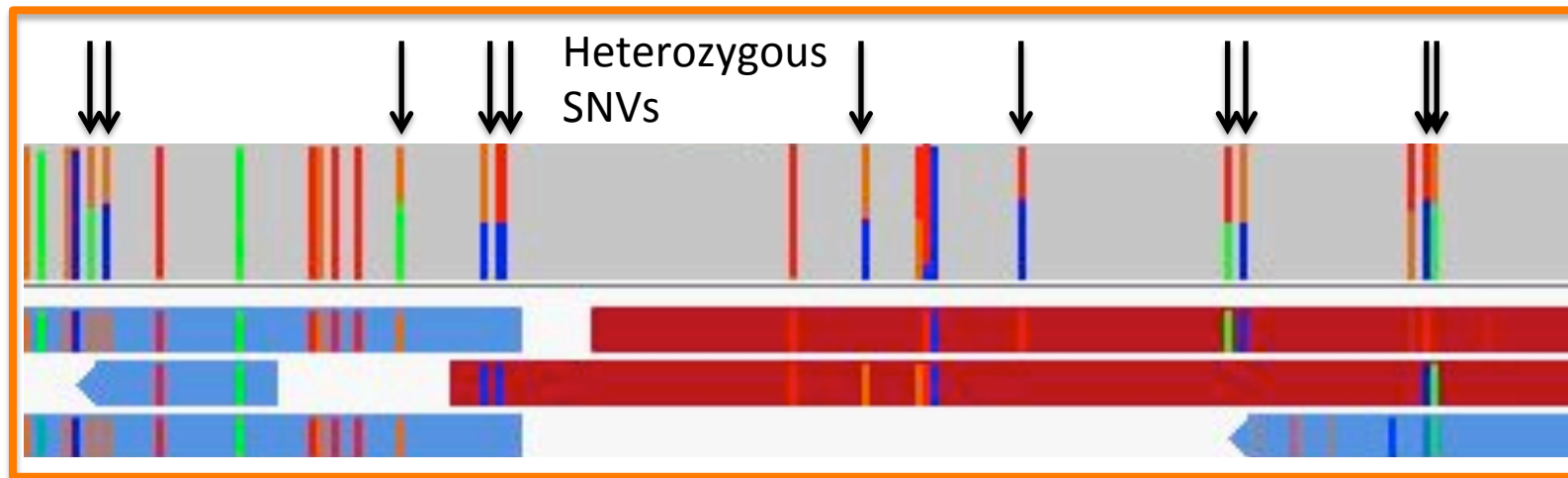
HLA遺伝子配列決定の概要



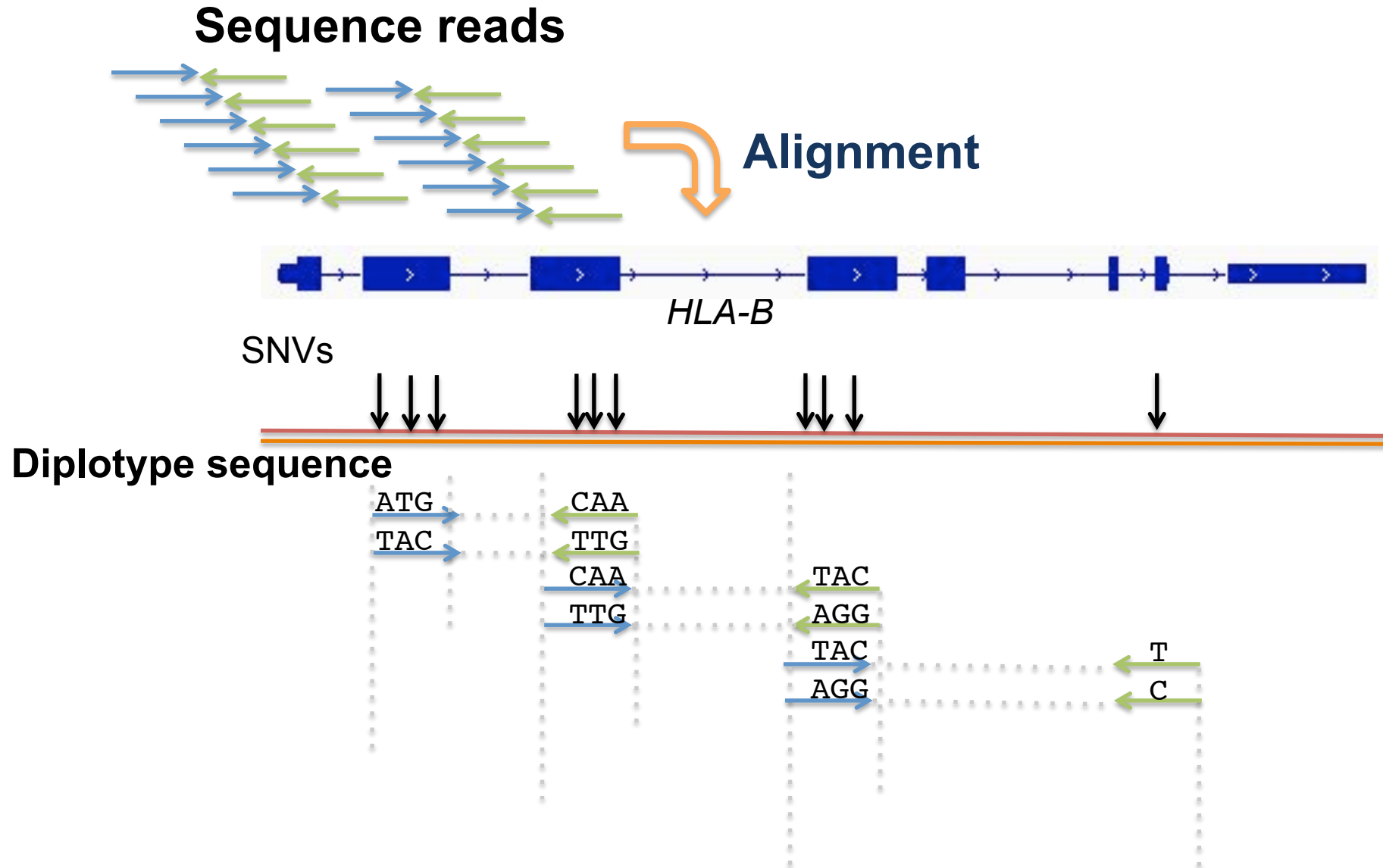
HLA-B エクソン2および3



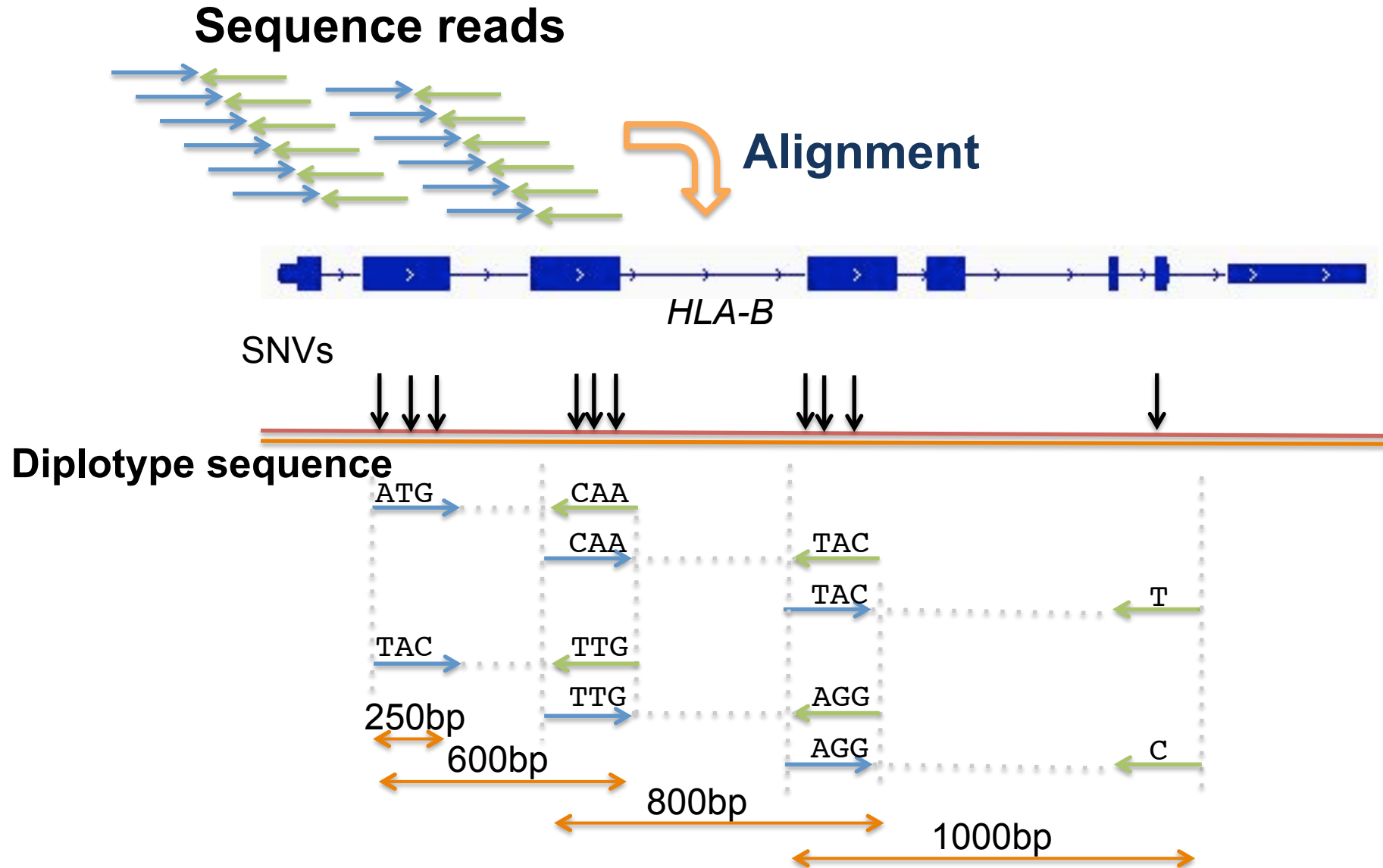
maximum differences = 80 in the read(250bp)



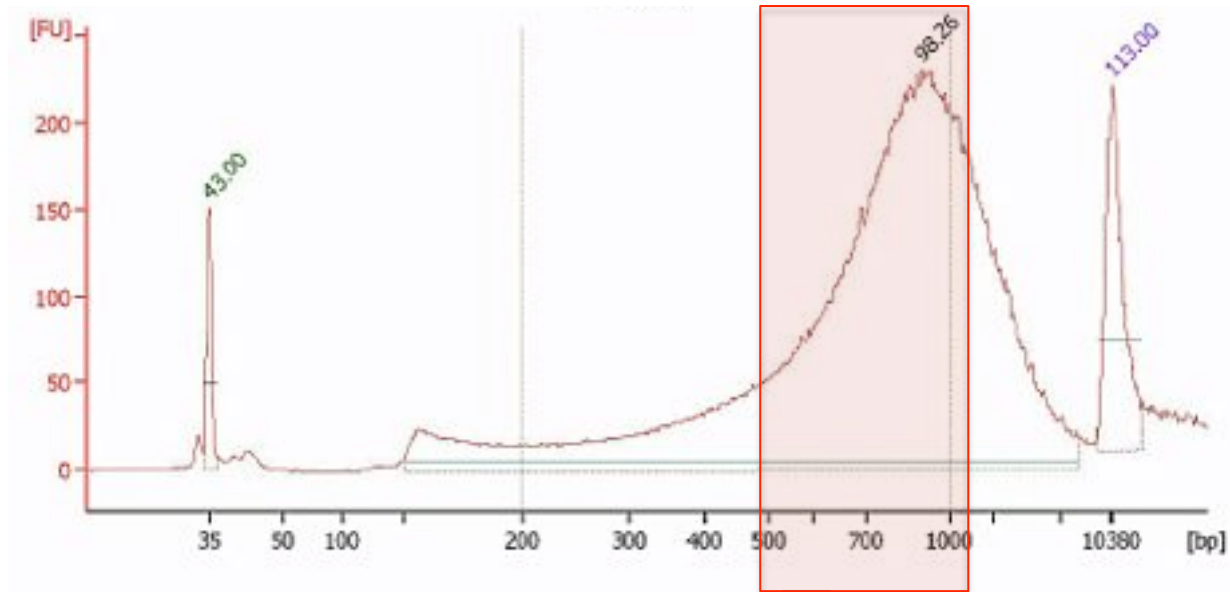
HLA遺伝子配列決定の概要



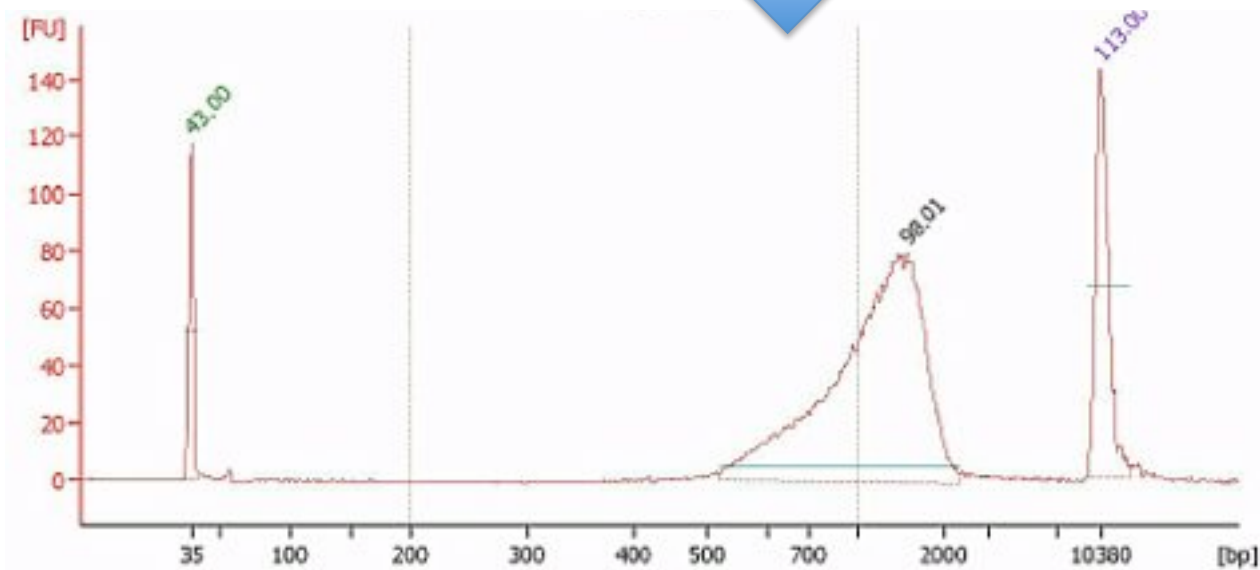
HLA遺伝子配列決定の概要



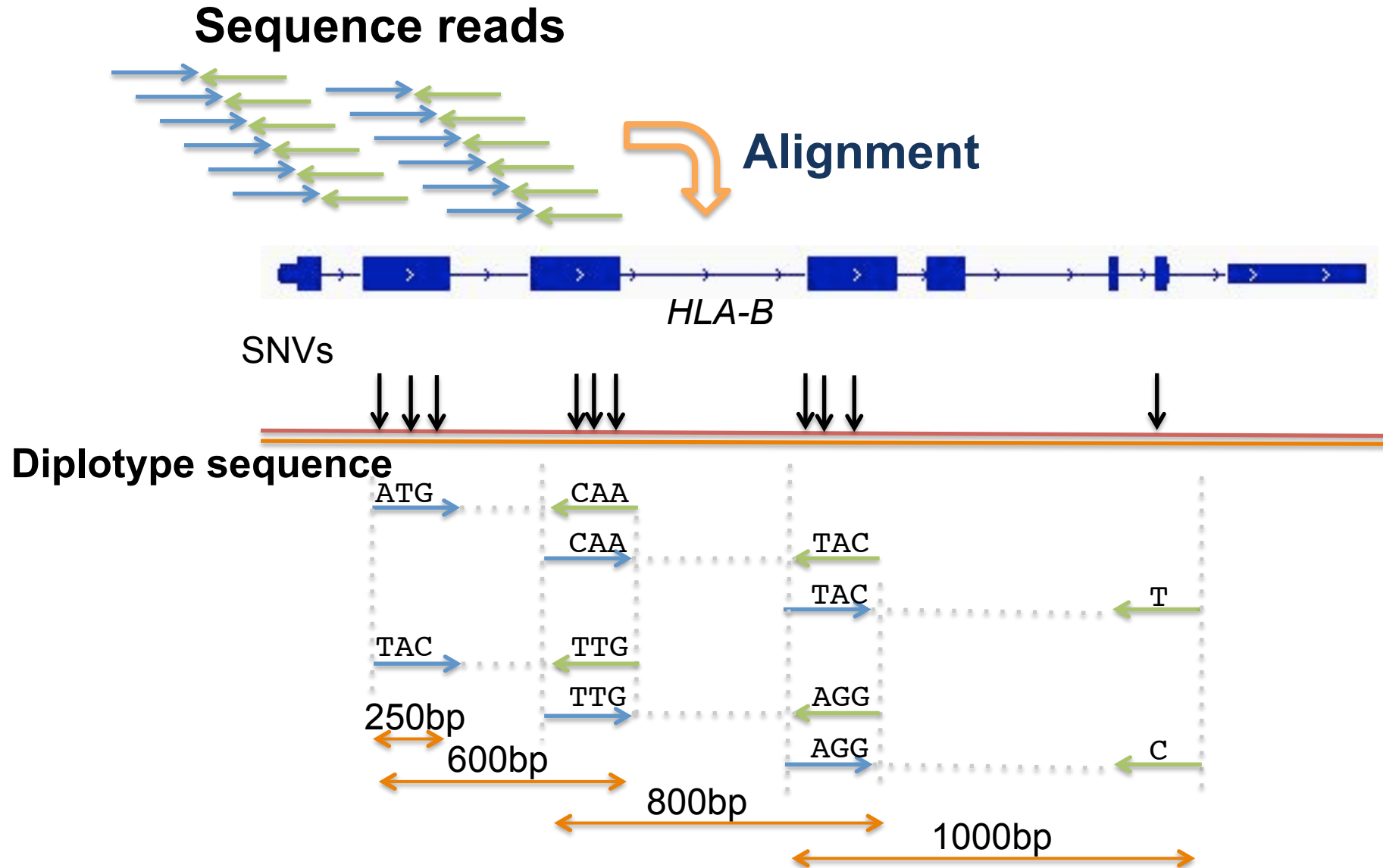
Nextera DNA ライブラリのサイズ選択



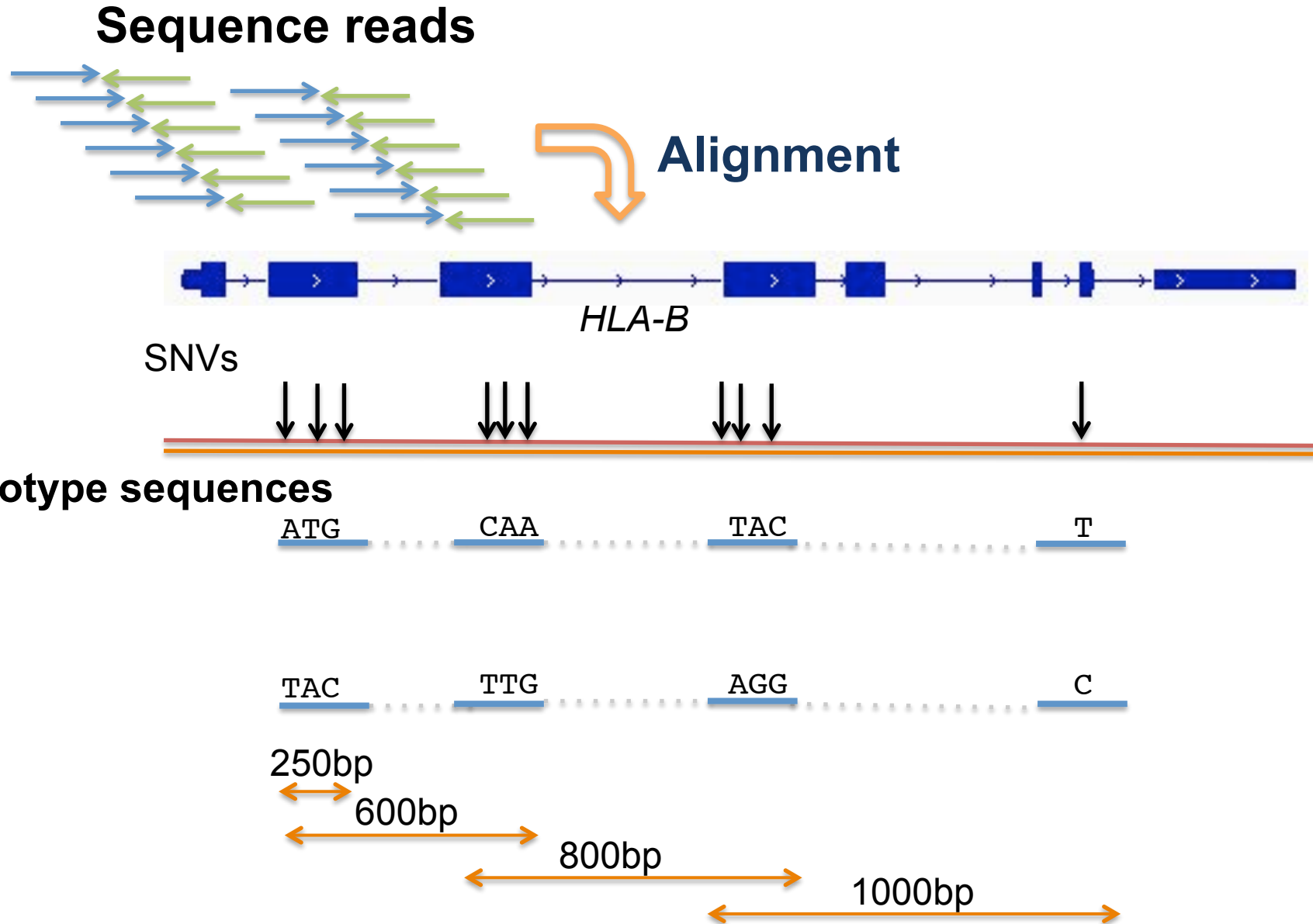
Agarose gel size selection



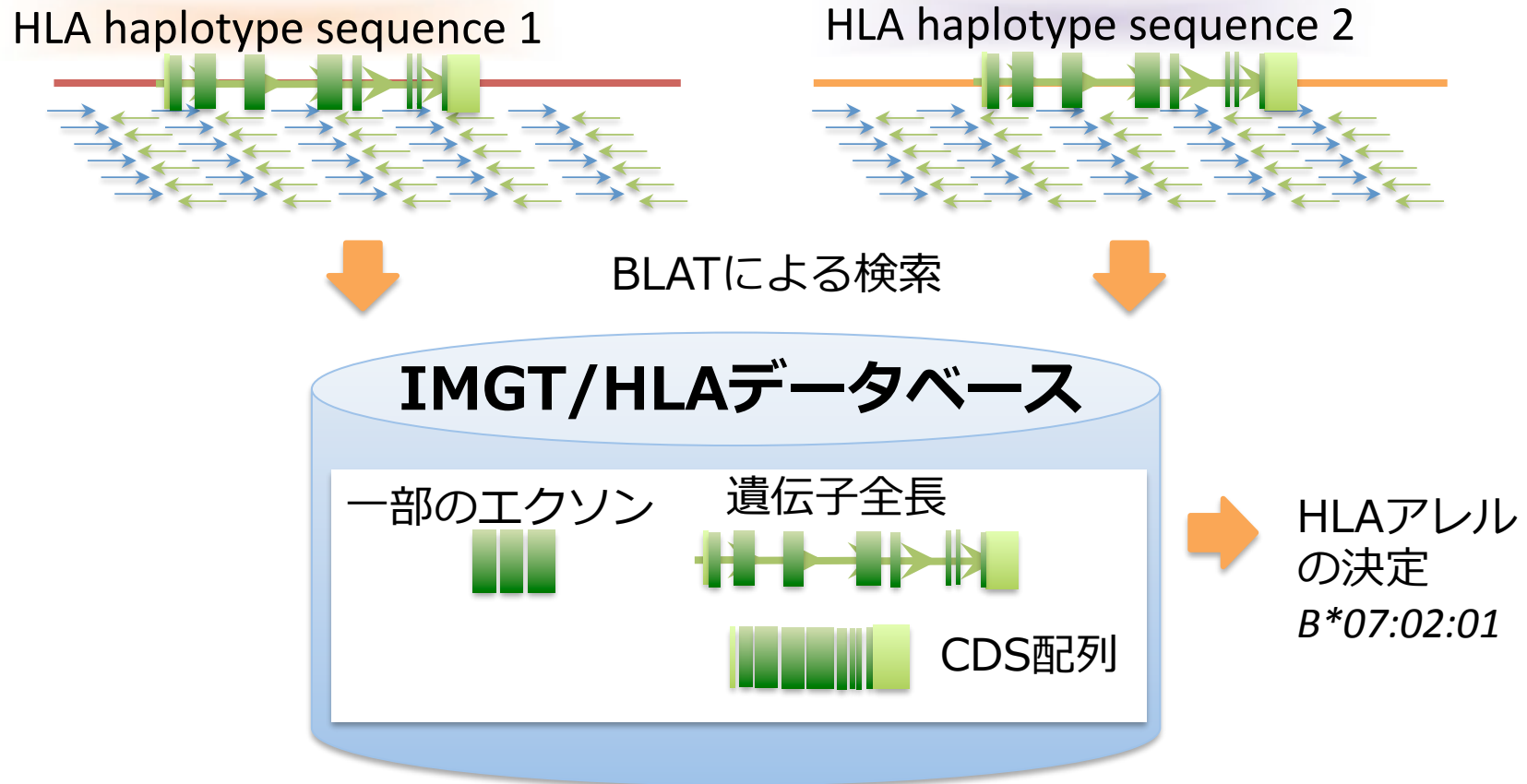
HLA遺伝子配列決定の概要



HLA遺伝子配列決定の概要

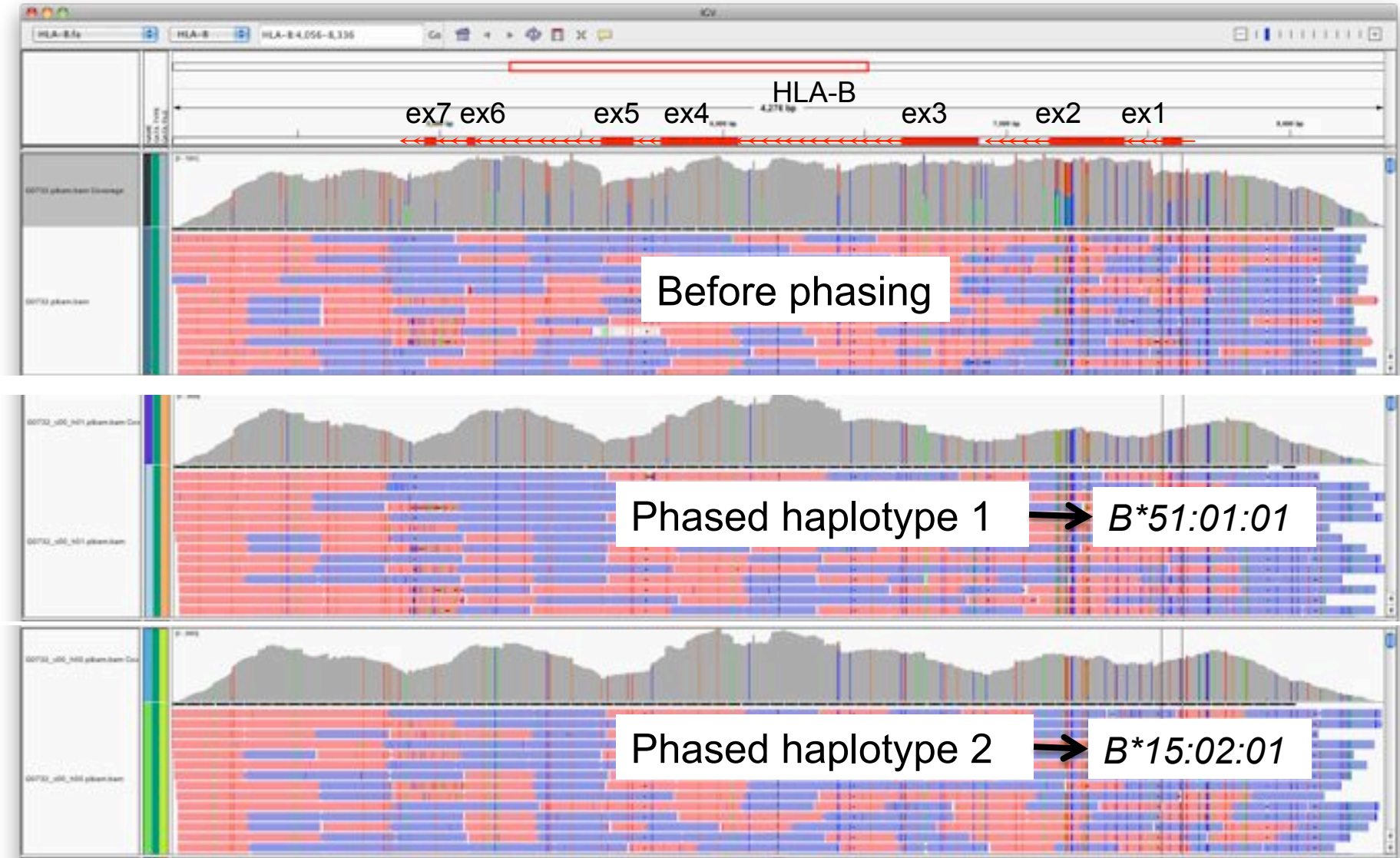


決定したHLA遺伝子完全配列の HLAアレル決定

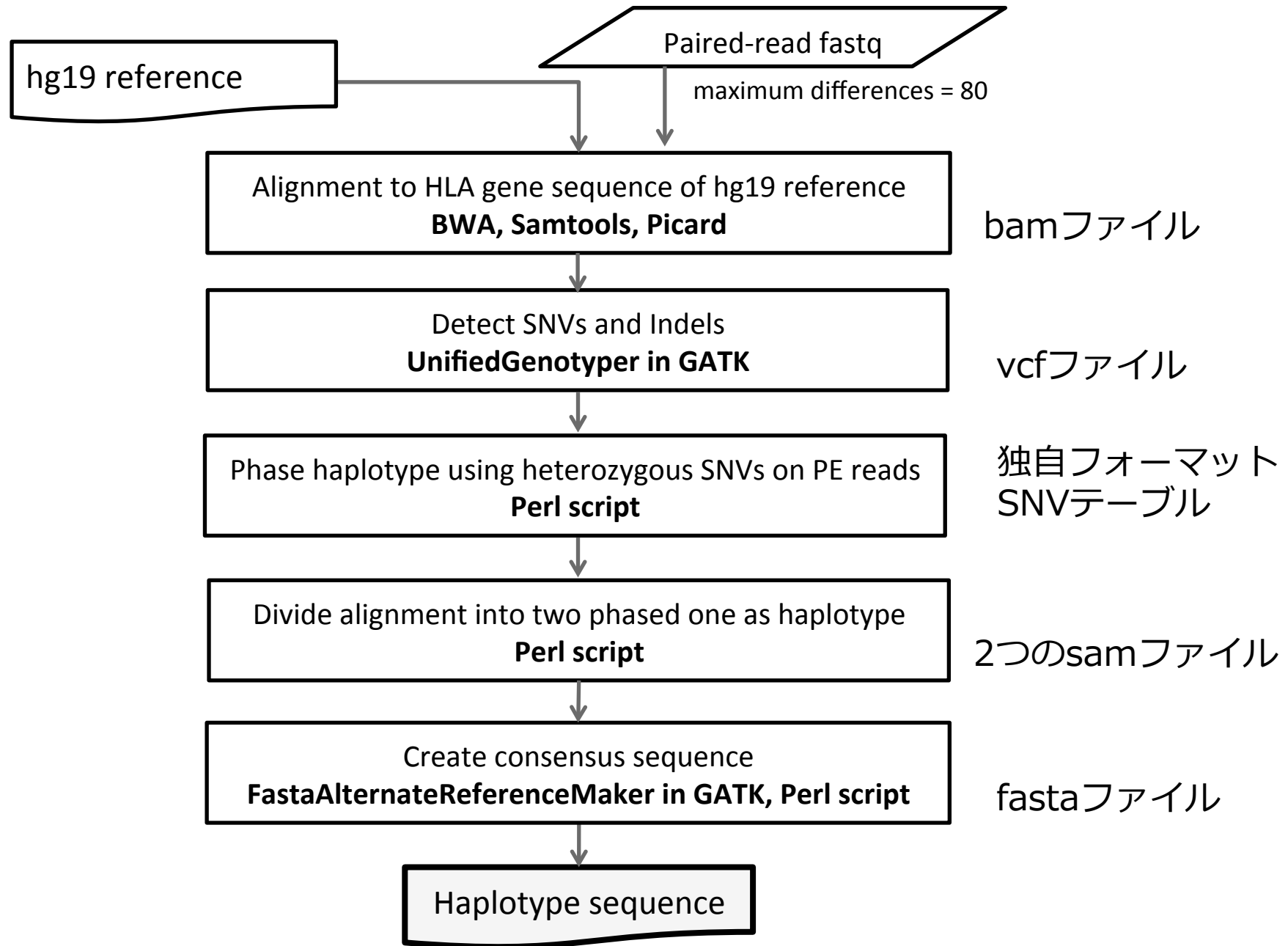


HLAアレルを決定することよりもHLA遺伝子の塩基配列を完全に決定することが本質

Phase-defined sequencingによる HLA遺伝子配列の完全決定



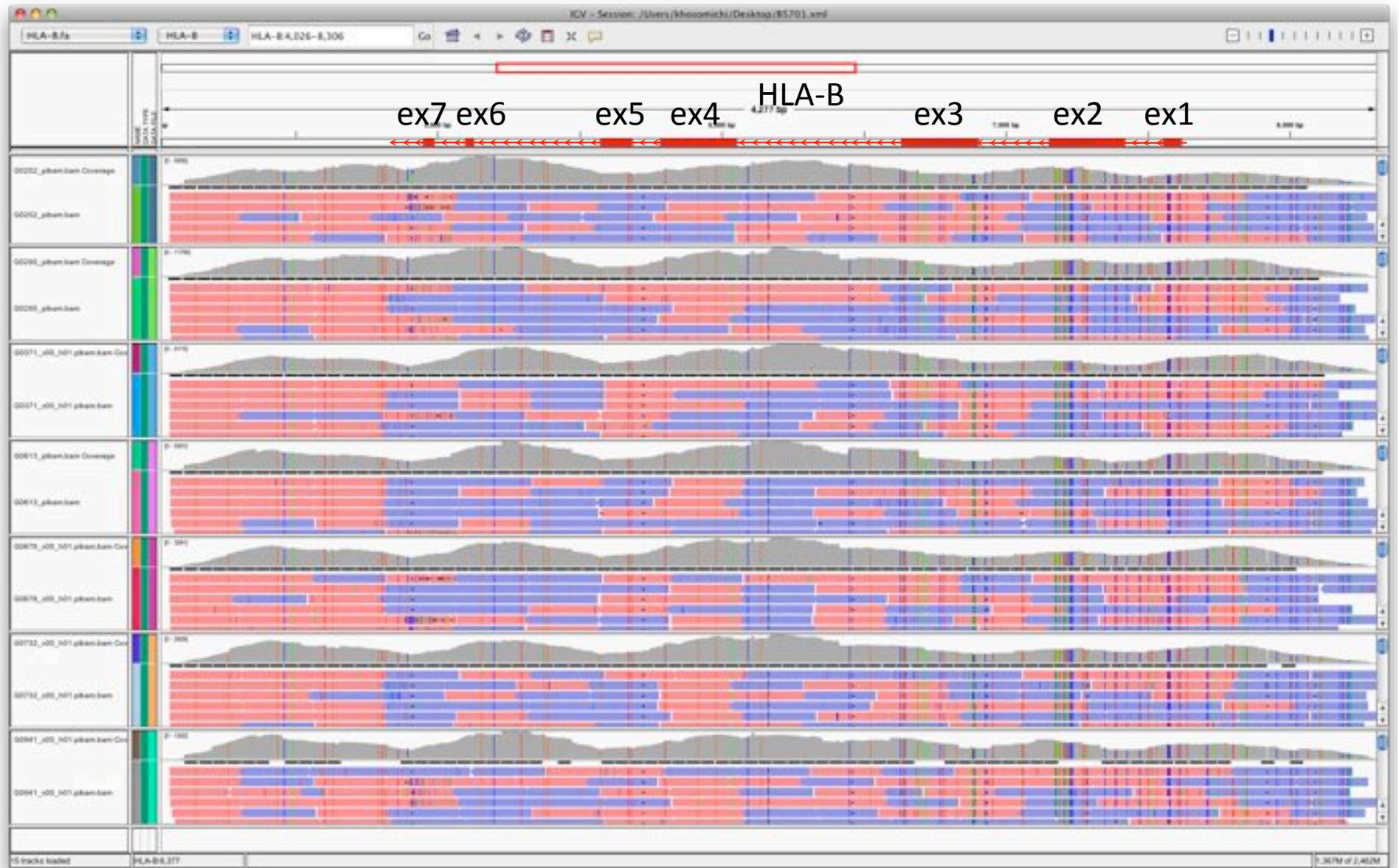
HLAハプロタイプシーケンス決定のワークフロー



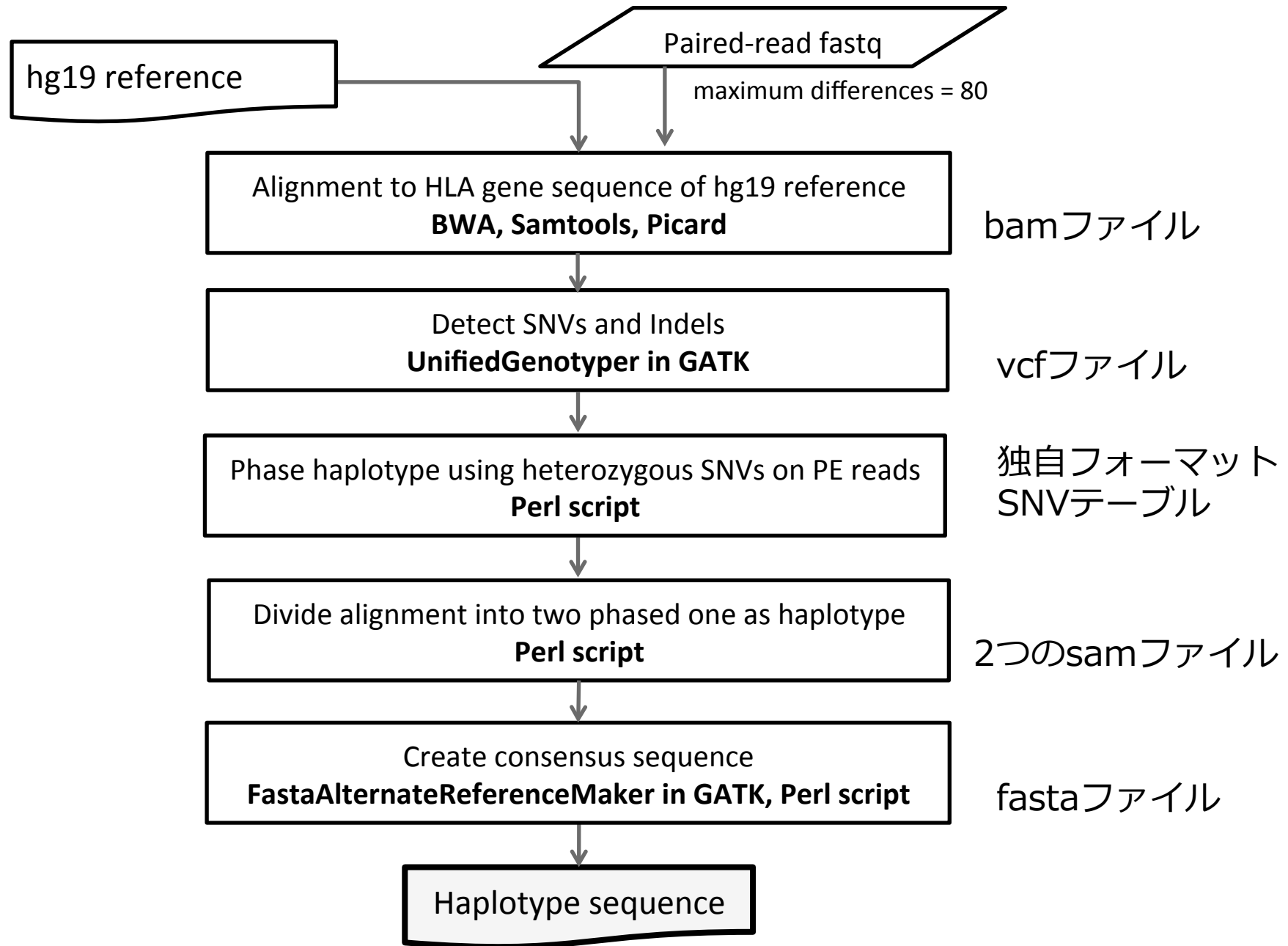
決定したHLA遺伝子配列

Sample ID	PCR-SSO		Next-gen SBT		Average read depth (allele1 / allele2)	Determined sequence (%) (allele1 / allele2)
1	B*51:01:01	B*15:02	B*51:01:01	B*15:02:01	2593.5 / 2431.3	100 / 100
2	B*07:02:01	B*15:02	B*07:02:01	B*15:02:01	3391.2 / 4206.2	100 / 100
3	B*35:01:01	B*15:02	B*35:01:01:01	B*15:02:01	2096.2 / 2431.8	100 / 100
4	B*15:21	B*15:02:01	B*15:21	B*15:02:01	4308.2 / 4693.6	100 / 100
5	B*35:05	B*15:02:01	B*35:05:01	B*15:02:01	4264.3 / 3620.2	100 / 100
6	B*15:18	B*15:02	B*15:18:01	B*15:02:01	1239.9 / 1277.3	100 / 100
7	B*40:06:01:01	B*15:02	B*40:06:01:01	B*15:02:01	5552.1 / 5139.5	100 / 100
8	B*52:01:01	B*15:02	B*52:01:01:01	B*15:02:01	4651.7 / 5085.9	100 / 100
9	B*15:01:01:01	B*57:01:01	B*15:01:01:01	B*57:01:01	2326.4 / 2195.1	100 / 100
10	B*54:01	B*57:01:01	B*54:01:01	B*57:01:01	1999.6 / 2034.6	100 / 100
11	B*40:06:01:01	B*57:01:01	B*40:06:01:01	B*57:01:01	3087.4 / 2833.9	100 / 100
12	B*15:11:01	B*57:01:01	B*15:11:01	B*57:01:01	1948.4 / 1721.9	100 / 100
13	B*15:01:01:01	B*57:01:01	B*15:01:01:01	B*57:01:01	1660.0 / 1482.0	100 / 100
14	B*44:03:01	B*57:01:01	B*44:03:01	B*57:01:01	806.0 / 668.2	100 / 100
15	B*55:07	B*58:01	B*55:07	B*58:01:01	771.6 / 1000.2	100 / 100
16	B*38:01:01	B*58:01	B*38:01:01	B*58:01:01	1565.1 / 2139.7	100 / 100
17	B*48:01:01	B*58:01	B*48:01:01	B*58:01:01	5555.3 / 4610.8	100 / 100
18	B*35:01:01	B*58:01:01	B*35:01:01:01	B*58:01:01	819.5 / 778.1	100 / 100
19	B*15:25:01	B*58:01:01	B*15:25:01	B*58:01:01	3441.2 / 3852.0	100 / 100
20	B*54:01	B*58:01:01	B*54:01:01	B*58:01:01	1308.3 / 1454.1	100 / 100
21	B*51:02:01	B*58:01:01	B*51:02:01	B*58:01:01	1464.0 / 1132.7	100 / 100
22	B*51:01:01	B*58:01:01	B*51:01:01	B*58:01:01	462.9 / 636.3	100 / 100
23	B*51:02:01	B*58:01:01	B*51:02:01	B*58:01:01	637.0 / 663.7	100 / 100
24	B*39:23	B*58:01:01	B*39:23	B*58:01:01	2112.6 / 2553.3	100 / 100

HLA-B*58:01 の塩基配列の比較

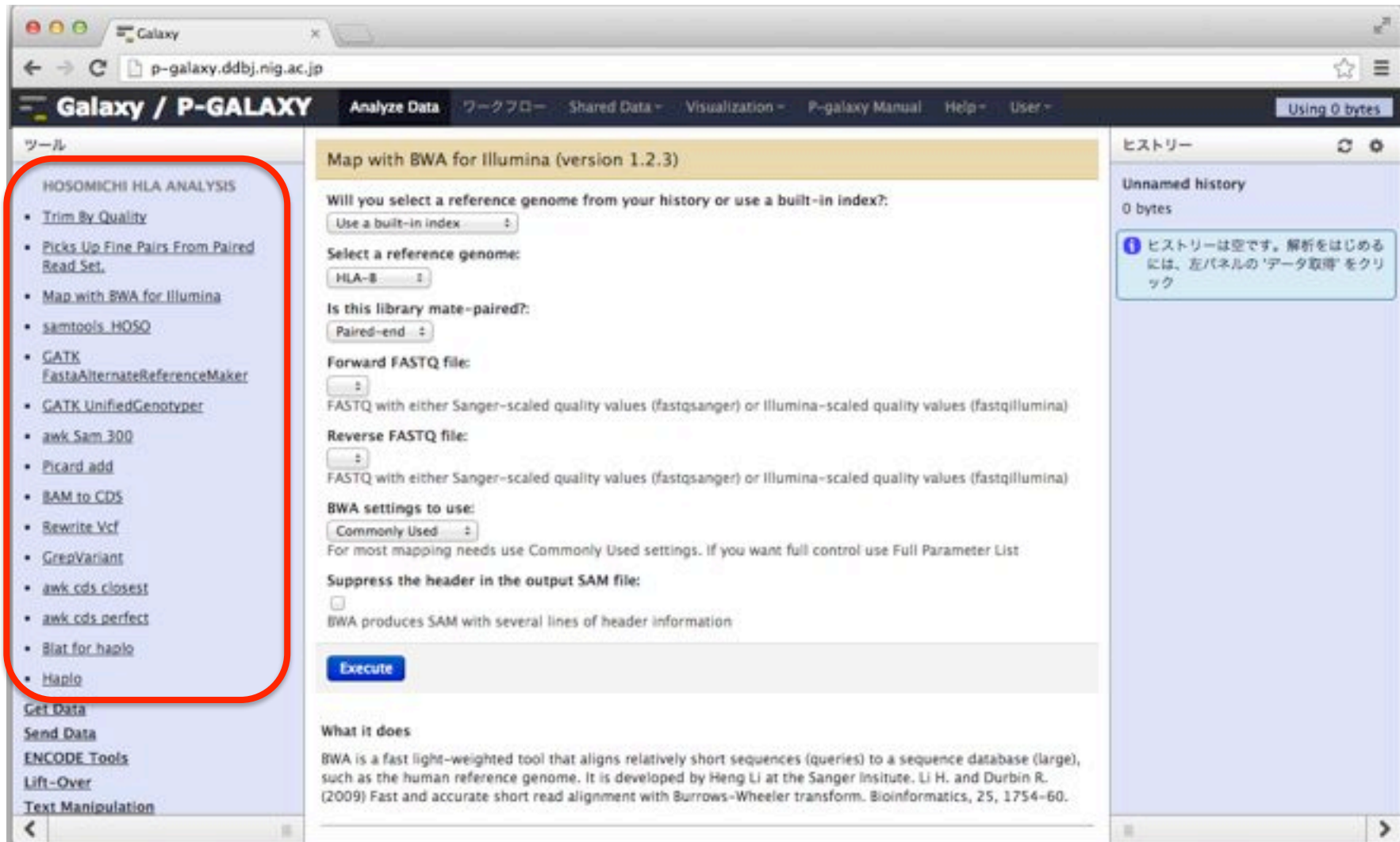


HLAハプロタイプシーケンス決定のワークフロー



HLA解析ツール (1)

- p-galaxy <http://p-galaxy.ddbj.nig.ac.jp>



Galaxy / P-GALAXY Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User Using 0 bytes

ツール

HOSOMICHI HLA ANALYSIS

- [Trim By Quality](#)
- [Picks Up Fine Pairs From Paired Read Set](#)
- [Map with BWA for Illumina](#)
- [samtools_HOSO](#)
- [GATK FaataAlternateReferenceMaker](#)
- [GATK UnifiedGenotyper](#)
- [awk_Sam_300](#)
- [Picard_add](#)
- [BAM to CDS](#)
- [Rewrite Vcf](#)
- [GrepVariant](#)
- [awk_cds_closest](#)
- [awk_cds_perfect](#)
- [Blat_for_haplo](#)
- [Haplo](#)

Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation

Map with BWA for Illumina (version 1.2.3)

Will you select a reference genome from your history or use a built-in index?:

Select a reference genome:

Is this library mate-paired?:

Forward FASTQ file:

FASTQ with either Sanger-scaled quality values (fastqsanger) or Illumina-scaled quality values (fastqillumina)

Reverse FASTQ file:

FASTQ with either Sanger-scaled quality values (fastqsanger) or Illumina-scaled quality values (fastqillumina)

BWA settings to use:

For most mapping needs use Commonly Used settings. If you want full control use Full Parameter List

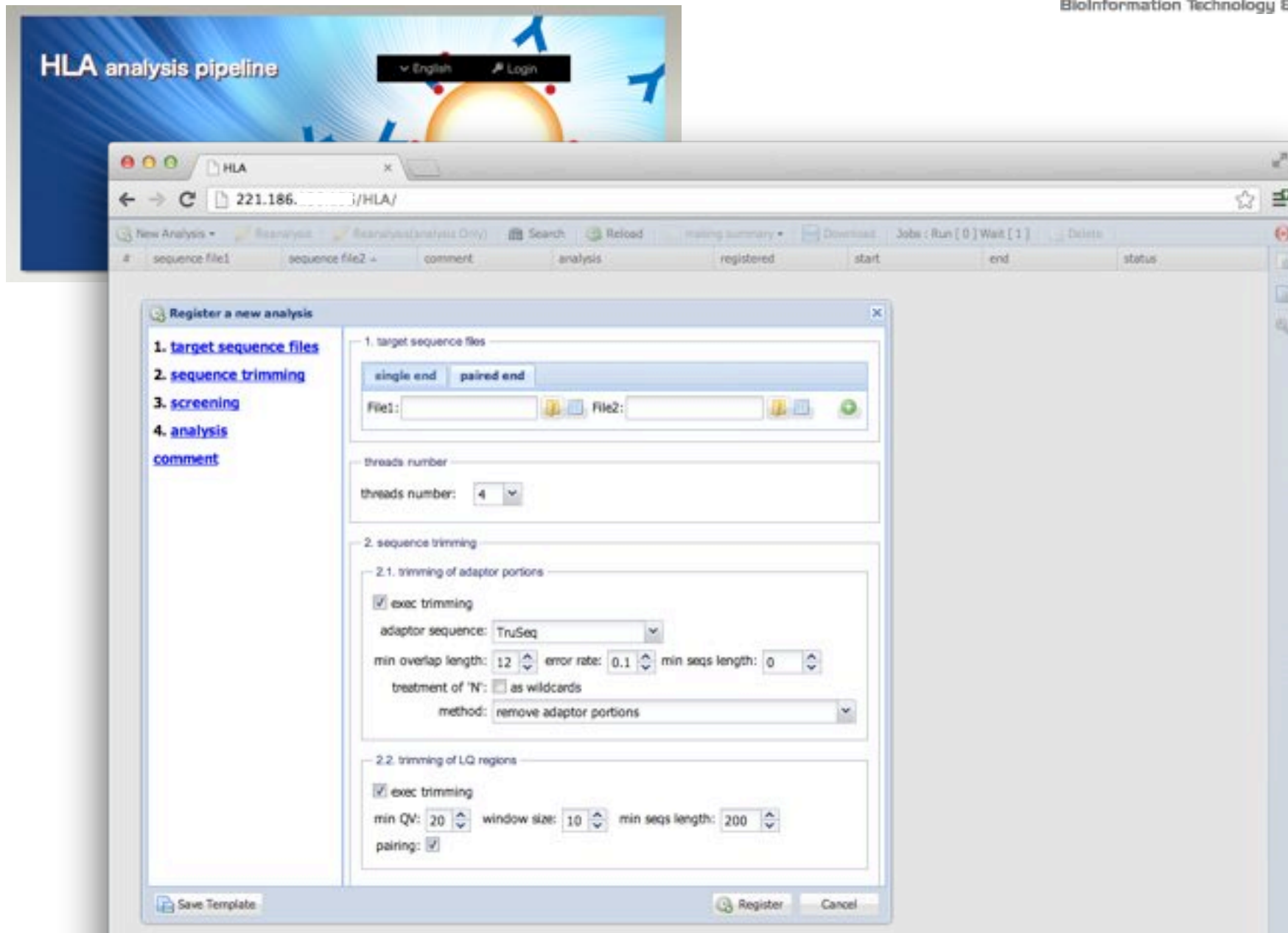
Suppress the header in the output SAM file:

BWA produces SAM with several lines of header information

What it does
BWA is a fast light-weighted tool that aligns relatively short sequences (queries) to a sequence database (large), such as the human reference genome. It is developed by Heng Li at the Sanger Institute. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-60.

ヒストリー
Unnamed history
0 bytes
ヒストリーは空です。解析をはじめするには、左パネルの 'データ取得' をクリック

HLA解析ツール (2)



The screenshot displays the HLA analysis pipeline web interface. At the top left, a banner reads "HLA analysis pipeline" with a "Login" button. Below it, a browser window shows the URL "221.186. /HLA/". The main content area features a "Register a new analysis" dialog box with the following settings:

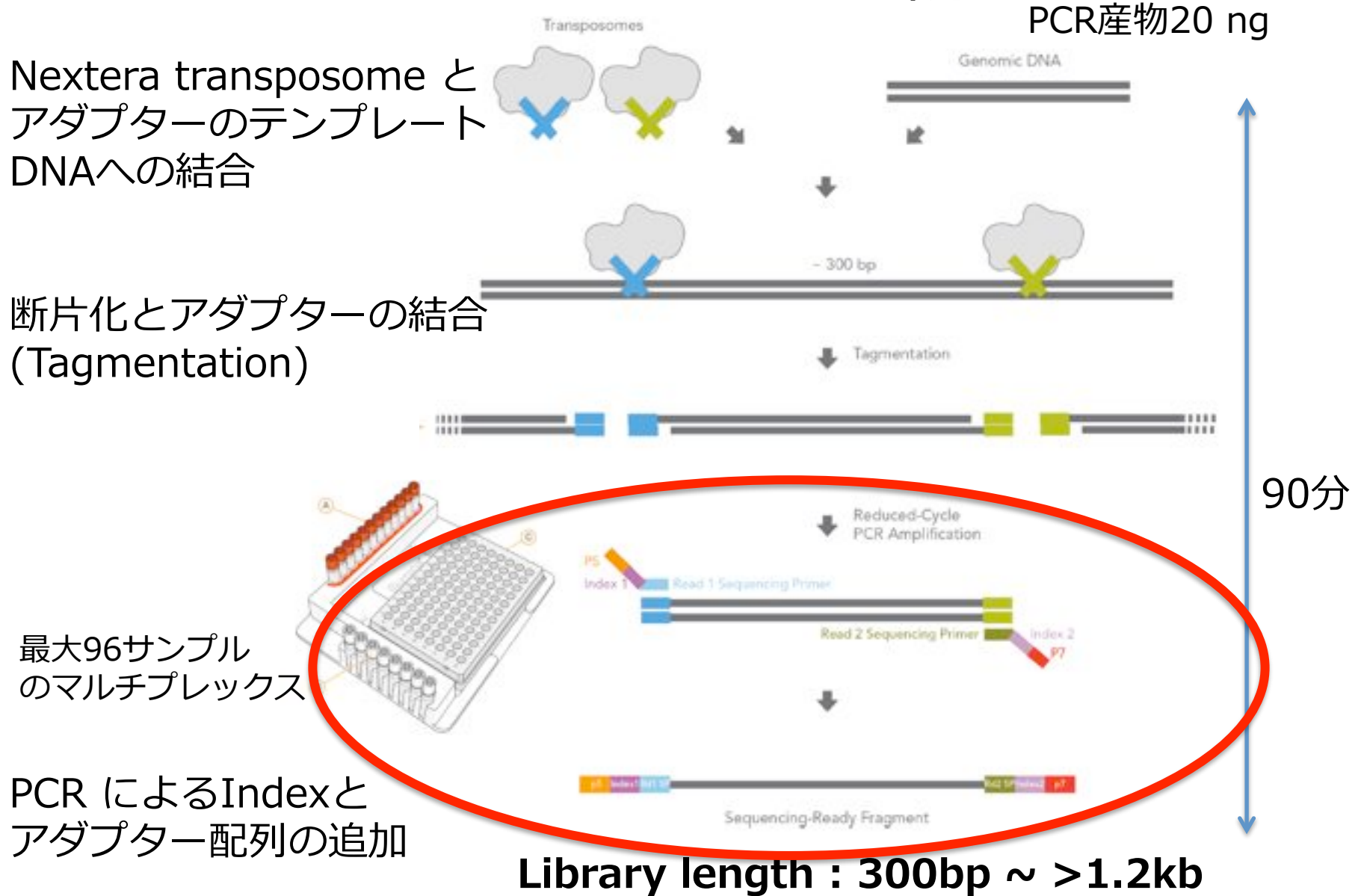
- 1. target sequence files**
 - single end | paired end
 - File1: [] File2: []
- threads number**
 - threads number: 4
- 2. sequence trimming**
 - 2.1. trimming of adaptor portions**
 - exec trimming
 - adaptor sequence: TruSeq
 - min overlap length: 12 error rate: 0.1 min seqs length: 0
 - treatment of 'N': as wildcards
 - method: remove adaptor portions
 - 2.2. trimming of LQ regions**
 - exec trimming
 - min QV: 20 window size: 10 min seqs length: 200
 - pairing:

Buttons at the bottom of the dialog include "Save Template", "Register", and "Cancel".

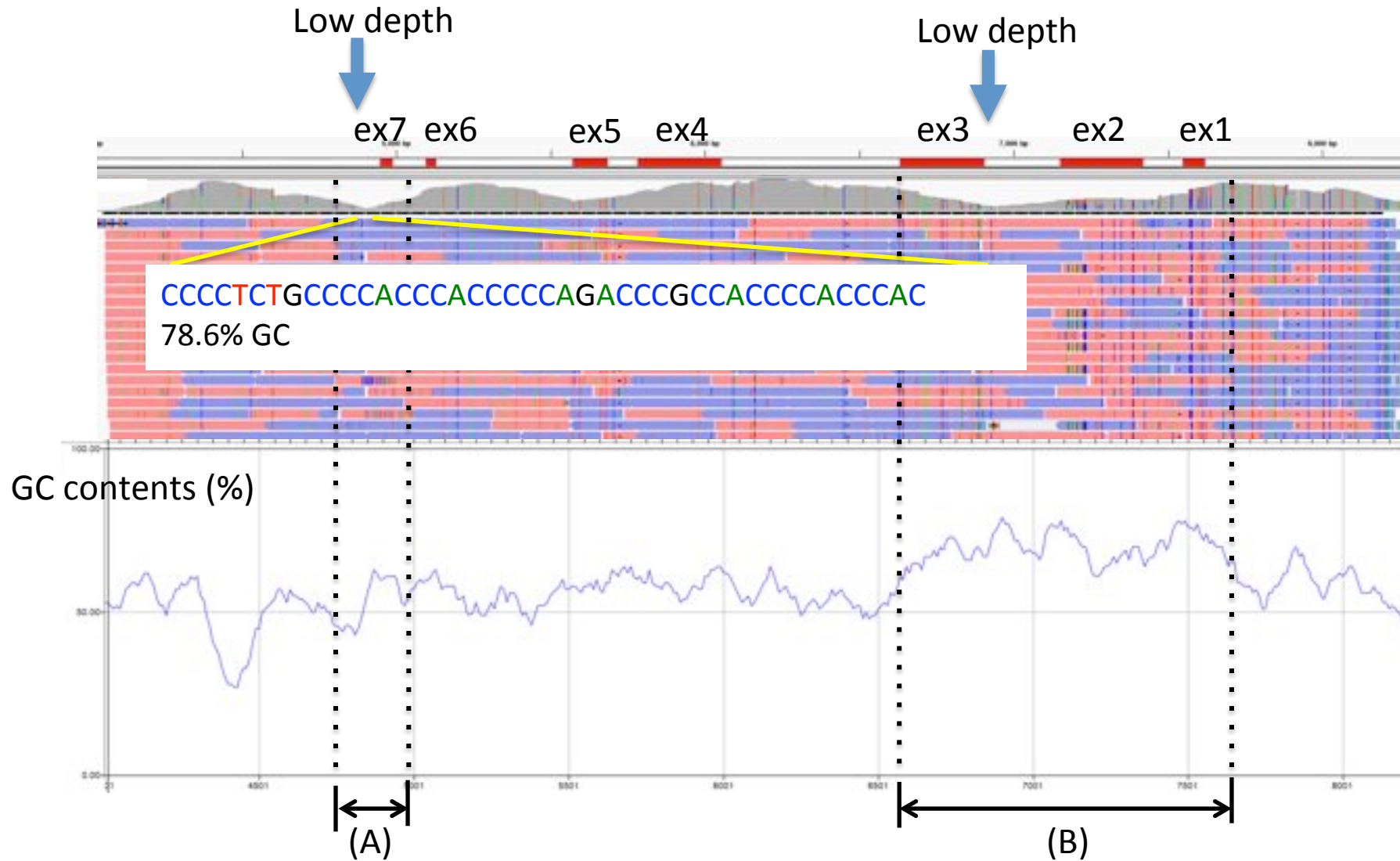
トピックス

- ロングPCRとNextera Kitによる
HLA遺伝子の配列決定
- Kapaエンジニア酵素による
ライブラリー調整における改良事例
- AMPure XPによる
ライブラリーのノーマライゼーション

Nextera DNA Sample Preparation Kit によるライブラリ調整



HLA遺伝子のGC%とRead depth



高GC%の領域では得られるread depthが低い

Kapaエンジニア酵素による プロトコールの改良

ロングPCR増幅

↓
Tagmentation

PCR産物(10ng)	4u μ l	55°C	5min
TD	5u μ l	10°C	∞
TDE1	1u μ l		

↓
ライブラリーのPCR増幅

Index1 primer	1u μ l	72°C	3min	} 7cycles
Index2 primer	1u μ l	98°C	30sec	
2xKapa HiFi HS RM	5u μ l	98°C	10sec	
PPC	1u μ l	63°C	30sec	
Tagmentation DNA	2u μ l	72°C	3min	
		10°C	∞	

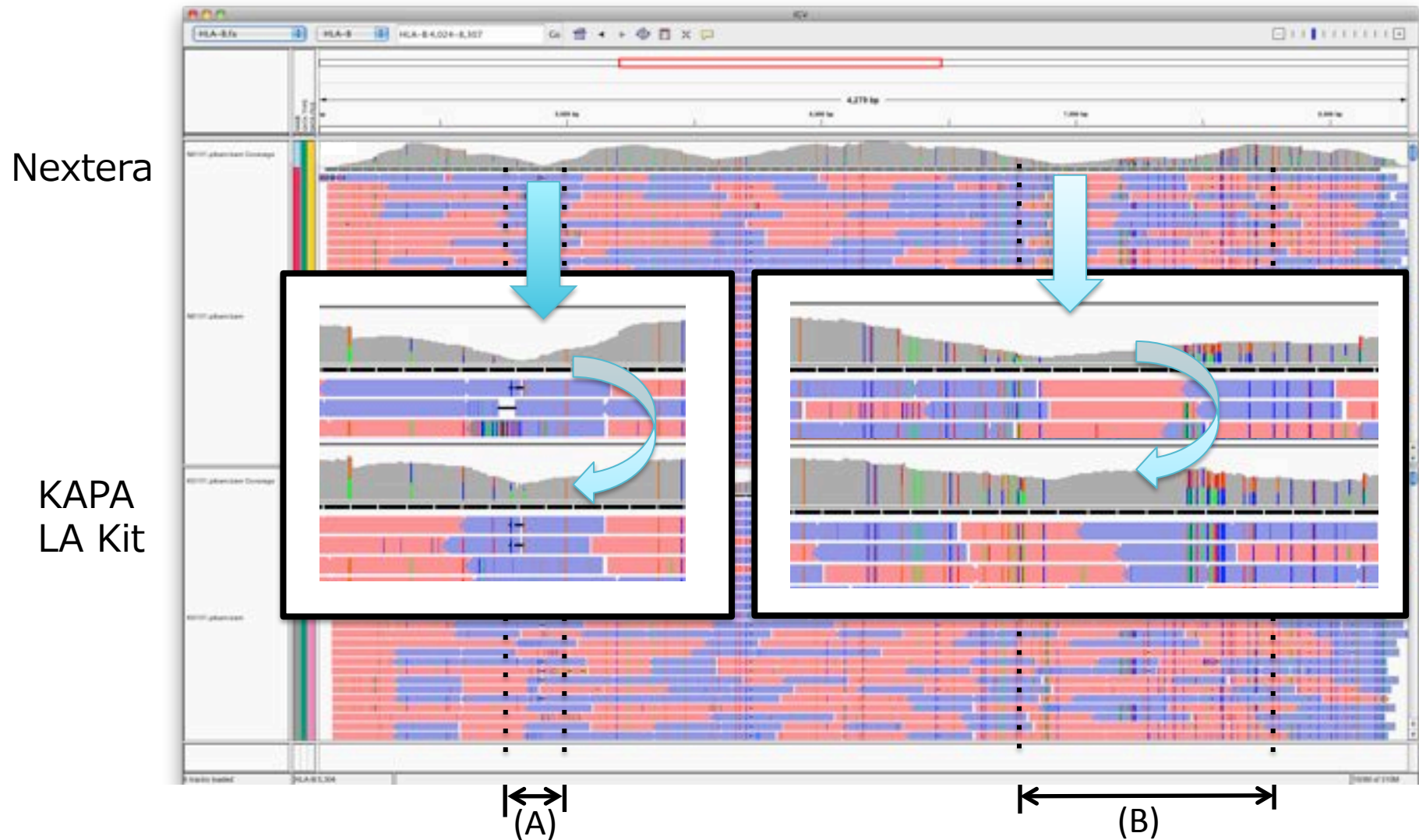
↓
精製 AMPure XP bead

PCR酵素

KAPA BIOSYSTEMS KAPA Library Amplification Kit KK2611

KAPA BIOSYSTEMS Kapa HiFi Hotstart mastermix KK2602

HLA遺伝子のGC%とRead depth



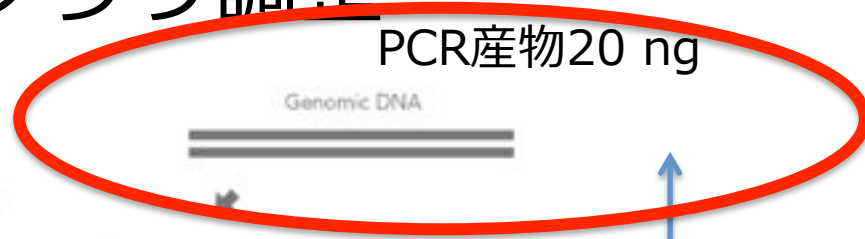
KAPA LA kitでは高GC%での低read depthが改善

トピックス

- ロングPCRとNextera Kitによる
HLA遺伝子の配列決定
- Kapaエンジニア酵素による
ライブラリー調整における改良事例
- AMPure XPによる
ライブラリーのノーマライゼーション

Nextera DNA Sample Preparation Kit によるライブラリ調整

Nextera transposome と
アダプターのテンプレート
DNAへの結合



断片化とアダプターの結合
(Tagmentation)

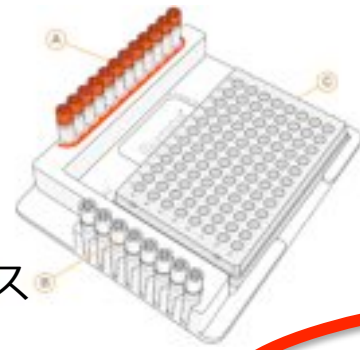


スタートPCR産物量と
インサート長の関係

PCR Amplification

90分

最大96サンプル
のマルチプレックス



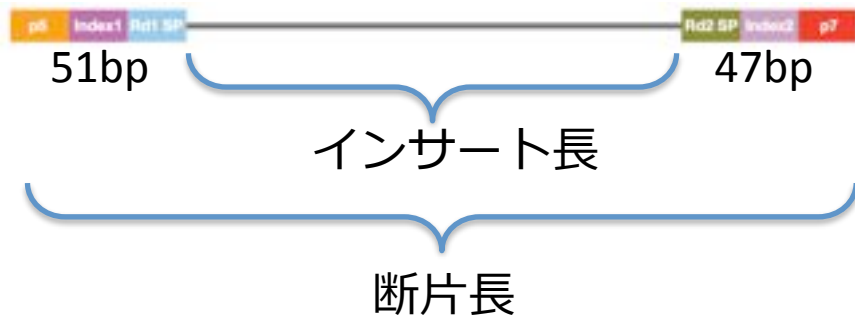
PCR によるIndexと
アダプター配列の追加



Library length : 300bp ~ >1.2kb

Nexteraライブラリー長

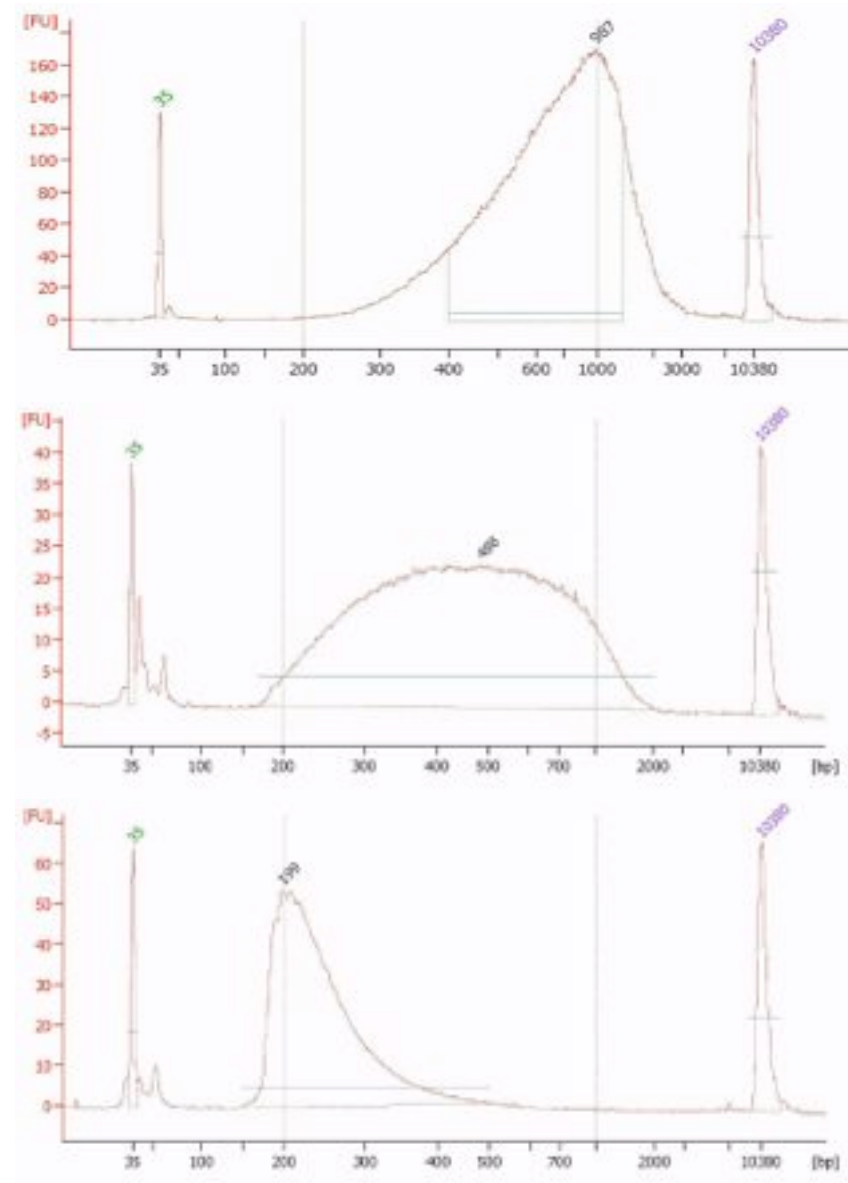
使用するPCR産物の量により
Nexteraライブラリーの長さが異なる



Index primer 1 (i7) 39bp + 8bp
5' CAAGCAGAAGACGGCATAACGAGAT[i7]GTCTCGTGGGCTCGG

Index primer 2 (i5) 43bp + 8bp
5' AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC

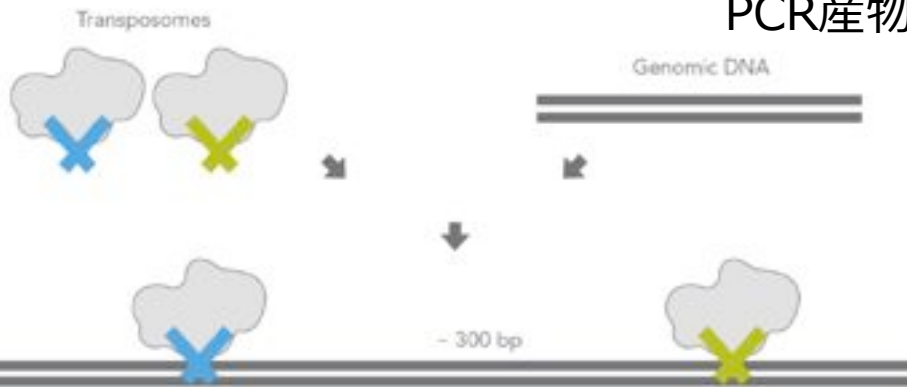
断片長 - 98bp = インサート長



Nextera DNA Sample Preparation Kit によるライブラリ調整

PCR産物20 ng

Nextera transposome と
アダプターのテンプレート
DNAへの結合

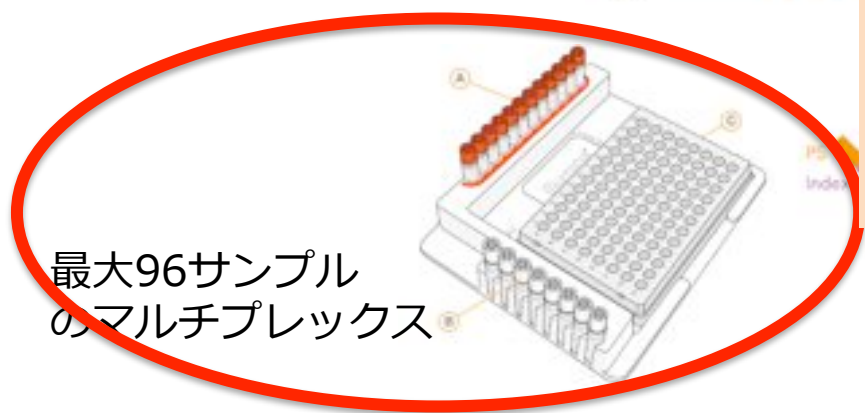


断片化とアダプターの結合
(Tagmentation)



プールする96サンプルの
ライブラリーのモル濃度
をそろえる必要

90分



最大96サンプル
のマルチプレックス

PCR によるIndexと
アダプター配列の追加



Library length : 300bp ~ >1.2kb

複数のNexteraライブラリーの モル濃度調整は手間がかかる

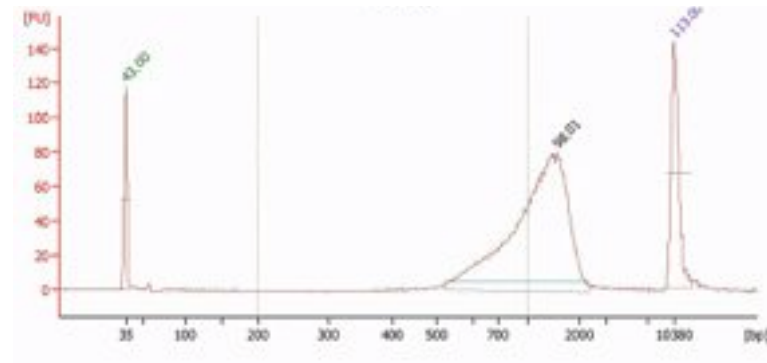
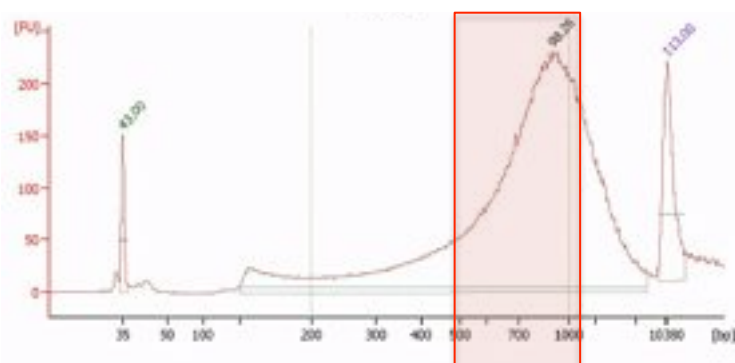
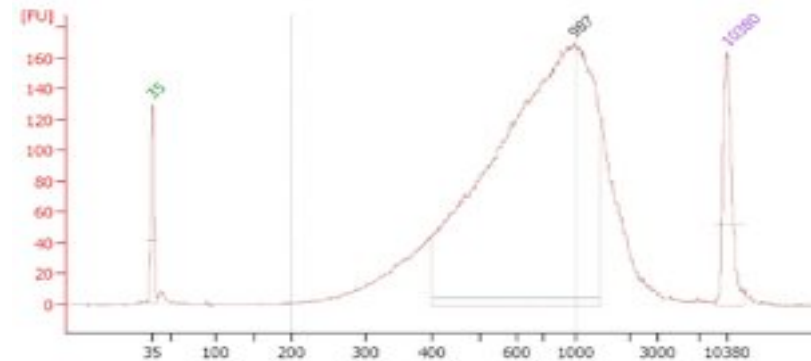
ゲル切り出しサイズでの
モル濃度の測定



モル濃度に基づく複数ライブラリーのプール

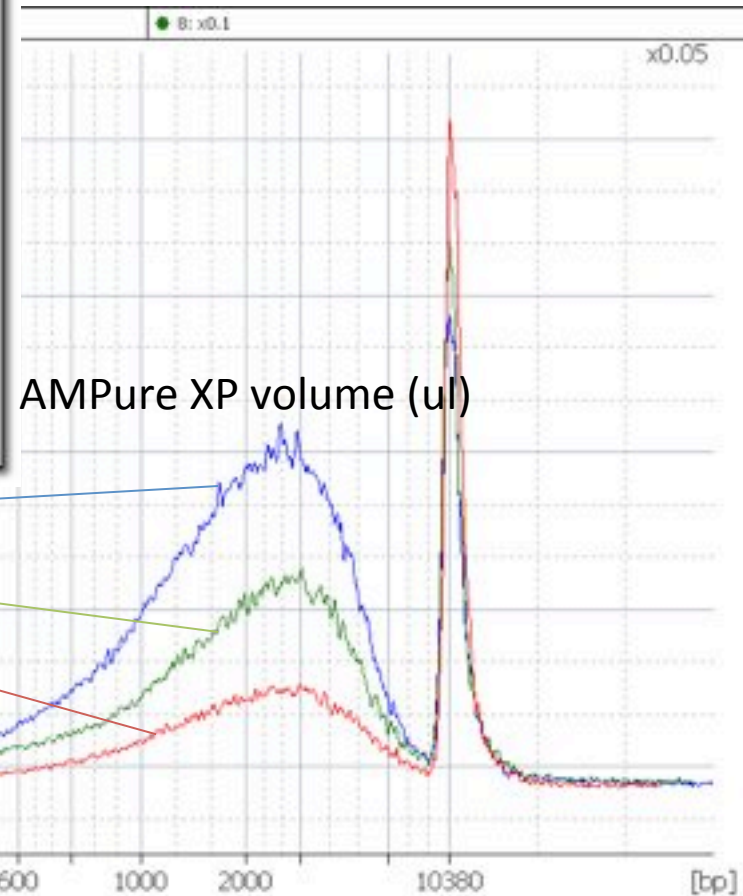
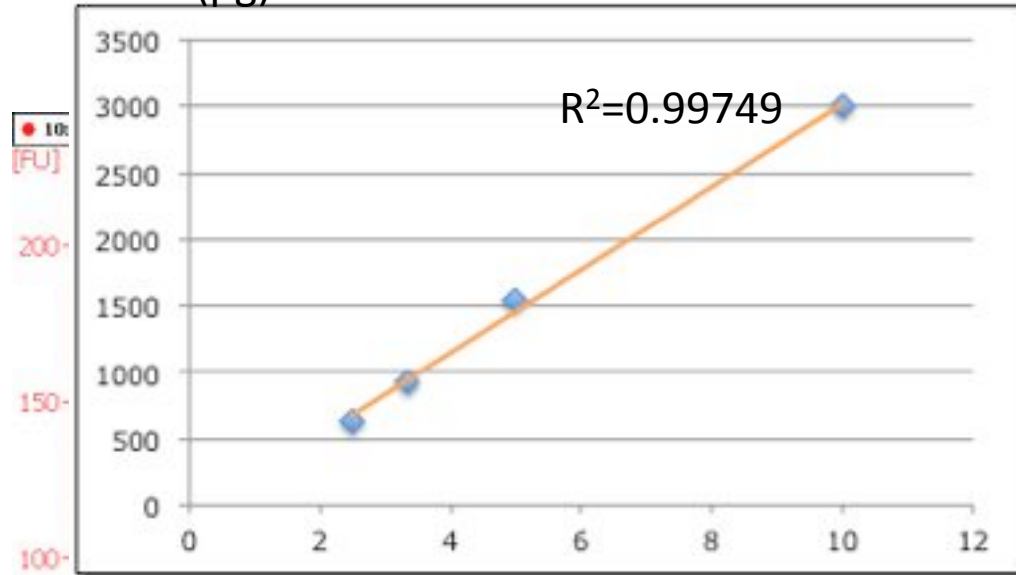


ゲル切り出しと切り出したライブラリープールの再測定



AMPure XPビーズによるDNA濃度の のノーマライゼーション

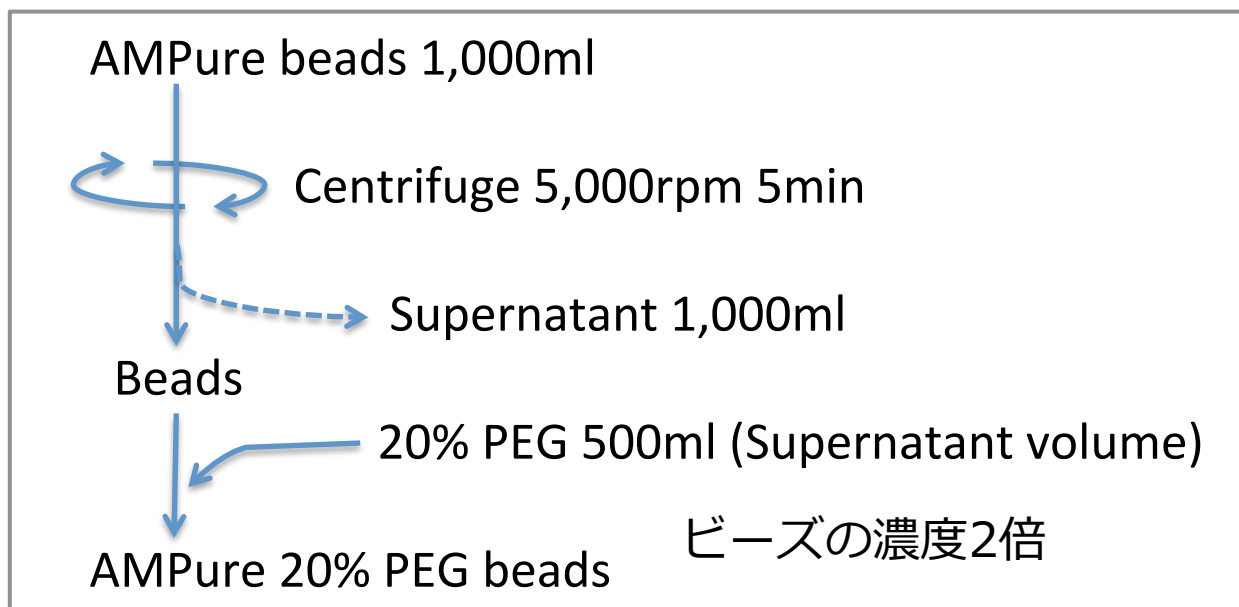
Bound DNA (pg)



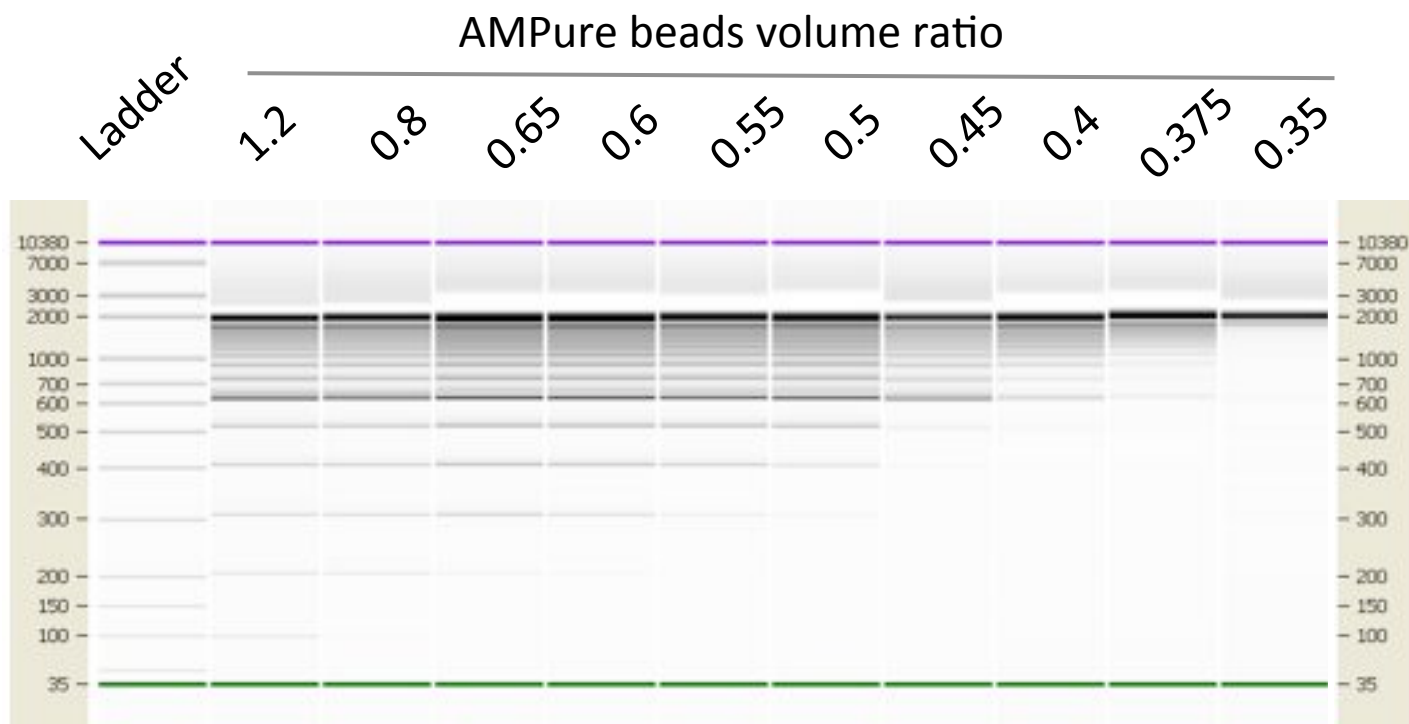
ライブラリーのモル濃度をそろえる為にはライブラリー長もそろえる必要がある

サイズセレクション用 AMPure XP beads浮遊液

- 20% PEG8000 and 2.5M NaCl
 - 10g polyethylene glycol (PEG) 8000
 - 7.3g NaCl
 - the volume to 50 ml with MilliQ water



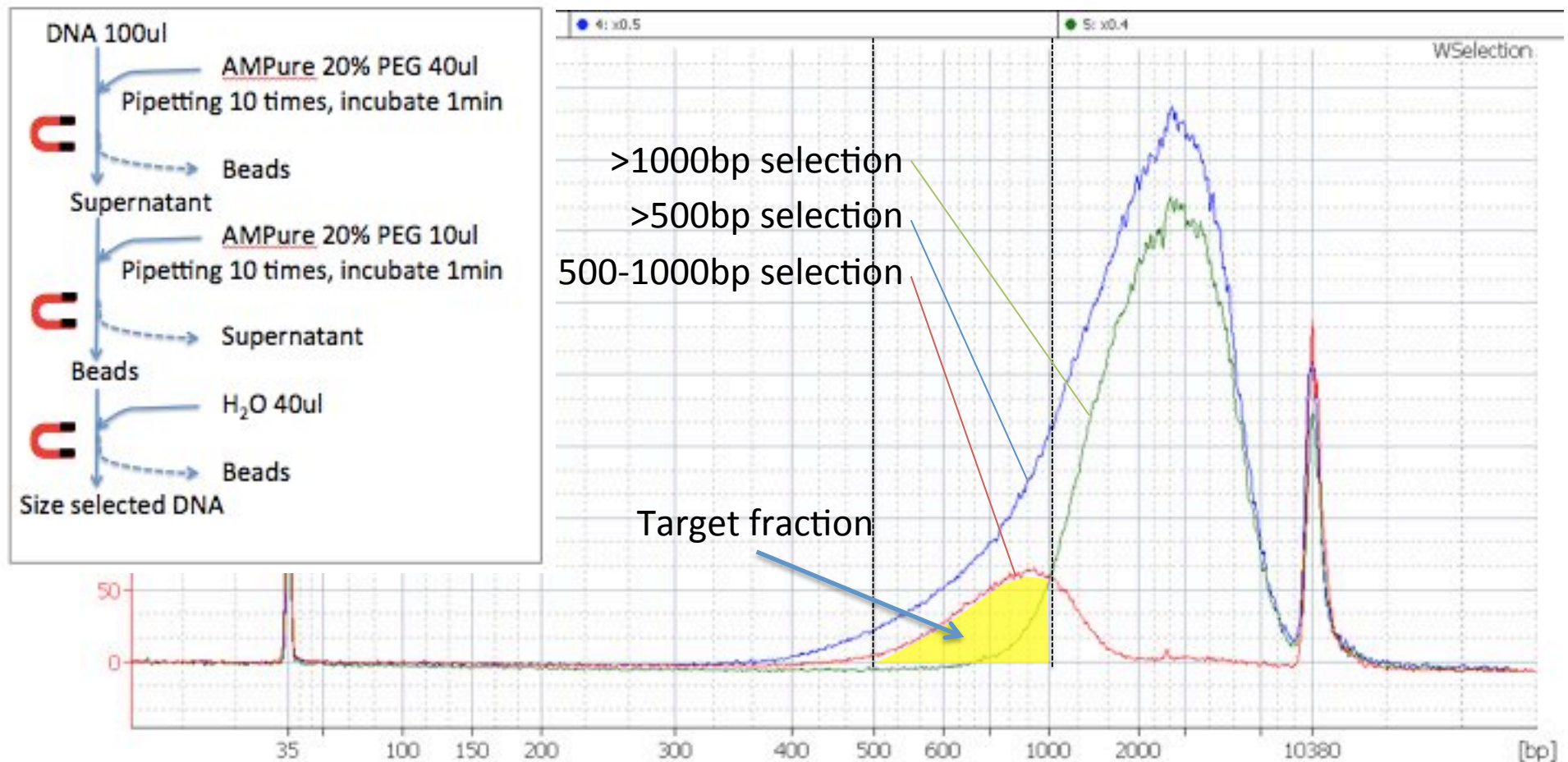
AMPure XP ビーズによるサイズ選択



Beads ratio	Selected length
1.2x	>100bp
0.8x	>200bp
0.6x	>300bp
0.5x	>500bp
0.45x	>600bp
0.35x	>1500bp

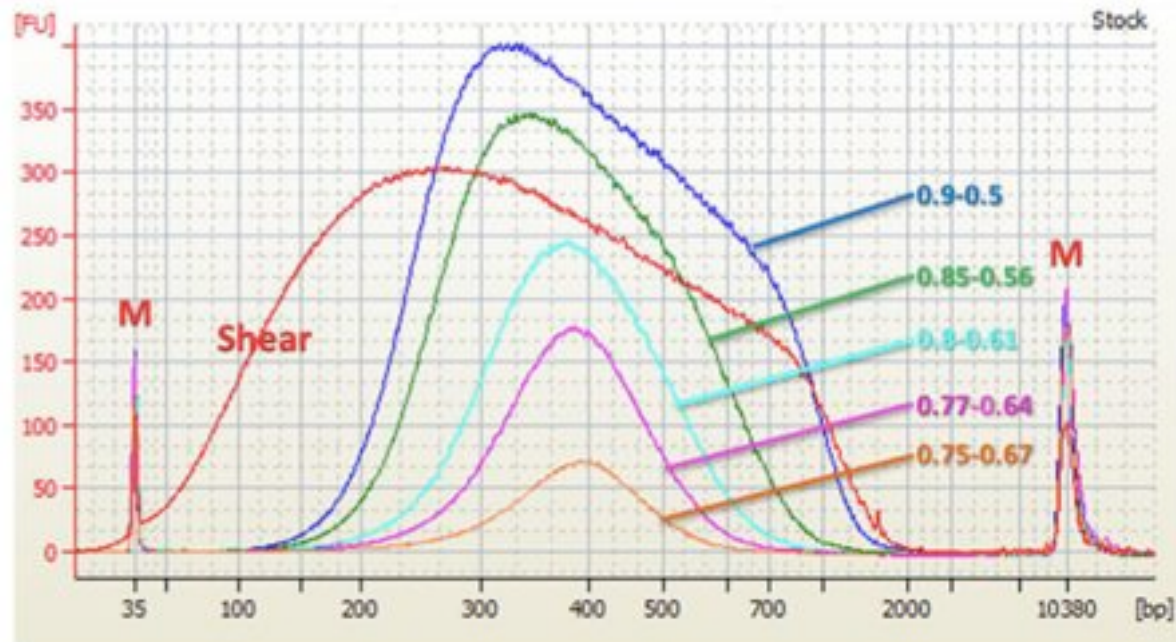
DNAに対するAMPure XP ビーズの割合を変えることで、結合するDNAの長さの範囲を選択することが可能

AMPure XP ビーズによるNexteraライブラリのサイズセレクション



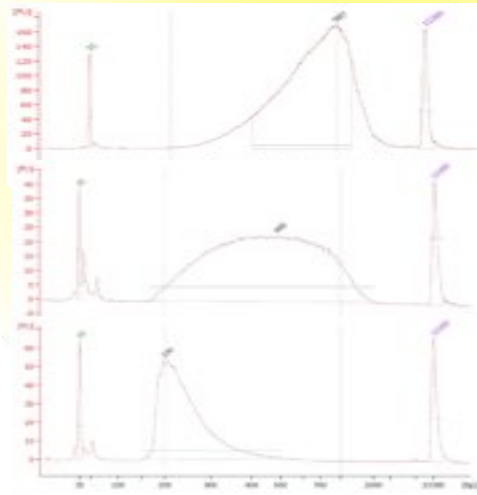
加えるビーズ量によりPEGの濃度が変わり、得られるDNA断片長をコントロールできる

SPRIselect reagent kit

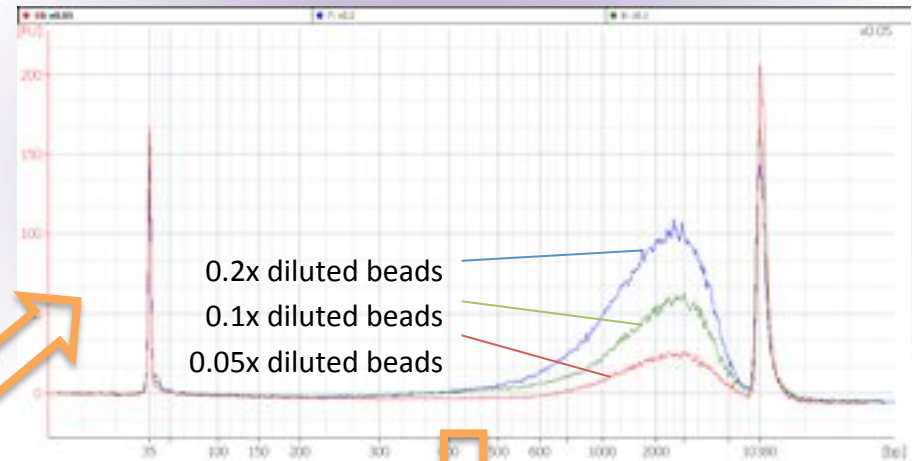


AMPure XP ビーズによるライブラリーのノーマライゼーション

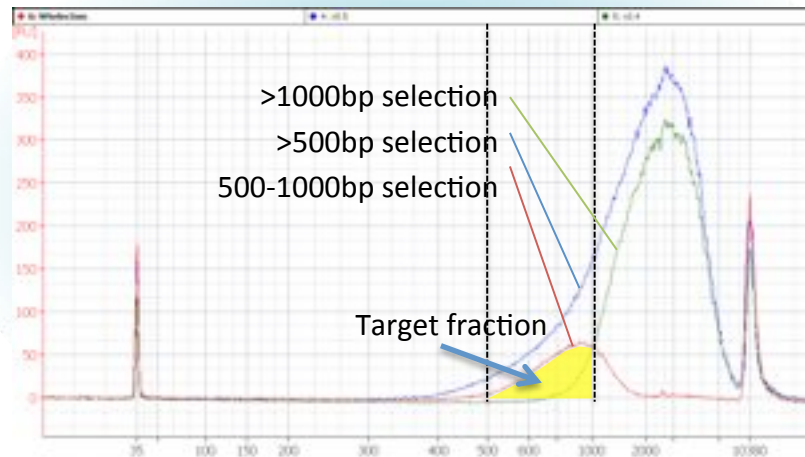
様々なインサートサイズのライブラリー



ライブラリーのDNA濃度のノーマライゼーション



ライブラリーのサイズセレクション



ライブラリーのサイズと濃度の調整することでモル濃度を均一化
シーケンスリード数を均一化

Collaborators and Acknowledgments

National Institute of Genetics Division of Human Genetics

Ituro INOUE,

Hirofumi NAKAOKA, Timothy A JINAM, Junko KAJIWARA,
Ryota SUGIMOTO

Genome Informatics Laboratory

Hideki NAGASAKI, Eli KAMINUMA

BITS(BioInformation Technology & Science)

Tadasu SHINI

Japanese Red Cross Society

Fumiaki NAKAJIMA, Marie SHIMIZU

Tokai University School of Medicine

Department of Molecular Life Sciences

Hidetoshi INOKO, Takashi SHIINA,

Shingo SUZUKI, Shigeki MITSUNAGA, Yuko OKUDAIRA

Education and Research Support Center

Masayuki TANAKA

