

次世代シーケンサーを用いた
HLA タイピングシステムの開発

瀬川 秀伸

奈良先端科学技術大学院大学

バイオサイエンス研究科 腫瘍細胞生物学研究室

(加藤 順也 教授)

平成 29 年 2 月 17 日 提出

1. 序論

1-1. HLA について

主要組織適合性複合体 (major histocompatibility complex : MHC) は、自己・非自己を決める、つまり移植の組織適合性を支配する遺伝子群としてマウスで 1930 年、1940 年代に Gorer¹⁾ と Snell²⁾ によって同定された。

そして、1950 年代に、Dausset³⁾ によりヒトの MHC は第 6 番染色体短腕部 6p21.31 上に存在することが明らかにされ、ヒト白血球抗原 (human leucocyte antigen : HLA) と呼ばれるようになった^{4)~6)}。その後の研究で、ヒトの細胞や組織の自己・非自己を決定する因子が HLA であることが分かり、T 細胞への抗原提示に関与することが明らかにされた。

1-2. HLA の構造・機能

HLA は T 細胞受容体を介し、T 細胞に抗原を提示する役割を担っている。T 細胞受容体は、HLA の一部と抗原を複合体として同時に認識し、細胞内へその情報を伝えて T 細胞を活性化させる。T 細胞への抗原提示には HLA クラス I と HLA クラス II が関与する⁷⁾ (図 1)。

HLA クラス I は、ほとんどの有核細胞の表面で発現し、内在性抗原を細胞傷害性 T 細胞へ提示する機能を持つ。内在性抗原とは、免疫細胞から攻撃を受けないようにするための自己ペプチド、または、ウイルスや細菌など細胞内に入り込み増殖する非自己ペプチドである。HLA クラス I は、細胞外領域が $\alpha 1, \alpha 2, \alpha 3$ ドメインからなる α 鎖と、 $\beta 2$ ミクログロブリン ($\beta 2m$) からなる β 鎖が非共有結合したヘテロ二量体である。 $\alpha 1, \alpha 2$ ドメインは細胞膜から最も外側に位置し、抗原が結合するペプチド収容溝を構成する^{7),8)} (図 1)。

HLA クラス II は、マクロファージ、樹状細胞などの抗原提示細胞の表面に発現し、外来性抗原をヘルパー T 細胞へ提示する機能を持つ。外来性抗原とは、抗原提示細胞により認識され取り込まれた細胞外に存在する非自己のペプチドである。HLA クラス II は、細胞外領域が $\alpha 1, \alpha 2$ ドメインからなる α 鎖と $\beta 1, \beta 2$ ドメインからなる β 鎖が非共有結合したヘテロ二量体である。 $\alpha 1, \beta 1$ ドメインは細胞膜から最も外側に位置し、抗原が結合するペプチド収容溝を構成する^{7),8)} (図 1)。

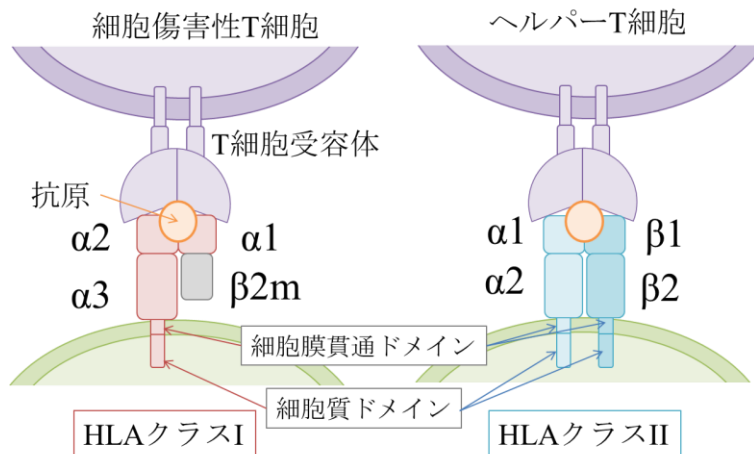


図1：HLAの構造とT細胞への抗原提示

HLAクラスI：α鎖（α1ドメイン、α2ドメイン、α3ドメイン）、β2m：β2ミクログロブリン
 HLAクラスII：α鎖（α1ドメイン、α2ドメイン）、β鎖（β1ドメイン、β2ドメイン）
 HLAクラスI, IIそれぞれで細胞膜貫通ドメインと細胞質ドメインを示す。

1-3. HLA 遺伝子座と多型

HLAクラスIにおいてT細胞への抗原提示に関与する遺伝子座としてHLA-A、-B、-Cが存在し、α鎖をコードする。HLAクラスIIにおいてT細胞への抗原提示に関与する遺伝子座として、α鎖をコードする遺伝子座（HLA-DRA、-DQA1、-DPA1）と、β鎖をコードする遺伝子座（HLA-DRB1、-DRB3/4/5、-DQB1、-DPB1）が存在する^{5),9)}。

HLA遺伝子はヒトの遺伝子のなかで多型性に富む遺伝子のうちのひとつである。1999年にHLA遺伝子群の約360万塩基対の塩基配列が決定された¹⁰⁾。それから様々な集団の解析により、異なる塩基配列をもつ多くのHLAタイプが次々と同定され、これまでに同定されたHLAタイプは、配列情報と共にIMGT/HLAデータベース (<http://www.ebi.ac.uk/ipd/imgt/hla/>) に登録されている（図2）。

HLAクラスI	A	B	C
	3657	4459	3290

HLAクラスII	DRB1	DRB3	DRB4	DRB5
	1977	127	53	43
	DQB1	DQA1	DPB1	DPA1
	978	76	716	44

図2：IMGT/HLAデータベースに登録されているアレル数

IMGT/HLAデータベース（ver-3.26）に登録されている遺伝子座ごとのアレル数を示す。

当データベースは3ヶ月ごとに更新され、登録数は年々増加している⁶⁾。多くの多型は抗原結合領域に集中する。つまり、HLAクラスIでは、 $\alpha 1$ と $\alpha 2$ ドメインをコードするエクソン2とエクソン3内に、HLAクラスIIでは、 $\alpha 1$ と $\beta 1$ ドメインをそれぞれコードするエクソン2内に多型が集中する^{8), 11)} (図3)。

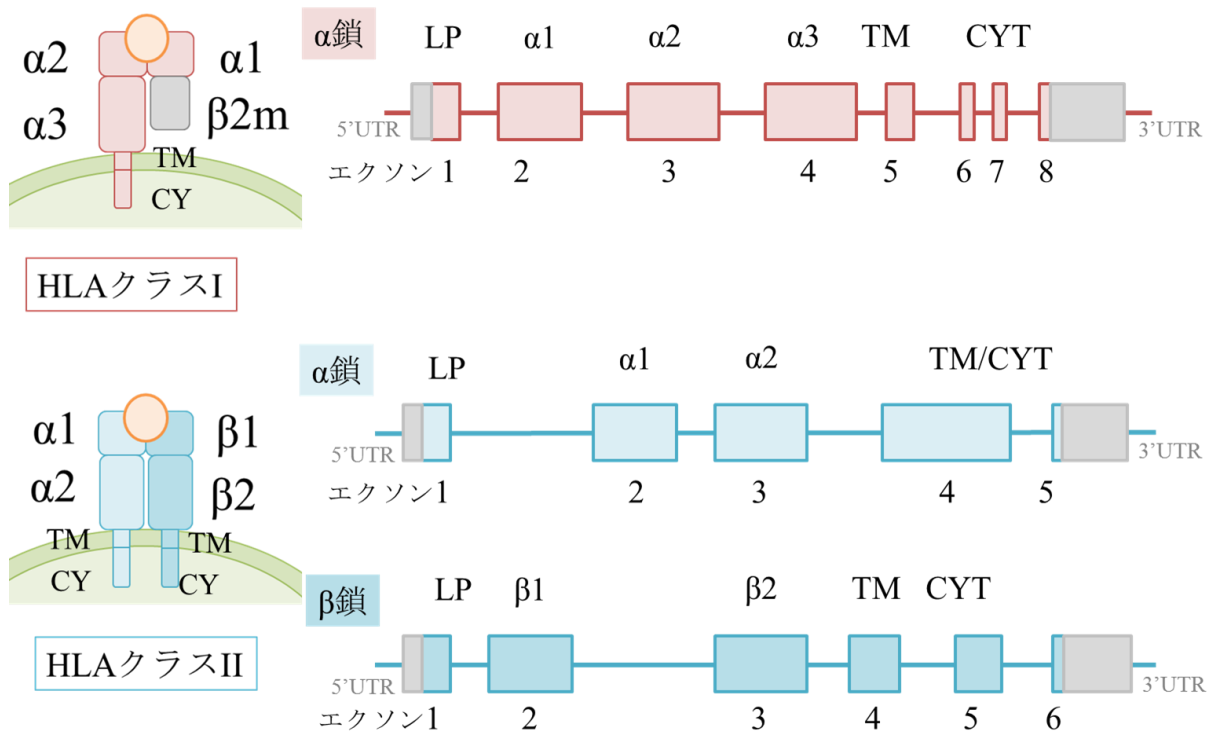


図3：各ドメインをコードするエクソン

HLAクラスIの α 鎖、HLAクラスIIの α 鎖、 β 鎖の各ドメインをコードするエクソンを示した。
TM：膜貫通ドメイン、CY：細胞質ドメイン、LP：リーダーペプチド

1-4. HLA タイプの命名法

HLA タイプの命名法は世界保健機関の HLA 命名委員会において規定されており、HLA 遺伝子座名の後にアスタリスクを挟み、コロンで区切られた 4 領域における数字と最後に付記されるアルファベットを用いて HLA タイプは表される (図 4)。

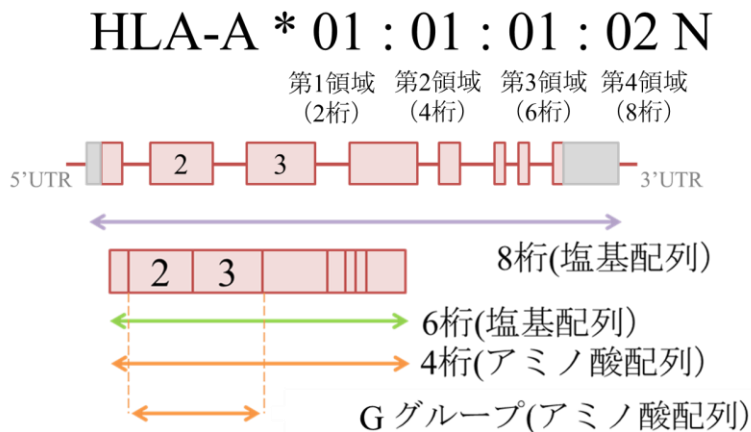


図 4 : HLA タイプの命名法と HLA タイピングの解像度

世界保健機関の HLA 命名委員会において規定されている HLA タイプの命名法についてその例を示した。また、HLA クラス I のイントロン、エクソンの図を用いて、各桁において表される範囲と意味を示した。

第 1 領域まででは血清型を表す。血清型は、レシピエントの血清中にドナーのリンパ球と反応する抗体の種類を示す。第 2 領域まででは、コーディング領域においてアミノ酸置換を伴う塩基置換によりタイプを区別する。つまり、4 桁タイプはコーディング領域のアミノ酸配列を表している。第 3 領域まででは、アミノ酸置換を伴わない塩基置換によりタイプを区別する。つまり、6 桁タイプはコーディング領域の塩基配列を表している。第 4 領域まででは、コーディング領域外における塩基置換によりタイプを区別する。つまり、8 桁タイプはイントロン領域と非翻訳領域も含めた領域の塩基配列を表している。

最後に付記されるアルファベット「N」は Null アレル (HLA が発現されていないアレル) を意味する。コーディング領域の途中で終止コドンとなる多型が存在するタイプである。アルファベット「G」は抗原結合領域の塩基配列が同じであるタイプグループを表す。これまでの HLA タイピングでは抗原結合領域をコードする範囲のみの解析が多く行われてきた。しかし、HLA データベースへのタイプ登録数の増加に伴い、抗原結合領域外でも多く多型が見つかってきており、4 桁以上の HLA タイピングが困難になってきていた。そこで、抗原結合領域のアミノ酸配列が判定できるように、「G」で表されるグループ (「G グループ」とする) が登録された^{11)~13)}。

1-5. HLA タイピングについて

1-5-1. 移植分野における HLA タイピング

HLA タイピングは父親、母親由来のアレルそれぞれの HLA タイプを決定することを意味し、主に臓器移植、造血幹細胞移植^{14)~17)}などの移植分野において行われる。また、感染症¹⁸⁾、自己免疫疾患¹⁹⁾、薬剤感受性²⁰⁾などの関連解析においても有用な検査技術となっている。移植時の移植片と宿主の HLA タイプが不一致であると強い T 細胞免疫応答が生じ、移植成績が左右され、移植片対宿主病などの合併症が生じる事がある⁵⁾。

抗原結合領域のアミノ酸配列レベル、つまり G グループレベルの HLA タイプの一致が移植分野において重要であることが報告されている²¹⁾。さらに、G グループレベル HLA タイピングより 4 桁レベル HLA タイピングで HLA タイプを調べた方が、移植成績が良い事も報告されている^{22)~24)}。また、造血幹細胞移植において HLA が発現しているかどうかも移植成績に影響することが報告されている²⁵⁾。そのため、Null アレルの検出が重要となる。さらに、HLA の発現に影響する多型がプロモーター領域において存在することが報告されており、8 桁レベル HLA タイピングも重要とされてきている^{26)~29)}。

1-5-2. 従来の HLA タイピング

現行の HLA タイピング手法には、SSP (Sequence Specific Primer) 法^{30), 31)}、SSOP (Sequence Specific Oligonucleotide Probe) 法^{32), 33)}、SBT (Sequence Based Typing) 法^{34), 35)}がある。SSP 法 (図 5A) は、各 HLA タイプ特異的なプライマーにより PCR を行い、ゲル電気泳動により DNA バンドのパターンを検出することで HLA タイプを判定する。SSOP 法 (図 5B) は目的とする HLA 配列の PCR 産物に対し、各 HLA タイプ特異的なプローブを反応させ、そのプローブを検出することより HLA タイプを判定する。SBT 法 (図 5C) はサンガー法により直接配列を決定することによりタイプを判定する方法である^{36), 37)}。

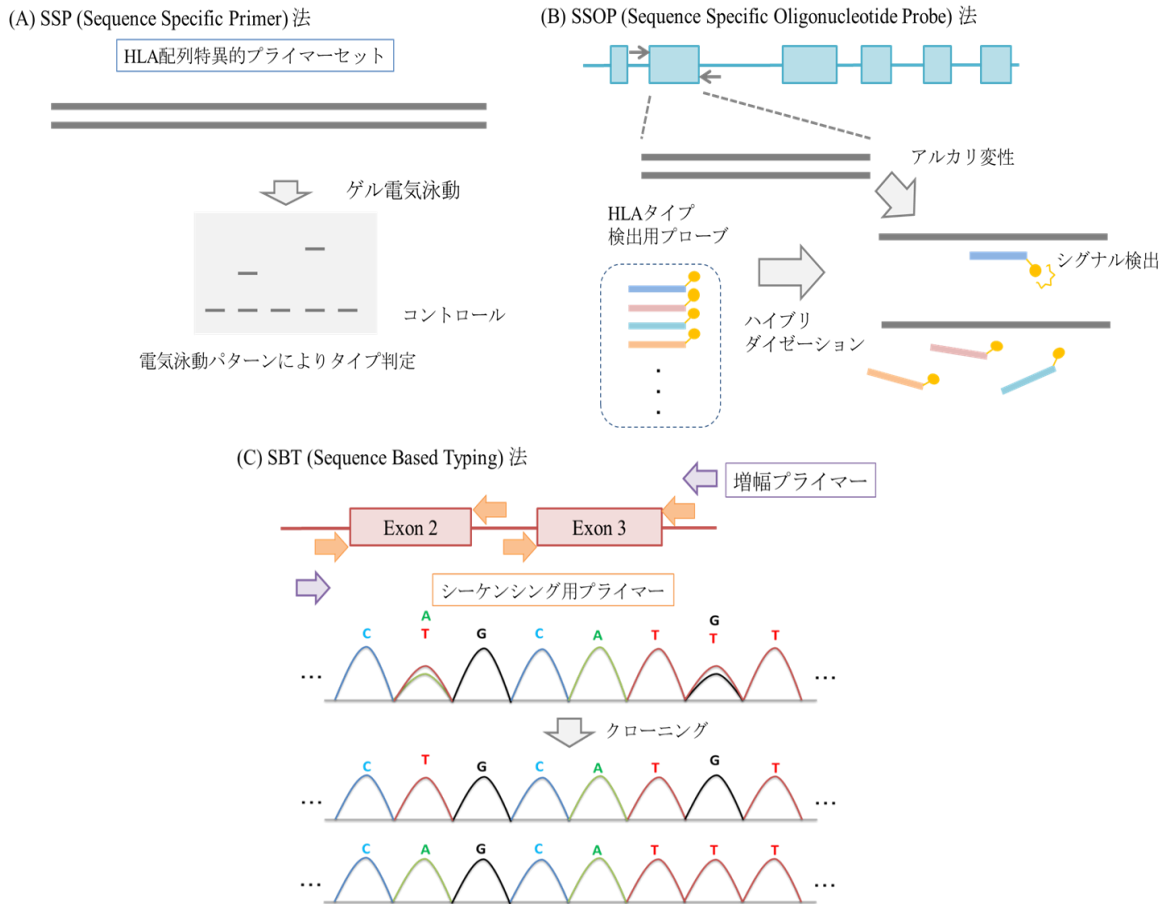


図 5 : 従来の HLA タイピングの概要

(A) SSP (Sequence Specific Primer) 法の概要を図示した。

コントロール：それぞれの PCR 産物に対して、混合して泳動する DNA。

(B) SSOP (Sequence Specific Oligonucleotide Probe) 法の概要を図示した。HLA クラス II のエクソン 2 の PCR 産物に対して検出用プローブのハイブリダイゼーションにより検出する例を示す。

(C) SBT (Sequence Based Typing) 法の概要を示した。増幅プライマーにより解析領域を増幅し、シーケンシング用プライマーで各エクソンの塩基配列を解析する例を示している。それぞれの塩基で表された波形はサンガーシーケンシングによって得られた一部領域のデータを示す。

1-5-3. 従来の HLA タイピングの問題点

SSP 法、SSOP 法では、年々同定される新規 HLA タイプに対応するためのプライマー、プローブが不足している。SSP 法、SSOP 法による HLA タイピング精度を維持するためには、新規 HLA タイプに対応するためのプライマーやプローブを新たに更新する必要があるが、多大なコストがかかってしまうため SSP 法、SSOP 法による HLA タイピングの解像度は、2 桁レベルもしくは G グループレベルまでにとどまっている。

一方、SBT 法は高解像度 HLA タイピング (4 桁以上の HLA タイピング) が可能である。しかし、HLA 遺伝子領域の高度な多型性が影響するため、2 つの染色体から増幅された混合産物のシーケンシングでは、各多型部位が同一の染色体上 (cis)、もし

くは異なる染色体上 (trans) に位置するののかというアレルの区別化ができない (phase ambiguity)。

高解像度 HLA タイピングを目的とした SBT 法は、phase ambiguity を解消するために、クローニングによりアレルを分けたハプロイド配列のシーケンシングを行う必要がある^{37), 38)} (図 5C)。その場合、検査工程が煩雑になり検査に要する時間と労力の消費が大きくなり、また、多数検体を処理することも困難となることから、多くの SBT 法による HLA タイピングでは G グループレベル HLA タイピングにとどまっている^{35), 39)}。

1-5-4. 次世代シーケンサーによる HLA タイピング

近年、上記の問題を解決できる次世代シーケンサー (NGS : Next Generation Sequencer) が HLA タイピング分野に導入されつつある。NGS では、1 分子の DNA 配列由来の配列情報を得ることができ、従来の方法 (項目 1-5-3 を参照) で問題になっていた phase ambiguity を解消できる。そのため、アレルごとの高解像度 HLA タイピングが可能となる^{37), 40), 41)}。また、サンガー法に比べて、NGS による配列解析は、解析される配列数が膨大であることから、複数サンプルの異なる遺伝子座を一度に解析することができる。以上の理由により、NGS を用いた HLA タイピングは従来の方法よりも簡便で、かつ効率の良い手法になり得る^{37), 41)}。

1-5-4-1. 報告されている HLA タイプ判定プログラムについて

2009 年より NGS を用いた HLA タイピング手法が報告されはじめた。当初は 454 GS FLX (Roche) による HLA タイピング手法の報告が主であったが、454 GS FLX よりも安価で多くのデータ量を出力できる MiSeq (Illumina)、Ion Torrent PGM (Thermo Fisher Scientific) のような次世代シーケンサーが登場、普及してからはこれらの機種での HLA タイピング手法の報告が行われてきた。それぞれの機種で異なる長所や短所があり²⁴⁾ (表 1)、それぞれの長所を活かした HLA タイプ判定プログラムが考えられている。

NGS機種名	会社名	データ量	配列決定できる鎖長	平均解析時間	本体価格	長所	短所
454 GS FLX	Roche	700 Mb	~1000 塩基	23時間	\$500,000	・長い鎖長	・読み取りエラーが多い ・データ量が少ない ・解析費用が高い ・解析時間が長い
MiSeq	Illumina	15 Gb	~300 塩基	56時間	\$125,000	・読み取りエラーが少ない ・データ量が多い ・解析費用が低い	・解析時間が長い
Ion Torrent PGM	Thermo Fisher Scientific	2 Gb	~400 塩基	3時間	\$80,000	・解析時間が短い ・データ量が多い ・解析費用が低い	・読み取りエラーが多い

Carapitoらのレビュー論文²⁴⁾を基に作成

表 1 : 次世代シーケンサー (NGS) 機種の特徴

以下に、各 NGS 機種による HLA タイプ判定プログラムの報告をいくつか紹介する。

- 454 GS FLX

Gabriel ら⁴²⁾、Holcomb ら⁴³⁾ が報告している 454 GS FLX を用いた手法では、配列決定できる鎖長が長いという特徴を活かした解析プログラムになっている。読み取りエラーが少なく、鎖長が長く、データ品質の高いデータを選択するという処理を行った後に、IMGT/HLA データベース (<http://www.ebi.ac.uk/ipd/imgt/hla/>) を調べ、完全一致したタイプを選定するといった比較的シンプルな HLA タイプ判定プログラムになっている。

一方、MiSeq、Ion Torrent PGM は、454 GS FLX よりも本体価格が安価で、データ量が多く、解析コストも低いため、これらの機種を用いた HLA タイピング手法の開発が重要である。しかし、これらの機種には配列決定できる鎖長が短い (MiSeq : ~300 bp、Ion Torrent PGM : 400 bp) という短所があり、短い配列データを用いた HLA タイプ判定プログラムが考えられてきている。

- Illumina システム (MiSeq、HiSeq)

Hosomichi ら⁴⁴⁾、Cao ら⁴⁵⁾ の方法は Illumina システムの出力データからアレルごとにつなぎ合わせて解析領域全体のコンセンサス配列を作成し、IMGT/HLA データベースとの類似性検索を行う。

- Ion Torrent PGM

Shiina ら^{46), 47)} は、SeaBass (Sequence Alignment Based Assignment Software) という HLA タイプ判定プログラムを開発している。SeaBass では、アライメントツールを用い、配列データと IMGT/HLA データベースとの類似性検索を行った後、候補となるアレルを絞り込む。そして、その候補アレルに配列データをマッピングさせ、算出された被覆率を参考にし、HLA タイプの判定が行われる。

1-5-4-2. 報告されている HLA タイプ判定プログラムの精度について

解析対象としている遺伝子座数 (6~11 遺伝子座) やサンプル数 (20~336 サンプル) が多く、NGS を用いて高解像度 HLA タイピング (4 桁以上の HLA タイピング) を行っている報告^{43), 45), 50) ~ 52)} の HLA タイピング正解率は 93~97 %代にとどまっている。99%代の HLA タイピング正解率を示す報告^{42), 46), 49), 53)} もあるが、解析対象としている遺伝子座数 (2~8 遺伝子座) やサンプル数 (4~11 サンプル) は少ない。

以上のように主要な HLA 遺伝子座すべてを解析対象とし、高い HLA タイピング正

解率を示しているような報告はまだない。各報告において 100 % の正解率とならなかった理由として以下のような問題点が挙げられている。

- ・ 不正確なアレルの区別化

ゲノム DNA から個々のエクソンを増幅する方法を用いライブラリー作成を行った場合、エクソン間の cis、trans 情報を得ることができないため、決定できないハプロタイプが存在する⁵⁾。

- ・ NGS による読み取りエラー

NGS を用いた配列解析では読み取りエラーが多いことが知られている^{24), 55)}。各報告では NGS の多くの配列データよりコンセンサス配列を作成することで、このようなエラーを克服する手法をとっているが、少なからず HLA タイピングに影響するケースがある^{5), 40)}。

- ・ PCR 増幅バイアス・片側対立遺伝子の欠落

一般に PCR ではアレル間において増幅バイアス⁵⁶⁾ やアレルの一方のみが増幅されない片側対立遺伝子の欠落 (PCR dropout) が知られている。NGS を用いた HLA タイピングにおいても、PCR の増幅バイアスや PCR dropout が原因で、片アレルの HLA タイプの決定が行えないことが報告されている^{5), 40), 52)}。

1-5-4-3. 癌治療における HLA タイピングを行う意義

HLA タイピングは腫瘍ペプチドワクチン療法を行う上でも重要である。腫瘍ペプチドワクチン療法⁵⁷⁾ とは、腫瘍抗原であるペプチドを投与し、そのペプチド特異的な細胞傷害性 T 細胞を活性化させ、腫瘍の増大を抑制させる治療方法である。腫瘍ペプチドワクチン療法では、ある特定の HLA タイプを有する患者に効果を示すように開発されるため、治療前に HLA タイピングを行っておく必要がある。

癌患者において、HLA の発現が低下または無い場合、抗原提示細胞の腫瘍抗原提示能が低下する。その結果、細胞傷害性 T 細胞の活性が低下し、抗腫瘍作用が低下する事が報告されている⁵⁸⁾。HLA の発現量に関わるアレル (Null アレルなど) を同定する HLA タイピングは腫瘍の免疫逃避機構に関する研究分野においても重要になっている。

さらに、HLA の発現低下は、免疫チェックポイント阻害剤による腫瘍の治療効果にも影響し得る。免疫チェックポイント阻害剤による腫瘍免疫療法は、細胞傷害性 T 細胞の活性化抑制に関与する免疫チェックポイント分子 (Programmed death-1 : PD-1,

Programmed death ligand-1 : PD-L1, Cytotoxic T-lymphocyte-associated protein 4 : CTLA-4) を阻害することにより、細胞傷害性 T 細胞を活性化させ、腫瘍を攻撃する治療法である^{59), 60)}。HLA の発現低下により抗原提示細胞の腫瘍抗原提示能が低下すると、細胞傷害性 T 細胞は活性化せず、免疫チェックポイント阻害剤は意味をなさず、治療効果を示さない事が報告されている^{61), 62)}。免疫チェックポイント阻害剤は、一部の患者では劇的な治療効果を示すが、治療効果のない患者も存在する。そのため、免疫チェックポイント阻害剤の効果がある患者を選定するためのバイオマーカー開発 (HLA 領域の Null アレルの同定など) が重要となっている。

当研究では、1 回の Ion Torrent PGM による配列解析で、最大 16 人の RNA サンプルから 11 遺伝子座 (HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1) それぞれのハプロタイプを自動的に決定できる HLA タイピングシステムを開発した。HLA タイピングシステム中の解析プログラムは、① 配列ファイル作成プログラム、② ハプロタイプ候補配列作成プログラム、③ HLA データベース検索プログラム、④ HLA タイプ決定プログラムから構成されている。Ion Torrent PGM の短所である読み取りエラーを除去した上で構成されたハプロタイプの候補となる配列を使用し、IMGT/HLA データベースを調べ上げ、HLA タイピングを行う。当プログラムは、Ion Torrent PGM のような鎖長が短い配列データからでも 6 桁レベル (コーディング領域の塩基配列レベル) での HLA タイピングが可能なプログラムとなっている。

2. 結果

2-1. HLA タイピングシステムの開発

当 HLA タイピングシステムは、「HLA タイピングプログラム用ライブラリー調製」、「NGS による配列解析」、「HLA タイピングプログラム」で構成されている。当項目では、HLA タイピングシステムの開発過程をその根拠と共に説明する。なお、NGS の出力配列が短いため、通常データ処理では HLA タイピング不能な検体があった。これらの検体に対応するため分子バーコード配列を利用して同一分子由来の配列を統合する「補完 HLA タイピングシステム」を別途開発した。まず、主体となる HLA タイピングシステムについて説明する。

2-1-1. HLA タイピングプログラム用ライブラリー調製

2-1-1-1. 一本鎖 cDNA 合成

全 RNA からランダムヘキサマーを用いて一本鎖 cDNA を合成した。RNA を用いた HLA タイピングではゲノム DNA を用いた HLA タイピングよりも解析領域が少なく、エクソン間での cis、trans の多型情報を取得しやすい。HLA 遺伝子座ごとのプライマーセット（項目 2-1-1-2 を参照）を用いて、一本鎖 cDNA を鋳型とする PCR を行うのだが、HLA 遺伝子領域全体を解析対象とするため、一本鎖 cDNA 合成では HLA 特異的なプライマーでなくランダムヘキサマーを使用した。詳細は材料と方法の項目 4-2-1 において説明する。

2-1-1-2. HLA プライマーの設計

解析対象とした各 HLA 遺伝子座に対して特異的なプライマーの設計を行った。HLA 解析用のプライマーセットを各 HLA 遺伝子座のコーディング領域をカバーするように設計した。IMGT/HLA データベース (<http://www.ebi.ac.uk/ipd/imgt/hla/>) にはコーディング領域の塩基配列タイプを表すデータベース（「HLA cDNA データベース」とする）と、イントロン領域も含めたゲノム DNA の塩基配列を表すデータベース（「HLA DNA データベース」とする）が存在する。それらの配列データを用いて各 HLA タイプにおいて保存されている領域にプライマーを設計した。プライマー設計時の HLA データベースのバージョンは version 3.16.0（2014 年 4 月更新）である。各 HLA 遺伝子座のコーディング領域をカバーするために HLA クラス I (HLA-A, -B, -C) に対してそれぞれ 4 領域、HLA クラス II (HLA-DRB1/3/4/5, -DQB1, -DQA1, -DPB1) に対してそれぞれ 3 領域を増幅するようなプライマーを 27 セット設計した（表 2）。

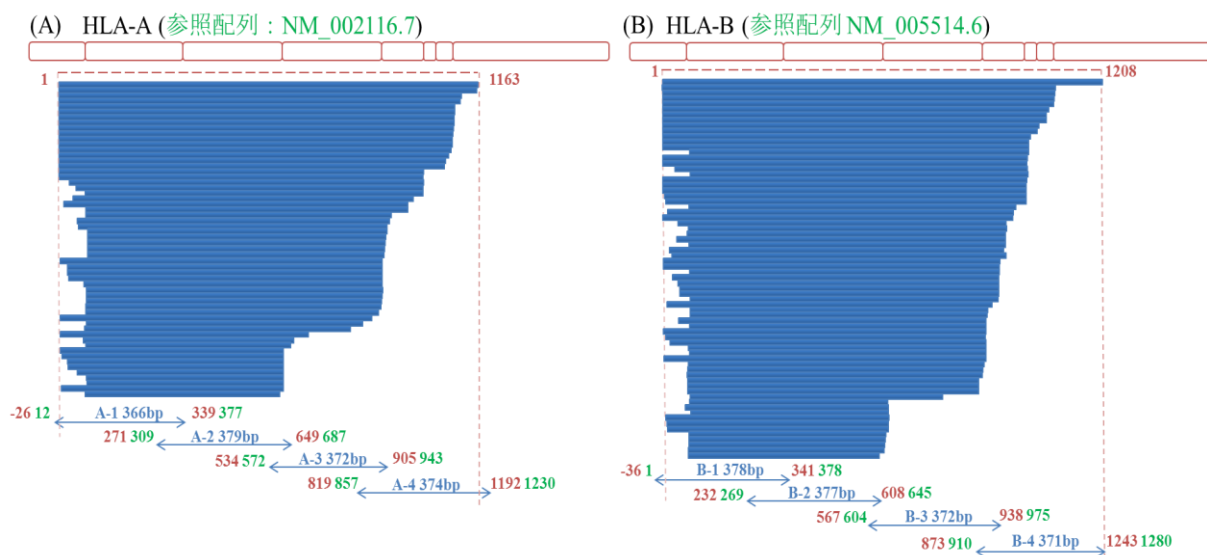
プライマー名	プライマー配列 (5'-3')	保存率1 (%)	不一致タイプ数	各プライマーが重なるHLAタイプ数	保存率2 (%)	不一致タイプ数	各プライマーが重なるHLAタイプ数 (アレル頻度: ≥ 0.001)
A-1F	ACTCAGATTCTCCCAGACG	99.3	1	145	100	0	67
A-1R	CCTCGCTCTGGTTGTAGTAGC	98.5	41	2735	100	0	225
A-2F	TGAAGGCCCASTCACAGAY	97.8	60	2735	100	0	225
A-2R	GBGTGGTGRGTCATATGYGT	98.5	15	990	100	0	145
A-3F	CAGTTGAGAGCCYACCTGSA	100	0	2735	100	0	225
A-3R	GGAAGA YRGCTCCCATCTCA	100	0	434	100	0	106
A-4F	CCTTCTGGARAGGAGCAGAG	100	0	434	100	0	145
A-4R	AGGTGCCTTTGCAGAAACAA	99.3	1	145	100	0	67
B-1F	CACCCACCCGACTCARA	100	0	234	100	0	115
B-1R	GCCTCGCTCTGGTTGTAGT	98.6	47	3453	99.2	3	360
B-2F	ARGAGGGGCC K GARTATT	98.7	46	3453	99.2	3	360
B-2R	CAGK T S T C C T T C C C G T T C T	99.7	12	3453	100	0	359
B-3F	GAGTSGCTCCGAGAYACC	98.3	60	3453	99.4	2	360
B-3R	AGCAACAATGCCACGAT	99.8	1	573	100	0	170
B-4F	AAGCCCCACACCTGAGAT	99.0	6	573	99.6	1	231
B-4R	GCTGTCTCTCCACCTCCTCA	100	0	234	100	0	115
C-1F	GGACTRSATTCTCCCAGA	100	0	192	100	0	67
C-1R	CTCGCTCTGGTTGTAGTAGCC	98.2	41	2257	99.3	1	148
C-2F	CGCGGCTACTACAACCAGA	98.3	39	2257	100	0	148
C-2R	ACCTCAGGTGGCCTCAT	98.8	8	658	100	0	83
C-3F	CAGYDSAGAGCCTACCTGGA	100	1	2257	100	0	72
C-3R	GATGGTRGGCTGGGAAGA	100	0	308	100	0	72
C-4F	TCTTCCAGCCYACCATC	100	0	308	100	0	72
C-4R	ACAGGTCAGTGTGGGACA	100	0	192	100	0	67
DRB1/3/4/5-1F	CCAGCATGGTGTGTCTGARG	100	0	53	100	0	51
DRB1/3/4/5-1R	RACTCGCKCTGCACAYGT	100	0	220	100	0	129
DRB1/3/4/5-2F	MCTGGAACAGCCAGAAGGAC	98.5	24	1561	98.9	4	362
DRB1/3/4/5-2R	ATCACCGRR TCTGGAAGGT	100	0	208	100	0	125
DRB1/3/4/5-3F	GAAGAGAAGRCKGGGR TGGT	100	0	208	100	0	125
DRB1/3/4/5-3R(v1)	GAAAGCTTTTCATC HTGCAA	84.9	8	53	84.3	8	51
DRB1/3/4/5-3R(v2)	TTGTGGAAGAATRASWGCCARG	100	0	53	100	0	51
DQB1-1F	TGCCAGSTACATCAGATCCA	100	0	21	100	0	19
DQB1-1R	TCCTTCTGNCTVTTCAGTA	98.5	9	591	98.5	1	68
DQB1-2F	CGCTT C B A C A G C G A C G T G G	96.3	22	594	100	0	70
DQB1-2R	CACCAGRR TCTGGAARGTCC	100	0	233	100	0	31
DQB1-3F	GAATGRY C A G G A G G A G A C A	98.7	3	233	100	0	31
DQB1-3R(v1)	GCATAAGCAGGCATCACAGA	4.8	20	21	31.6	13	19
DQB1-3R(v2)	GCAGGCRTYACAGAA R A G T G	100	0	21	100	0	19
DQA1-1F	CAGCTCAGAACACCACTGC	100	0	28	100	0	22
DQA1-1R	GACCTCAGGAACCTCATTGG	100	0	46	100	0	29
DQA1-2F	CCCATGAATTTGATGGAGAT	100	0	52	100	0	32
DQA1-2R	GGAATGATCACTCTGGAGAGG	100	0	46	100	0	29
DQA1-3F	AGTCACAGAAGGTGTTTCTGAGA	100	0	46	100	0	29
DQA1-3R(v1)	GATGCACCTTCCCTCCAG	50.0	14	28	36.4	14	22
DQA1-3R(v2)	AGGATGGGATTCACAA K G G	100	0	28	100	0	22
DPB1-1F	TCATTTGCCATCCITTTCC	100	0	7	100	0	7
DPB1-1R	CCAGCTCGTAGTTGTGCTG	100	0	344	100	0	74
DPB1-2F	GCGCTTCTGGAGAGATACA	96.2	13	344	100	0	74
DPB1-2R	CCATTACGGATCAGGTWGGT	100	0	59	100	0	35
DPB1-3F	CCTGAATGGACAGGAGGAAA	100	0	59	100	0	35
DPB1-3R	GGARCCAGRTGCTAACRAAA	100	0	7	100	0	7
DPA1-1F	TCATCAATTATAGACCCACAAC	100	0	5	100	0	5
DPA1-1R	GTCACCTCAGGGGATCRT	100	0	38	100	0	38
DPA1-2F	AAGGAGACCGTCTGKCATCT	100	0	38	100	0	38
DPA1-2R	CCCTGCAGTCATAGAAGTCCTC	100	0	38	100	0	38
DPA1-3F	TCTTCTGCCAGAACAGAT	100	0	11	100	0	11
DPA1-3R	GGARCCAGRTGCTAACRAAA	100	0	5	100	0	5

 : DNAデータベース総登録タイプ数
 : cDNAデータベース総登録タイプ数
 : cDNAデータベースに登録されているタイプのうち各種プライマーが重なるタイプの総数 (cDNAデータベースには図6で示したように様々な長さの配列が登録されているため、プライマーの設計位置によっては重ならないタイプも存在する)
 : DNAまたはcDNAデータベース中で各種プライマーが重なるタイプのうち、アレル頻度が0.001以上のタイプ数
各タイプのアレル頻度はAllele Frequencies (<http://www.allelefrequencys.net/>)を参考にした
各混合塩基の表記について
S : C,G, Y : C,T, R : A,G, K : G,T, M : A,C, W : A,T, B : G,T,C, D : G,A,T, B : G,T,C, V : A,C,G, N : A,C,G,T

表2: プライマー配列と保存率

各遺伝子座の両端に位置させたプライマー（HLA クラス I：1F と 4R、HLA クラス II：1F と 3R）は、コーディング領域外（HLA cDNA データベースに登録されているタイプはコーディング領域内の配列データである）に設計した為、HLA DNA データベースの配列を用いて保存率 1、保存率 2（下記）を計算した。なお、各遺伝子座の両端に位置するプライマー以外は HLA cDNA データベースの配列を用いて保存率を計算した。保存率 1 はプライマー設計領域の配列が保存されているタイプの割合を表し、保存率 2 はアレル頻度（Allele Frequencies に登録されている人種）が 0.001 以上のタイプにおいてプライマー設計領域の配列が保存されているタイプの割合を表す。アレル頻度は、各人種においてアレル頻度が調べられているデータベース Allele Frequencies (<http://www.allelefreqencies.net/>) を用いて調べた。改良前プライマー（以下で説明）を除いたプライマーの保存率 1 は、96.2 ～100 %、保存率 2 は、98.5 ～100 % となった。したがって、設計したプライマーは、アレル頻度の少ないタイプを除くと、ほとんどの HLA cDNA データベースに登録されているタイプを特異的に増幅することができる。

なお、設計した 3 プライマーでバージョン 1 (v1) とバージョン 2 (v2) の 2 種類が存在しているが、v2 は v1 の改良版である。v2 は v1 の設計段階において、一部のタイプにおいて保存領域の見落としがあることが判明し HLA タイピングが行えなかったため、設計の見直しを行ったものである。したがって、一部のサンプルにおいて、改良前プライマーではなく改良後プライマーから出力されたデータを使用している。詳細な設計方法は項目 4-2-2 において説明する。解析対象とした 11 遺伝子座における各増幅領域と HLA cDNA データベースとの位置関係は図 6 に示した。



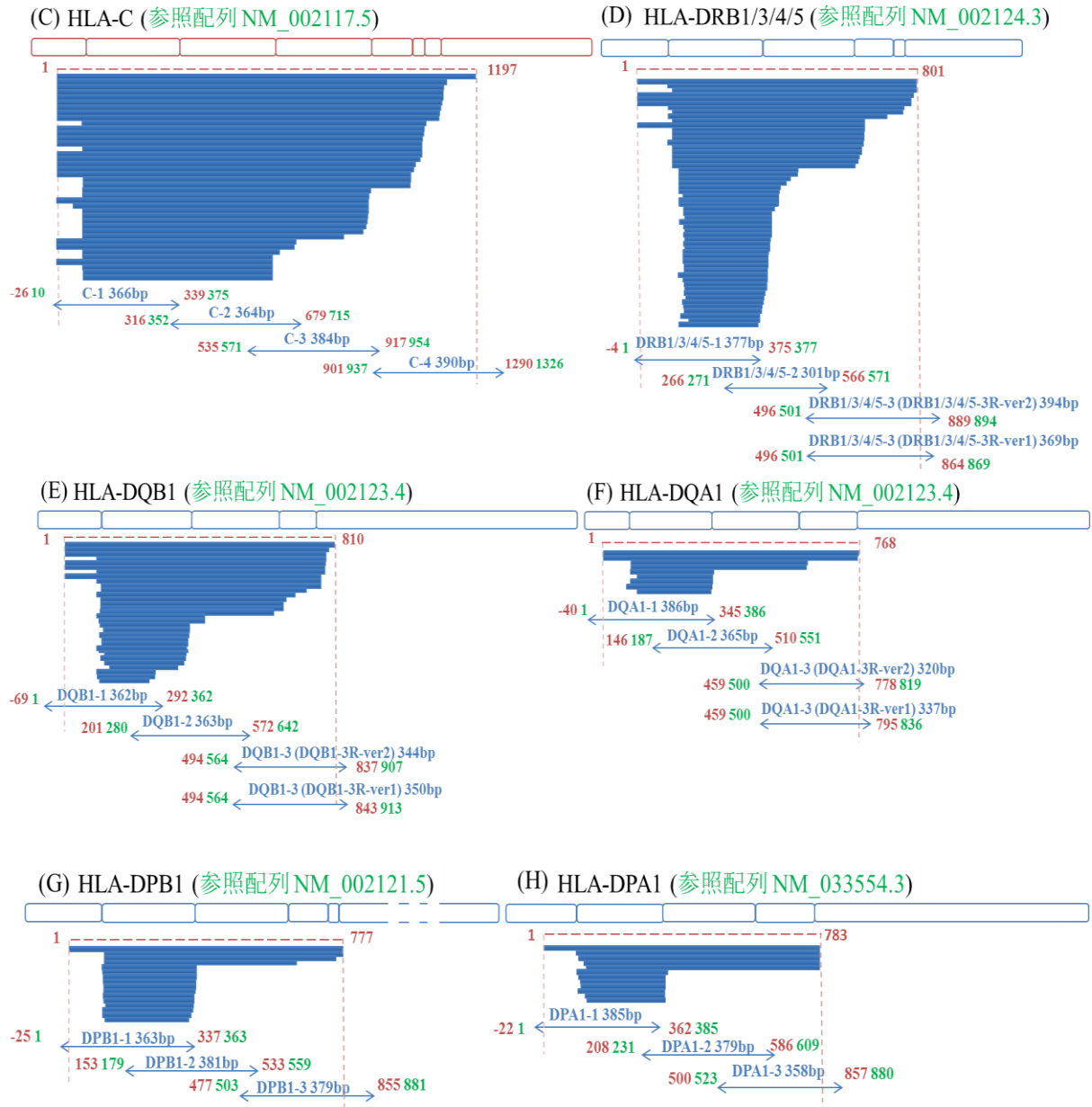


図 6 : PCR による増幅領域と HLA cDNA データベース/参照配列との位置関係
 (A) HLA-A, (B) HLA-B, (C) HLA-C, (D) HLA-DRB1/3/4/5, (E) HLA-DQB1, (F) HLA-DQA1,
 (G) HLA-DPB1, (H) HLA-DPA1

青色バー： HLA cDNA データベースに登録されている配列を長さの種類ごとに並べたもの。

青矢印：設計したプライマー（表 2）による増幅領域を表す。

数字（赤）：開始コドン（ATG）の A を 1 とした場合の各増幅領域の位置を表す。

数字（緑）：マッピングに使用する参照配列の最初の塩基を 1 とした場合の位置を表す。

2-1-1-3. NGS 解析用アダプター配列の付加

調製した一本鎖 cDNA (項目 2-1-1-1 を参照) を鋳型とし、設計したプライマーセット (項目 2-1-1-2 を参照) により PCR を行なった後、27 種類の PCR 産物をサンプルごとで等量ずつ混合し、精製した。そして、混合 PCR 産物にアダプター配列を付加した。このアダプター配列には、個人を識別するための配列と、NGS による配列解析を可能とする配列が含まれている。個人を識別するための配列が含まれているアダプターを付加することで、複数サンプルを一回の配列解析で解析することができる。詳細な方法は項目 4-2-3, 4 において説明する。

2-1-2. NGS による配列解析

NGS により解析されたデータは、Torrent Server (Ion PGM™ Sequencer による解析用サーバー) 上でデータ品質のフィルタリングが行われ、個人識別配列により各サンプルが識別され、各遺伝子座に対して 1 種類ずつ用意した参照配列へのマッピングが行われる。マッピングは解析された配列がどの遺伝子座のどの領域由来の配列なのかを明らかにするために行う。参照配列データベースである NCBI Reference Sequence (<http://www.ncbi.nlm.nih.gov/refseq/>) に登録されている各 HLA 遺伝子座の cDNA の配列から 1 種類選択し、各遺伝子座の参照配列とした (図 6)。詳細な方法は項目 4-2-5 において説明する。

2-1-3. HLA タイピングプログラムの開発

NGS によるテストセット (12 サンプル: 健常人血漿由来 RNA サンプル 3 例、乳癌組織由来 RNA サンプル 9 例) の配列解析を行なった出力データを用いて HLA タイピングプログラムの開発を行なった。

HLA タイピングプログラムは「配列ファイル作成プログラム」、「ハプロタイプ候補配列作成プログラム」、「HLA データベース検索プログラム」、「HLA タイプ決定プログラム」から構成されている。上記プログラムを順に実行させることにより自動的に、各サンプルの遺伝子座ごとの HLA タイプを決定する。図 7 に HLA タイピングプログラムの全体像を示す。当項目では各プログラムの説明を行うと共に、そのような解析方法にした根拠をテストセットのデータを基に行う。

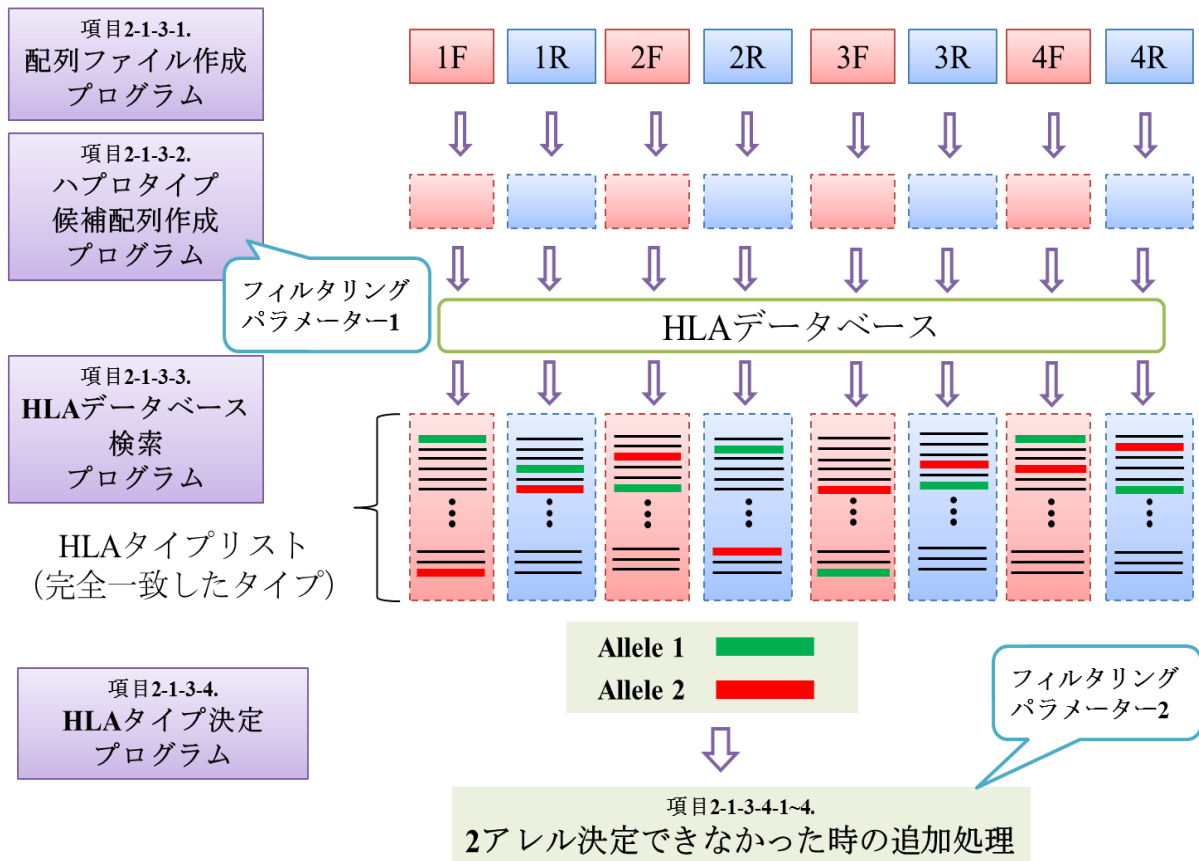


図7：HLA タイピングプログラムの全体像

HLA タイピングプログラムは4つのプログラム（配列ファイル作成プログラム、ハプロタイプ候補配列作成プログラム、HLA データベース検索プログラム、HLA タイプ決定プログラム）から構成されている。

1F～4R：図8で示す各解析領域のシーケンスデータを表している。

緑線、赤線：各解析領域でハプロタイプ候補配列が完全一致した共通のHLAタイプを表す。

フィルタリングパラメーター1及び2：当プログラムで設定するパラメーターである。

2-1-3-1. 配列ファイル作成プログラム

「配列ファイル作成プログラム」は、各解析領域（HLA クラス I：1F~4R、HLA クラス II：1F~3R）の配列データを取得するプログラムである。NGS により生成された個々の配列データのことを「リード」と呼ぶことにする。図 8 は、増幅領域が 4 か所ある HLA クラス I の各リードが参照配列へマッピングした際のイメージ図である。配列解析の開始を示すアダプターが各ライブラリーに対して両方向に付加し、配列が読み取られるため、図 8 のように赤い矢印で示した右向きのリード（「フォワードリード」とする）と青い矢印で示した左向きのリード（「リバースリード」とする）が存在する。プライマー配列を除き、解析領域の長さの 55% もしくは 60%（遺伝子座ごとに設定）までの配列領域（図 8 の赤枠、青枠で示した領域）を解析対象配列とした。塩基読み取り精度によりデータとして出力されるリードの長さは様々であるが、出力データを確認したところ、約 350 塩基以上のリードでは後半の塩基読み取りエラーが多く見られる傾向にあり、以後の解析に悪影響が生じたため、途中までの配列を使用した。詳細な方法は項目 4-3-1 において説明する。

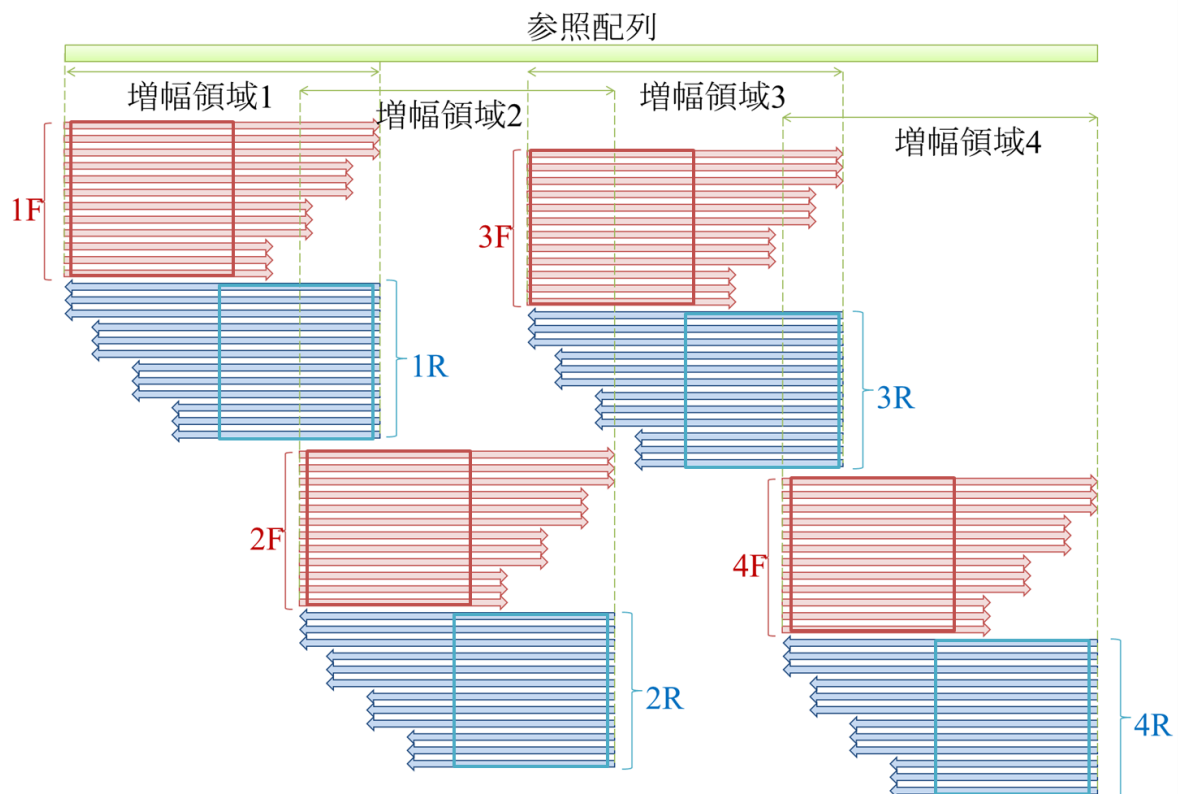


図 8：マッピングイメージ（HLA クラス I）

HLA クラス I において各増幅領域のリードが参照配列へマッピングされた時のイメージ図。増幅領域 1 のフォワードリードを 1F、リバースリードを 1R としている。増幅領域 2~4 も同様に 2F~4R とそれぞれのリードを表した。

赤枠、青枠：それぞれフォワードリード、リバースリードにおける解析領域を表す。

2-1-3-2. ハプロタイプ候補配列作成プログラム

「ハプロタイプ候補配列作成プログラム」では、マルチプルアライメントツールである Multiple Sequence Alignment : MUSCLE (<http://www.drive5.com/muscle>)⁶³⁾ を使用し、2つのアレル上の HLA 配列の候補（「ハプロタイプ候補配列」とする）の作成と、各候補配列の信頼性を示すスコア（「ハプロタイプスコア」とする）を付加する。ハプロタイプ候補配列作成過程において、読み取りエラーを除去し、かつ、アレルの区別化を行う。

当項目では、作成した配列ファイル（項目 2-1-3-1 を参照）を用い、NGS による読み取りエラーを除去しつつ、ハプロタイプ候補配列を作成する過程と、ハプロタイプスコアの算出方法について説明する。

まず、作成した配列ファイルをマルチプルアライメントツールにより整列化させ、ポジションごとに各塩基の割合を計算する（「縦」の情報）（図 9）。2つのアレルを分けた配列を作成するためには、「縦」の情報に加え、「横」の情報（各多型の組み合わせ）が必要である。「横」の情報を自動的に得るために、取得したポジションごとの塩基割合情報から 2 種類以上の塩基が存在するポジションを認識し、それらの塩基を用いて考えられる全てのハプロタイプを羅列する。図 9 の例の場合、アレル間で 3 箇所塩基が異なるので 8 パターン（ $2 \times 2 \times 2$ ）のハプロタイプ候補配列（候補 1~8）が羅列できる。そして、各ハプロタイプパターンを有する配列のリード数をマルチプルアライメントの出力データからカウントする。このリード数を「ハプロタイプスコア」とする。図 9 の場合、候補 1 と候補 2 のハプロタイプスコアはそれぞれ 6、候補 3~8 のハプロタイプスコアは 0 となる。図 9 では、NGS による読み取りエラーが全くない場合として図示しているが、実際のデータでは読み取りエラーが少なからず生じるため、候補 3~8 のようなパターンでも 1 以上のハプロタイプスコアとなる場合がある。さらに、2 アレルで塩基が異なるポジションでの読み取りエラーにより 3 種類以上の塩基が存在することも考えられるためハプロタイプの候補がさらに増える事もある。

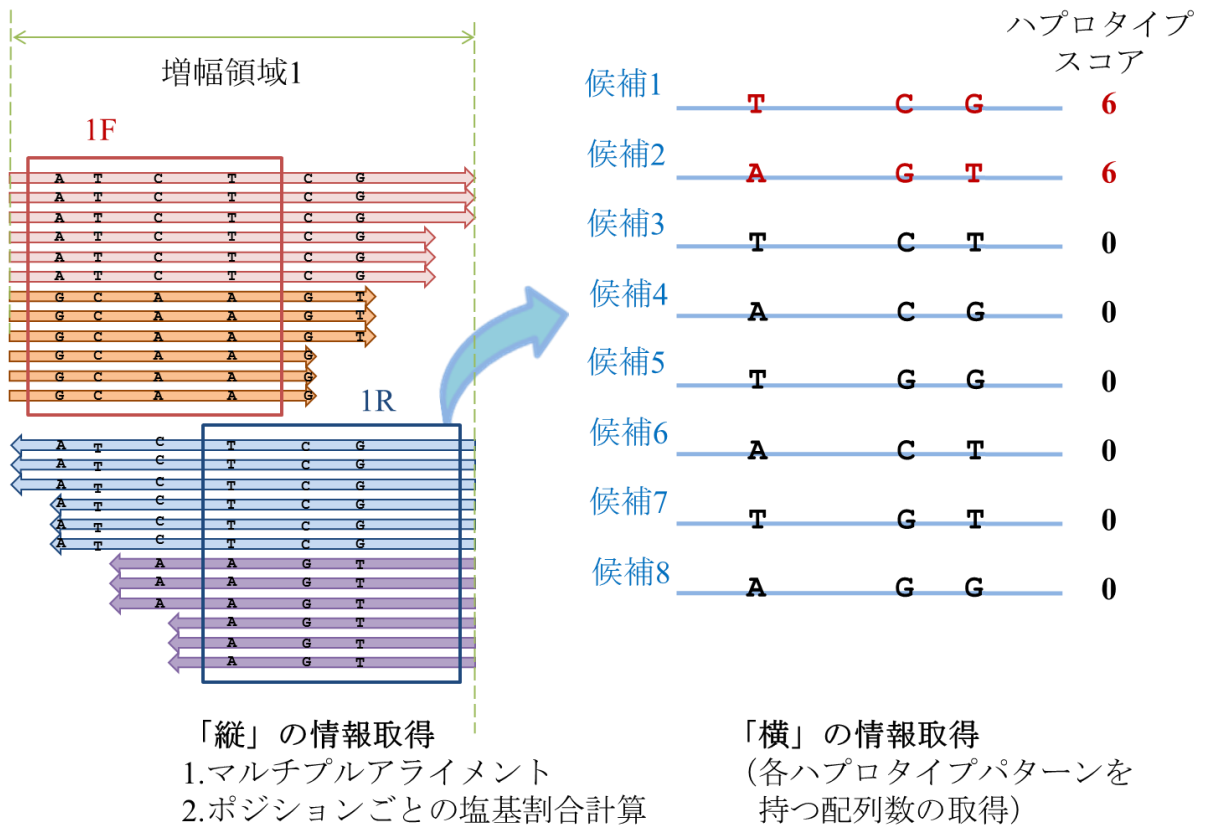


図9：ハプロタイプ候補配列作成プログラムの概要

増幅領域1に6箇所の多型が存在する配列例を示す。

赤枠：1F解析領域を表す。

青枠：1R解析領域を表す。

候補1~8: 1R解析領域の配列データから考えられるハプロタイプ候補配列とハプロタイプスコア。

NGS による読み取りエラーにより生じたハプロタイプ候補配列を除去するためにパラメーター（「フィルタリングパラメーター1」とする）を設定した。例として図 10 に BC63 サンプルの HLA-A で作成された 1F、1R 領域のハプロタイプ候補配列のハプロタイプスコアを降順で示した。フィルタリングパラメーター1 を「40」とした場合、最も高いハプロタイプスコアの 1/40 以下のスコアを有する候補配列を除去し、残りの候補配列をハプロタイプ候補配列として選択する。フィルタリングパラメーター1 は遺伝子座ごとに設定した。各遺伝子座の設定値に関しては項目 2-1-3-4-3 において説明する。ハプロタイプ候補配列作成の詳細な方法は項目 4-3-2 において説明する。

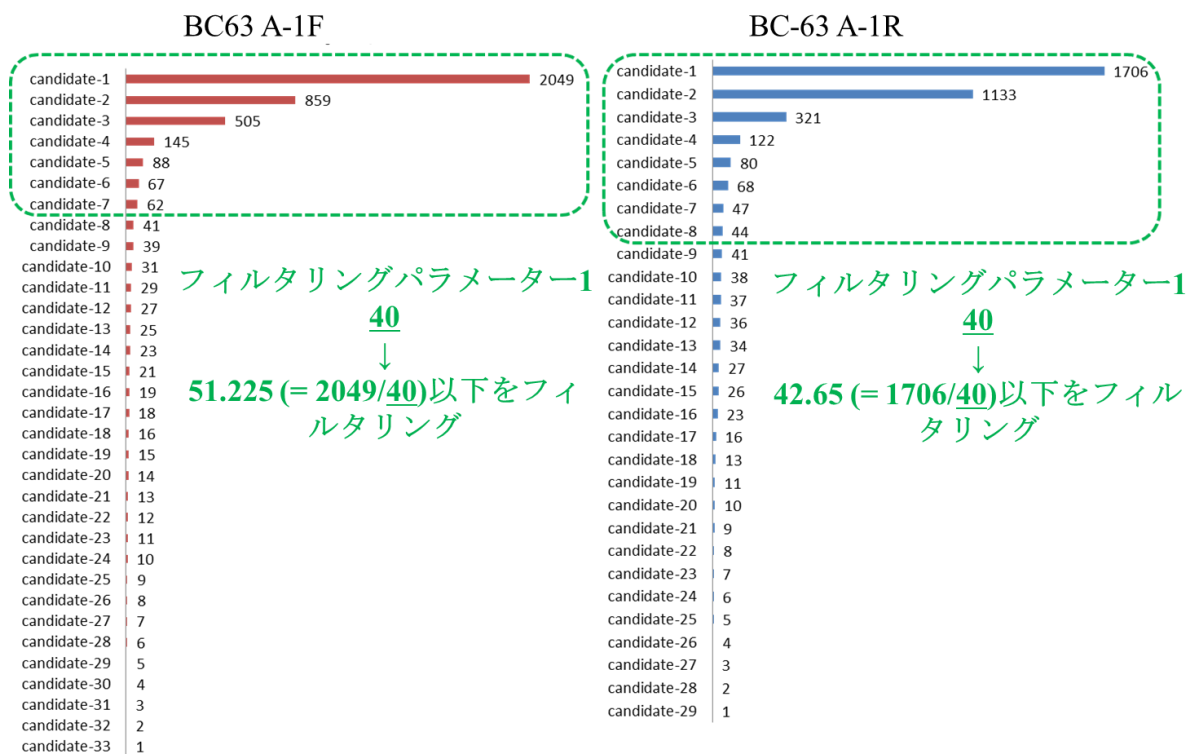


図 10 : フィルタリングパラメーター1 によるフィルタリング例
 BC63 サンプルの HLA-A の 1F、1R 領域で作成された各ハプロタイプ候補配列 (A-1F : Candidate-1~33、A-1R : Candidate-1~29) のハプロタイプスコアを示す。
 緑字 : フィルタリングパラメーター1 を用いたフィルタリングの計算例を示す。

2-1-3-3. HLA データベース検索プログラム

「HLA データベース検索プログラム」ではアライメントツールを用いて、HLA cDNA データベース (<http://www.ebi.ac.uk/ipd/imgt/hla/>) への検索を行う。

作成したハプロタイプ候補配列ファイル (項目 2-1-3-2 を参照) を HLA cDNA データベースに対してアライメントツールである BLAST-like alignment tool : BLAT (<http://genome.ucsc.edu/>) を使い、アライメントさせた。配列の一致度からタイプを選定するための指標として「MR (Matched Rate)」を設定した。MR はハプロタイプ候補配列が HLA cDNA データベースの各 HLA タイプ配列と重なっている領域 (「重複領域」とする) 中で一致した塩基数の割合である (図 11A)。

BLAT (<http://genome.ucsc.edu/>) は、連続した類似配列内の塩基置換、挿入欠失のみを検出する。そのため図 11B の場合ハプロタイプ候補配列 A はタイプ 1、タイプ 2 ともに塩基置換は 0、挿入欠失も 0 というアライメント結果になる。しかし、実際はタイプ 1 の塩基配列に対しては 40 塩基の完全一致、タイプ 2 の塩基配列は 32 塩基と部分一致である。即ち BLAT の出力のみでは一致不一致を判定できない場合があるため、MR を導入し $MR = 1$ を完全一致の指標とした。タイプ 1 は $MR = 1$ 、タイプ 2 では $MR = 0.8$ となる。MR の計算方法の詳細は項目 4-3-3 において説明する。



図 11 : MR (Matched Rate) について

(A) 一致率の計算方法

赤枠で示した領域 : 2 配列 (タイプとハプロタイプ候補配列) で塩基配列が一致している領域。

(B) ハプロタイプ候補配列 A のタイプ 1、2 それぞれへのアライメント例

タイプ 1 : 不一致塩基数、欠失・挿入塩基数が 0 で、重複領域で全ての塩基が一致している例。

タイプ 2 : 不一致塩基数、欠失・挿入塩基数が 0 であるが、重複領域で部分一致している例。

2-1-3-4. HLA タイプ決定プログラム

「HLA タイプ決定プログラム」では、HLA データベース検索結果（項目 2-1-3-3 を参照）を用いて、全解析領域（HLA クラス I：1F~4R、HLA クラス II：1F~3R）において共通して完全に配列が一致した HLA タイプを選定する処理を先ず行う。大部分のサンプルでは、全領域（HLA クラス I：1F~4R、HLA クラス II：1F~3R）で共通して完全一致したタイプ（MR=1）を出力させると 2 タイプの出力データとなり、ハプロタイプを決定できた（図 12）。

BC63	1F	1R	2F	2R	3F	3R	4F	4R
A*02:06:01	3025	1706	417	453	1885	1026	2599	878
A*24:02:01	650	1133	483	279	303	325	293	902

BC63	1F	1R	2F	2R	3F	3R	4F	4R
B*56:03	356	638	902	479	580	314	245	287
B*54:01:01	242	129	263	76	276	260	245	287

図 12：追加処理を必要とせずハプロタイプを決定できた例

BC63 サンプルの HLA-A, -B において出力された共通タイプと各領域のハプロタイプスコア。

しかし、一部サンプルは、当処理のみではハプロタイプを決定できず、以下で説明するような追加処理が必要であった。追加処理が必要なケースは以下のとおりである。

1. NGS の読み取りエラーのため正解となる配列を絞り込めない場合。
2. NGS の読み取りエラーの中で特に欠失エラーのため MR ≠ 1 になる場合。
3. 別コンティグに多型がある場合。
4. HLA cDNA データベースに登録された HLA 配列が不完全な場合。

2-1-3-4-1. NGS の読み取りエラーのため正解となる配列を絞り込めない場合

NGS の読み取りエラーのためフィルタリングパラメーター1（項目 2-1-3-2 を参照）による処理では、目的とするアレル以外のハプロタイプ候補配列が残ってくる場合がある。このようなケースに対し、フィルタリングパラメーター2 を導入して解決した。

ハプロタイプスコアは当該タイプと完全に合致したリード数になるため、ハプロタイプスコアが同一の場合、同じ配列を持っている（項目 2-1-3-2 を参照、もちろん非常に低い確率で異なる配列である可能性はある）。ハプロタイプスコアが一致する領域が多数を占める場合、一致しない領域は読み取りエラーによる可能性が高い。

このようなケースを判別する方法を図 13 の例を用い説明する。BC92 サンプル HLA-A において 1F~4R で完全に一致した HLA タイプは、3 つ A*02:01:01、A*31:01:02、

A*31:14N である。A*02:01:01、A*31:01:02 は全領域に渡ってハプロタイプスコアが異なるため、すべての領域の配列が異なっている。それに対し、A*31:01:02、A*31:14N は 2R, 3F 領域以外は同スコアで、A*31:14N の 2R, 3F 領域はハプロタイプスコアが低いことから NGS の読み取りエラーのため生じたと考えられる。

A*02:01:01、A*31:01:02、A*31:14N の 3 タイプをマルチプルアライメントさせた結果 (図 14) をみると、A*02:01:01 と A*31:01:02 は 1F~4R の全領域に渡り、塩基が異なる箇所が存在する。それに対し、A*31:01:02 と A*31:14N のマルチプルアライメントデータは、A*31:14N は、A*31:01:02 と比べ 2R、3F 領域のうち、黄色の四角で示している所にシトシンの挿入がみられた。

BC92 サンプル等の例から、領域数の半数以上が同スコアである場合、同じグループ (「アレルグループ」とする) とする規則を設けた。図 13 の場合、8 領域の半数は 4 であるため、6 領域で同スコアである A*31:01:02 と A*31:14N は同じアレルグループとなる。アレルグループ内のフィルタリング処理のためフィルタリングパラメーター 2 を設置し、項目 2-1-3-2-2 と類似の操作を領域ごとで行い、ハプロタイプスコアが低いアレル (HLA-A の場合、フィルタリングパラメーター 2 は「5」としたため、最も高いハプロタイプスコアの 1/5 以下のスコアを有するアレル) を除去した。

BC92 MR=1	1F	1R	2F	2R	3F	3R	4F	4R
A*02:01:01	5273	3816	925	1291	1771	1191	4322	1658
A*31:01:02	138	2996	436	523	249	1323	2836	1483
A*31:14N	138	2996	436	53	82	1323	2836	1483

フィルタリングパラメーター 2 により除去されるタイプ



 アレルグループ 1
 アレルグループ 2

図 13 : NGS の読み取りエラーのため正解となる配列を絞り込めない場合の例

BC92 サンプルにおいてフィルタリングパラメーター 2 による追加処理が必要な例を示す。

A*31:14N は、フィルタリングにより除去されるタイプである。

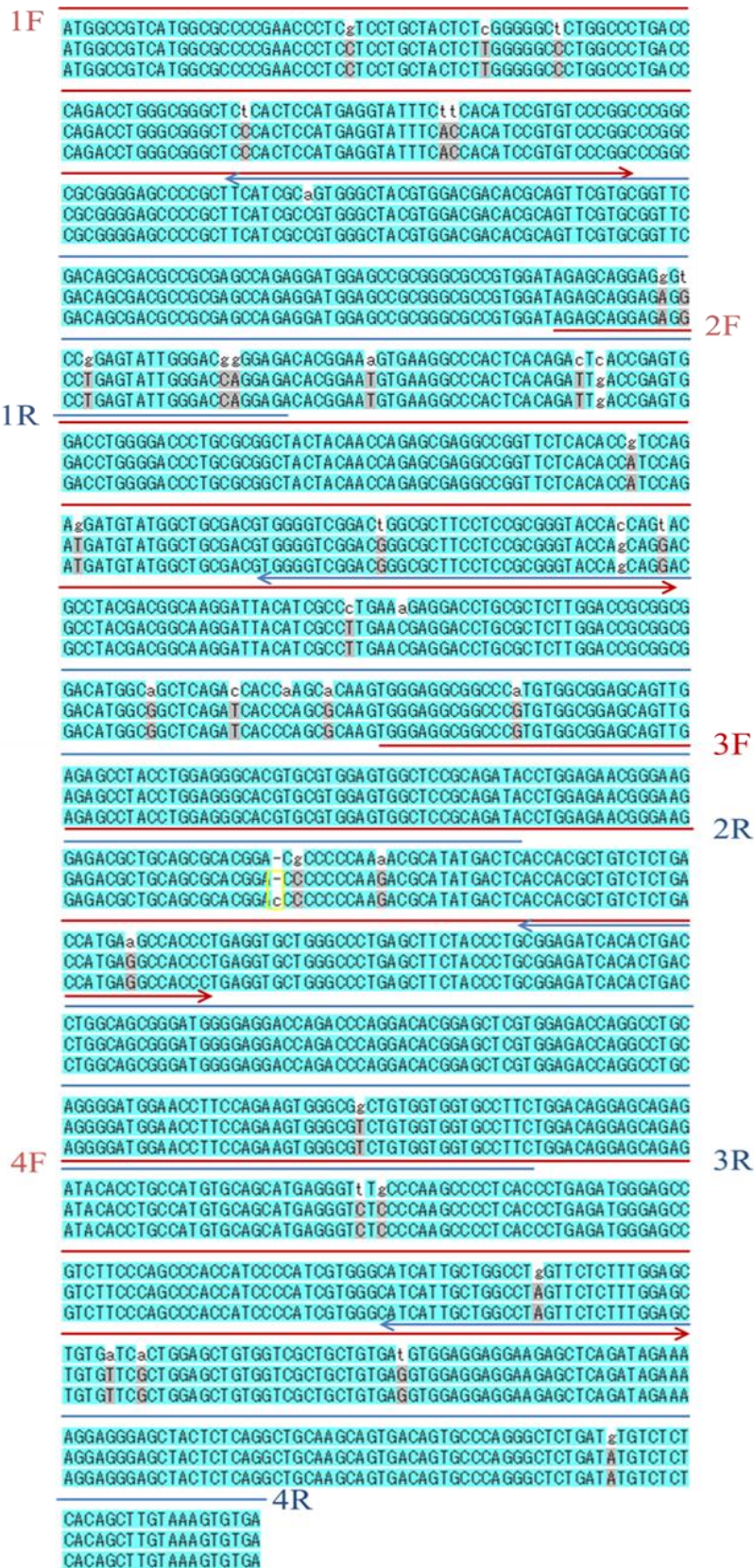


図 14 : マルチプルアライメントの結果例

BC92 サンプル、HLA-A 遺伝子座のマルチプルアライメントの結果。
 1 段目 : A*02:01:01、2 段目 : A*31:01:02、3 段目 : A*31:14N
 黄色枠 : A*31:14N と A*31:01:02 とで配列が異なる箇所 (シトシンの挿入)。

各遺伝子座のフィルタリングパラメーター1、2は、表3で示した組み合わせごとに正解率を算出し、最も正解率が高い組み合わせを採用した。正解率は、サンガー法（項目4-4を参照）で決定したHLAアレル（212アレル）の内、HLAタイピングプログラムで決定できたアレル数の比率である。HLA-Aでは、フィルタリングパラメーター1、2がそれぞれ40、5である組み合わせ、HLA-B、-C、-DPA1では、フィルタリングパラメーター1、2がそれぞれ100、10、HLA-DRB1/3/4/5、-DQB1、-DQA1では、80、10、HLA-DPB1では、40、10となった。

2種類のパラメーターの組み合わせ (フィルタリングパラメーター1、2)		100、10	100、5	80、10	80、5	40、10	40、5
遺伝子座ごとの タイピング 正解率 (%)	A	83	96	83	96	88	100
	B	100	100	100	100	100	100
	C	100	100	92	92	92	92
	DRB1	100	100	100	100	100	100
	DRB3/4/5	90	90	95	95	95	95
	DQB1	96	96	100	100	96	96
	DQA1	100	100	100	100	96	96
	DPB1	92	92	92	92	100	100
	DPA1	100	100	100	100	100	100

：各遺伝子座で設定したパラメーター値の組み合わせ

表3：2種類のパラメーターの組み合わせによるHLAタイピング正解率

2-1-3-4-2. NGSの読み取りエラーの中で特に欠失エラーのためMR≠1になる場合

項目2-1-3-1で述べたようにNGSの出力データを確認したところ、約350bp以上解析されたリードでは後半の塩基読み取りエラーが多く見られる傾向にあったが、図15Aで示すように、フォワードリード、リバースリード相互排他的に配列ファイル作成領域内で、欠失エラーが生じているサンプルがいくつか確認された（表4）。タイプごとで決まった位置に図15Aのような欠失が生じているため配列依存的なエラーの可能性が高い。解析する各領域（HLAクラスI：1F~4R、HLAクラスII：1F~3R）で1領域でもこのような欠失エラーが生じると、全領域で共通に完全一致したタイプを選定することはできない。

このような欠失エラーに対応するために以下の処理（図15B）を行なった。MR=1で共通タイプ（図15B、①）を出力できなかった場合は、MR≧0.99、欠失が1以下、不一致塩基数が0で共通タイプを検索（図15B、②）、この条件で2アレル決定できなければ、MR≧0.99、欠失が2以下、不一致塩基数が0で共通タイプを検索（図15B、③）、というように検索条件を順次変更することにより配列ファイル作成領域内の欠失エラーに対応させる手法を考案した（図15B）。テストセットの配列データを確認したところ、このような欠失エラーが同じ領域内に2箇所出現する例が確認されたので、(HLA-C*07:02:01、3F領域)、当HLAタイピングプログラムでは欠失を2つまで許すこととした。

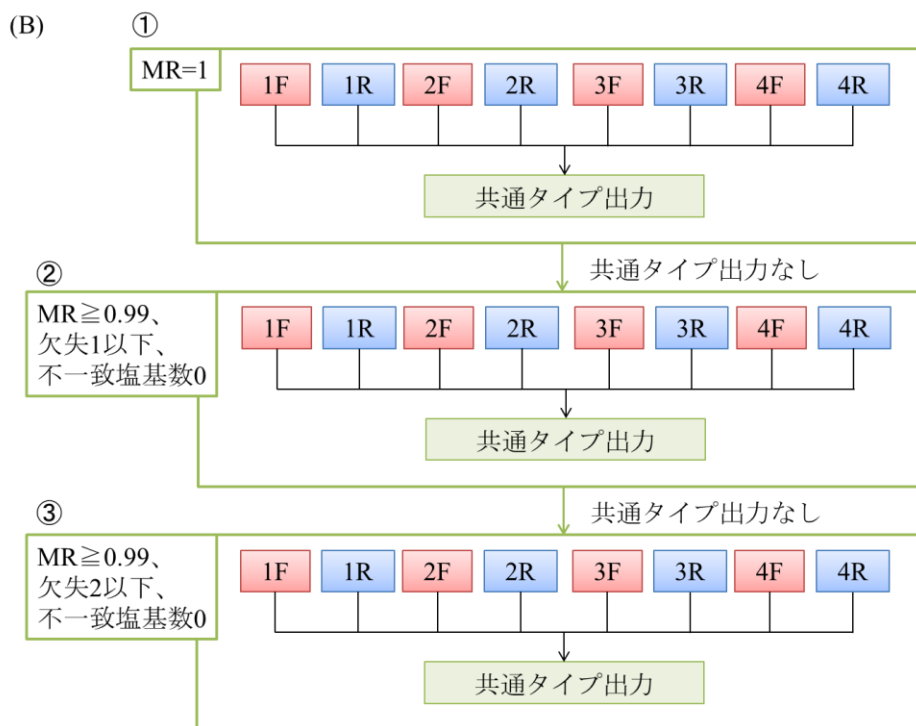
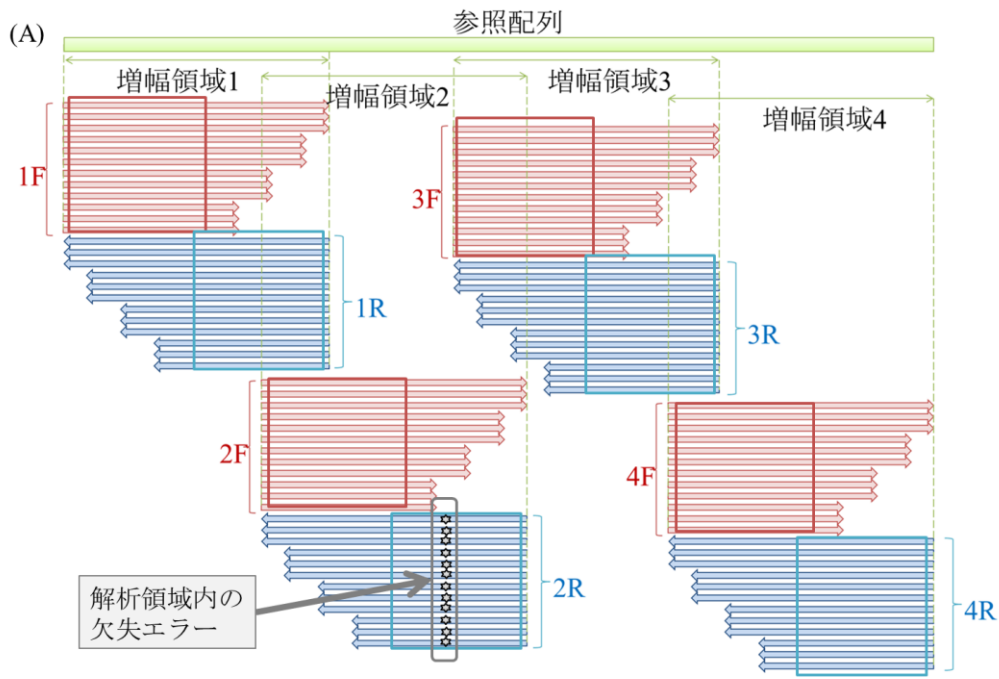


図 15 : NGS の読み取りエラーの中で特に欠失エラーのため MR \neq 1 になる場合

(A) NGS で解析されたリードの中に欠失エラーが存在する際のイメージ図。

赤枠、青枠：配列ファイル作成領域。

(B) 欠失エラーが生じた場合の追加処理の流れ

- ① MR=1 で共通タイプ。
- ② MR \geq 0.99、欠失が 1 以下、不一致塩基数が 0 で共通タイプを検索。
- ③ MR \geq 0.99、欠失が 2 以下、不一致塩基数が 0 で共通タイプを検索。

タイプ	欠失領域	欠失数	サンプル数
HLA-B*15:11:01	1F	1	1
HLA-B*46:01:01	1F	1	1
HLA-C*7:02:01	3F	2	1
HLA-DQA1*01:01:01	3R	1	1
HLA-DQA1*01:03:01	3R	1	1
HLA-DQA1*03:01:01	3R	1	2
HLA-DQA1*03:02	3R	1	3
HLA-DQA1*03:03:01	3R	1	2
HLA-DQA1*05:08	3R	1	1

表 4：解析領域内の欠失エラー

2-1-3-4-3. 別コンティグに多型がある場合

NGS で配列決定できる鎖長はおよそ 400 塩基であるため、2 つ以上の配列領域にまたがる多型の場合、関係のないハプロタイプが選択されてくる場合がある。図 16A に一例を示す。この例の場合、318 番と 664 番の塩基に多型があったが、別々のコンティグにあったため、3 つの候補が残ってしまった。前項で述べたハプロタイプスコアが同一の場合同じ配列を持つ、という特徴をこの場合も活用した。

各領域に 1 種類しかないハプロタイプスコアを下線で強調した (図 16B)。DPB1*02:01:02、DPB1*04:02:01 は固有の配列を持つが、DPB1*105:01 は全 2 者と共通の配列しかもたない。従って DPB1*105:01 は DPB1*02:01:02、DPB1*04:02:01 の部分配列から生成されたものと結論付けることができる。

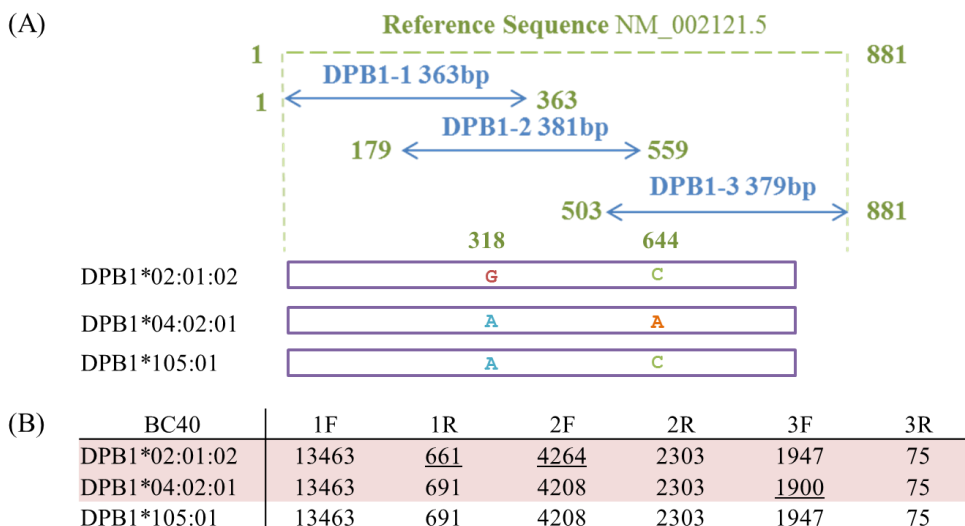


図 16：別コンティグに多型がある場合の追加処理が必要なサンプル例

- (A) 各増幅領域において共通にマッチした 3 つの HLA タイプの多型の位置関係を示す。
 数字 (緑)：マッピングに使用する参照配列の最初の塩基を 1 とした場合の位置。
 (B) 出力された 3 タイプと各領域のハプロタイプスコアを表す。

2-1-3-4-4. HLA cDNA データベースに登録された HLA 配列が不完全な場合

HLA cDNA データベースに登録されている多くのアレルは調べられていない領域を含む不完全な配列である。そのため、図 6 に示したように HLA cDNA データベースに登録されているデータの長さは様々である。2 アレルのうち、一方が長い配列を有するタイプで、もう一方が短い配列を有するタイプである場合、解析領域 1F~4R (HLA クラス I) もしくは 1F~3R (HLA クラス II) で共通のタイプを選択すると長い配列タイプしかデータとして出力されないケースが想定される。そのため HLA クラス I では 1F~4R で 2 アレル決定できなければ、1F~4F で共通のタイプを検索、1F~4F で 2 アレル決定できなければ、1F~3R で検索、1F~3R で 2 アレル決定できなければ、1F~3F で検索、というように検索領域を順次縮小することにより HLA cDNA データベースに登録されている様々な長さの配列を有する HLA タイプに対応させる手法を考案した。HLA クラス II でも同様に、1F~3R で 2 アレル決定できなかった場合、1F~2F まで検索した。

このような追加処理を必要としたハプロタイプ例を図 17 に示した。図 17A は、BC63 サンプルの HLA-DQB1 におけるタイプと各解析領域との位置関係を示している。このサンプルの HLA-DQB1 のタイプは、長さが 786 bp である DQB1*05:02:01 と長さが 552 bp である DQB1*04:08 である。3F 領域、3R 領域のハプロタイプ候補配列は DQB1*04:08 には重ならないため、1F~3R もしくは 1F~3F で完全に一致したタイプを検索すると、図 17B で示すように DQB1*05:02:01 は出力されるが、DQB1*04:08 は出力されない。DQB1*04:08 をカバーする 1F~2R で共通のタイプを検索すると DQB1*04:08 のデータが出力され、最終的にこのサンプルのハプロタイプは DQB1*05:02:01 と DQB1*04:08 なる。

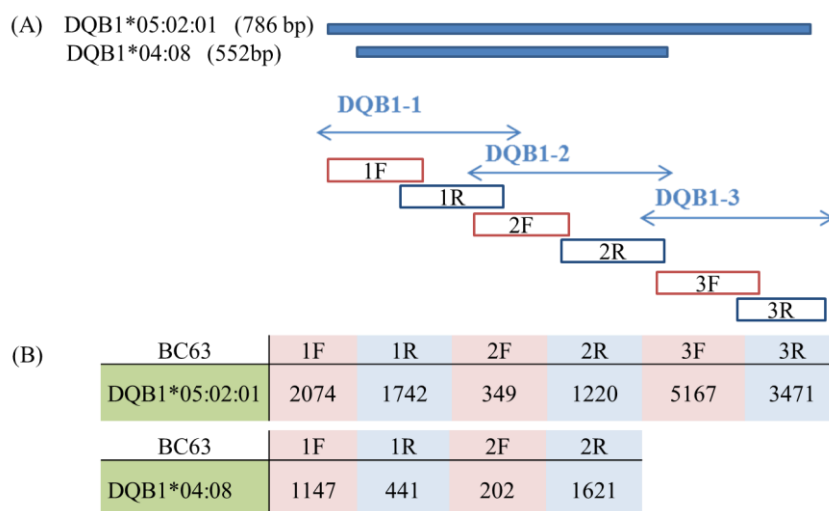


図 17 : HLA cDNA データベースに登録された HLA 配列が不完全な場合の例

(A) BC63 サンプルの HLA-DQB1 のハプロタイプと、1F~3R 領域の位置関係を表す。

(B) BC63 サンプルの DQB1 において出力された共通タイプと各領域のハプロタイプスコアを表す。

以上 4 つの追加処理を組み合わせた解析ラインを図 22 (HLA タイピングシステムと補完 HLA タイピングシステムを組み合わせた解析フローチャート) に示した。図 22 の解析フローチャートのうち HLA タイピングプログラムを詳細に示したものが図 25 であり、図 25 に関する詳しい説明は項目 4-3-4 で行う。

2-2. 開発した HLA タイピングシステムの検証

設定したパラメーターによる HLA タイピング結果 (テストセット) を表 5 に、正解率を表 6 に示した。開発した HLA タイピングプログラムにより 6 桁レベルで決定した全遺伝子座のアレル数は、211 アレル、サンガー法により決定した全遺伝子座のアレル数は 212 アレルであるため、正解率は 99.5 % となった。また、抗原結合領域 (G グループレベル) の HLA タイピング精度 (表 6) も示した。G グループレベル HLA タイピングを行う際は、HLA タイプ検索プログラム終了後、G グループが存在しているタイプは、その G グループ名に置き換え、HLA タイピング決定プログラムを実行させて出力させた。増幅バイアスが影響しサンガー法により決定した HLA タイプが出力されず、決定できなかったアレルが 6 桁レベル、G グループレベル HLA タイピング共に 1 アレル存在した (S サンプル、HLA-DRB3/4/5 遺伝子座)。

sample	A		B		C		DRB1		DRB3/4/5	
	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2
BC63	A*24:02:01	A*02:06:01	B*56:03	B*54:01:01	C*01:02:01	C*01:02:01	DRB1*14:54:01	DRB1*04:05:01	DRB3*02:02:01	DRB4*01:03:01
BC92	A*02:01:01	A*31:01:02	B*40:02:01	B*48:01:01	C*03:04:01	C*03:03:01	DRB1*11:01:01	DRB1*04:06:01	DRB3*02:02:01	DRB4*01:03:01
BC102	A*26:01:01	A*24:02:01	B*52:01:01	B*40:02:01	C*03:04:01	C*12:02:02	DRB1*09:01:02	DRB1*15:02:01	DRB4*01:03:02	DRB5*01:02
BC159	A*26:03:01	A*02:07:01	B*51:01:01	B*51:01:01	C*01:02:01	C*14:02:01	DRB1*14:03:01	DRB1*08:03:02	DRB4*01:03:01	
BC40	A*24:02:01	A*11:01:01	B*52:01:01	B*15:01:01	C*04:01:01	C*12:02:02	DRB1*04:06:01	DRB1*04:05:01	DRB4*01:03:01	
BC70	A*26:03:01	A*33:03:01	B*51:02:01	B*15:01:01	C*07:02:01	C*15:02:01	DRB1*14:06:01	DRB1*14:54:01	DRB3*02:02:01	DRB3*02:02:01
BC74	A*26:01:01	A*24:02:01	B*52:01:01	B*40:02:01	C*03:04:01	C*12:02:02	DRB1*09:01:02	DRB1*15:02:01	DRB4*01:03:02	DRB5*01:02
BC35	A*24:02:01	A*24:02:01	B*40:01:02	B*07:02:01	C*07:02:01	C*04:82	DRB1*12:01:01	DRB1*01:01:01	DRB3*01:01:02	
BC109	A*26:01:01	A*11:01:01	B*07:02:01	B*15:01:01	C*07:02:01	C*04:01:01	DRB1*04:06:01	DRB1*09:01:02	DRB4*01:03:01	DRB4*01:03:02
N	A*31:01:02	A*02:06:01	B*40:01:02	B*51:01:01	C*03:04:01	C*14:02:01	DRB1*09:01:02	DRB1*04:06:01	DRB4*01:03:01	DRB4*01:03:02
S	A*02:01:01	A*02:07:01	B*15:11:01	B*46:01:01	C*03:03:01	C*01:02:01	DRB1*12:01:01	DRB1*04:03:01	DRB3*01:01:02	DRB4*01:03:01
K	A*24:02:01	A*24:02:01	B*07:02:01	B*07:02:01	C*07:02:01	C*07:02:01	DRB1*01:01:01	DRB1*01:01:01		

sample	DQB1		DQA1		DPB1		DPA1	
	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2
BC63	DQB1*04:08	DQB1*05:02:01	DQA1*03:03:01	DQA1*01:04:01	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
BC92	DQB1*03:01:01	DQB1*03:02:01	DQA1*03:01:01	DQA1*05:05:01	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
BC102	DQB1*03:03:02	DQB1*06:01:01	DQA1*03:02	DQA1*01:03:01	DPB1*09:01:01	DPB1*05:01:01	DPA1*02:01:01	DPA1*02:02:02
BC159	DQB1*03:01:01	DQB1*06:01:01	DQA1*05:07	DQA1*01:03:01	DPB1*05:01:01	DPB1*02:02	DPA1*02:02:02	DPA1*02:02:02
BC40	DQB1*03:02:01	DQB1*04:08	DQA1*03:01:01	DQA1*03:03:01	DPB1*04:02:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
BC70	DQB1*03:01:01	DQB1*05:02:01	DQA1*05:03	DQA1*01:04:01	DPB1*02:01:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
BC74	DQB1*06:01:01	DQB1*03:03:02	DQA1*01:03:01	DQA1*03:02	DPB1*09:01:01	DPB1*04:02:01	DPA1*01:03:01	DPA1*02:01:01
BC35	DQB1*05:01:01	DQB1*03:03:02	DQA1*01:01:01	DQA1*03:02	DPB1*04:02:01	DPB1*03:01:01	DPA1*01:03:01	DPA1*02:02:02
BC109	DQB1*03:03:02	DQB1*03:02:01	DQA1*03:01:01	DQA1*03:02	DPB1*02:01:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
N	DQB1*03:03:02	DQB1*03:02:01	DQA1*03:01:01	DQA1*03:02	DPB1*02:01:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
S	DQB1*03:02:01	DQB1*03:02:01	DQA1*05:08	DQA1*03:01:01	DPB1*05:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
K	DQB1*05:01:01	DQB1*05:01:01	DQA1*01:01:01	DQA1*01:01:01	DPB1*04:02:01	DPB1*135:01	DPA1*02:02:02	DPA1*01:03:01

黒字: HLAタイピングプログラムにより決定できたアレル
 赤字: HLAタイピングプログラムにより決定できなかったアレル
 紫字: DRB3/4/5が同一アレル上に存在しないDRB1タイプ (DRB1の2桁目が01, 08, 10であるタイプ)
 :ホモ接合体サンプルとして判定したアレル
 :増幅バイアスが原因でタイプを決定できなかったアレル

表 5: 設定したパラメーターでの HLA タイピング結果 (テストセット)

6桁レベル	Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
Program (alleles)	211	24	24	24	24	19	24	24	24	24
Manually (alleles)	212	24	24	24	24	20	24	24	24	24
Accuracy (%)	99.5	100	100	100	100	95.0	100	100	100	100

Gグループレベル	Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
Program (alleles)	211	24	24	24	24	19	24	24	24	24
Manually (alleles)	212	24	24	24	24	20	24	24	24	24
Accuracy (%)	99.5	100	100	100	100	95.0	100	100	100	100

Program : HLAタイピングプログラムにより決定したアレル数

Manually : サブクローニングを要したサンガー法により決定したアレル数

Accuracy : サンガー法により決定したアレルのうちHLAタイピングプログラムにより決定できた割合

表 6 : 設定したパラメーターでの HLA タイピング結果 (テストセット)

次に、開発した HLA タイピングシステムの性能を評価するために、検証セット (30 サンプル : 乳癌組織由来 RNA サンプル 9 例、大腸癌組織由来 RNA サンプル 21 例) で HLA タイピングを行った。NGS による配列解析は 2 回 (14 サンプルと 16 サンプル) にわけて行った。表 7 に、検証セットにおいて、サンガー法により決定した HLA タイプを示した。そのうち、赤字で示した HLA タイプが開発した HLA タイピングプログラムで決定できなかったタイプとなる。当結果より、正解率を計算した結果、HLA タイピングプログラムにより 6 桁レベルで決定した全遺伝子座のアレル数は、516 アレル、サンガー法により決定した全遺伝子座のアレル数は 532 アレルであるため、正解率は 97.0 % となった (表 8)。G グループレベルの HLA タイピングでは、523 アレル決定することができ、正解率は 98.3 % となった (表 8)。以上のように開発した HLA タイピングシステムは、高い精度を示す方法であることを証明することができた。

sample	A		B		C		DRB1		DRB3/4/5	
	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2
BC9	A*24:02:01	A*31:01:02	B*40:01:02	B*55:02:01	C*03:04:01	C*01:02:01	DRB1*04:05:01	DRB1*15:01:01	DRB4*01:03:01	DRB5*01:01:01
BC55	A*24:02:01	A*24:20	B*54:01:01	B*35:01:01	C*01:02:01	C*03:03:01	DRB1*11:01:01	DRB1*04:05:01	DRB3*02:02:01	DRB4*01:03:01
BC2	A*02:10	A*26:01:01	B*40:06:01	B*40:02:01	C*03:03:01	C*08:01:01	DRB1*04:06:01	DRB1*04:05:01	DRB4*01:03:01	DRB4*01:03:01
BC1	A*24:02:01	A*24:02:01	B*51:01:01	B*40:01:02	C*15:02:01	C*07:02:01	DRB1*14:06:01	DRB1*09:01:02	DRB3*02:02:01	DRB4*01:03:02
BC46	A*24:02:01	A*24:02:01	B*52:01:01	B*52:01:01	C*12:02:02	C*12:02:02	DRB1*04:05:01	DRB1*15:02:01	DRB4*01:03:01	DRB5*01:02
BC148	A*26:02:01	A*24:02:01	B*40:02:01	B*55:02:01	C*03:04:01	C*01:02:01	DRB1*14:54:01	DRB1*09:01:02	DRB3*02:02:01	DRB4*01:03:02
BC68	A*24:02:01	A*02:06:01	B*35:01:01	B*51:01:01	C*03:03:01	C*14:02:01	DRB1*09:01:02	DRB1*09:01:02	DRB4*01:03:02	DRB4*01:03:02
BC27	A*31:01:02	A*11:01:01	B*54:01:01	B*15:01:01	C*03:03:01	C*01:02:01	DRB1*15:01:01	DRB1*08:03:02	DRB5*01:01:01	
BC18	A*02:06:01	A*11:01:01	B*54:01:01	B*15:01:01	C*03:03:01	C*01:02:01	DRB1*12:02:01	DRB1*04:05:01	DRB3*03:01:03	DRB4*01:03:01
CC1	A*24:02:01	A*24:03:01	B*40:06:01	B*52:01:01	C*08:01:01	C*12:02:02	DRB1*09:01:02	DRB1*15:02:01	DRB4*01:03:02	DRB5*01:02
CC2	A*02:01:01	A*11:01:01	B*54:01:01	B*15:01:01	C*04:01:01	C*01:02:01	DRB1*04:06:01	DRB1*04:05:01	DRB4*01:03:01	DRB4*01:03:01
CC3	A*24:02:01	A*24:02:01	B*52:01:01	B*40:06:01	C*08:01:01	C*12:02:02	DRB1*09:01:02	DRB1*09:01:02	DRB4*01:03:02	DRB4*01:03:02
CC4	A*24:02:01	A*31:01:02	B*40:01:02	B*56:01:01	C*04:01:01	C*01:02:01	DRB1*12:01:01	DRB1*09:01:02	DRB3*01:01:02	DRB4*01:03:02
CC5	A*02:01:01	A*24:02:01	B*52:01:01	B*15:18:01	C*07:04:01	C*12:02:02	DRB1*04:01:01	DRB1*15:02:01	DRB4*01:02	DRB5*01:02
CC6	A*02:06:01	A*11:01:01	B*40:06:01	B*44:03:01	C*14:03	C*08:01:01	DRB1*11:01:01	DRB1*12:01:01	DRB3*01:01:02	DRB5*02:02:01
CC7	A*02:01:01	A*24:02:01	B*52:01:01	B*15:11:01	C*03:03:01	C*12:02:02	DRB1*09:01:02	DRB1*15:02:01	DRB4*01:03:02	DRB5*01:02
CC11	A*02:01:01	A*26:01:01	B*59:01:01	B*40:02:01	C*03:04:01	C*01:02:01	DRB1*14:05:01	DRB1*09:01:02	DRB3*02:02:01	DRB4*01:03:01
CC12	A*02:01:01	A*24:02:01	B*35:01:01	B*07:02:01	C*08:01:01	C*07:02:01	DRB1*12:01:01	DRB1*15:02:01	DRB3*01:01:02	DRB5*01:02
CC59	A*24:02:01	A*24:02:01	B*51:01:01	B*07:02:01	C*01:02:01	C*07:02:01	DRB1*12:01:01	DRB1*01:01:01	DRB3*01:01:02	
CC14	A*02:06:01	A*24:02:01	B*51:01:01	B*54:01:01	C*14:02:01	C*01:02:01	DRB1*04:05:01	DRB1*09:01:02	DRB4*01:03:01	DRB4*01:03:02
CC15	A*24:02:01	A*11:01:01	B*35:01:01	B*07:02:01	C*03:03:01	C*07:02:01	DRB1*15:01:01	DRB1*01:01:01	DRB5*01:01:01	
CC20	A*24:02:01	A*24:02:01	B*59:01:01	B*54:01:01	C*01:02:01	C*01:02:01	DRB1*04:05:01	DRB1*04:05:01	DRB4*01:03:01	DRB4*01:03:01
CC21	A*24:02:01	A*24:02:01	B*40:02:01	B*52:01:01	C*12:02:02	C*03:03:01	DRB1*14:05:01	DRB1*04:03:01	DRB3*02:02:01	DRB4*01:03:01
CC23	A*33:03:01	A*33:03:01	B*44:03:01	B*44:03:01	C*01:02:01	C*14:03	DRB1*13:02:01	DRB1*08:03:02	DRB3*03:01:01	
CC27	A*02:06:01	A*26:01:01	B*40:02:01	B*40:01:02	C*03:04:01	C*03:04:01	DRB1*09:01:02	DRB1*04:05:01	DRB4*01:03:02	DRB4*01:03:01
CC33	A*02:06:01	A*02:06:01	B*46:01:01	B*46:01:01	C*01:02:01	C*01:12:02	DRB1*08:03:02	DRB1*08:03:02		
CC36	A*02:01:01	A*31:01:02	B*40:01:02	B*51:01:01	C*03:04:01	C*14:02:01	DRB1*09:01:02	DRB1*04:05:01	DRB4*01:03:01	DRB4*01:03:02
CC39	A*24:20	A*31:01:02	B*40:02:01	B*51:01:01	C*03:04:01	C*14:02:01	DRB1*04:05:01	DRB1*04:04:01	DRB4*01:03:01	DRB4*01:03:01
CC44	A*26:01:01	A*24:02:01	B*40:01:02	B*52:01:01	C*12:02:02	C*15:02:01	DRB1*08:03:02	DRB1*01:01:01		
CC38	A*02:01:01	A*24:02:01	B*40:06:01	B*40:06:01	C*08:01:01	C*01:02:01	DRB1*09:01:02	DRB1*09:01:02	DRB4*01:03:02	DRB4*01:03:02

sample	DQB1		DQA1		DPB1		DPA1	
	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2	Allele 1	Allele2
BC9	DQB1*06:02:01	DQB1*04:08	DQA1*01:02:01	DQA1*03:03:01	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
BC55	DQB1*03:01:01	DQB1*04:08	DQA1*05:05:01	DQA1*03:03:01	DPB1*03:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*01:03:01
BC2	DQB1*03:02:01	DQB1*04:08	DQA1*03:01:01	DQA1*03:03:01	DPB1*02:01:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
BC1	DQB1*03:03:02	DQB1*03:01:01	DQA1*05:03	DQA1*03:02	DPB1*02:01:02	DPB1*05:01:01	DPA1*01:03:01	DPA1*02:02:02
BC46	DQB1*06:01:01	DQB1*04:08	DQA1*01:03:01	DQA1*03:03:01	DPB1*04:02:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
BC148	DQB1*05:03:01	DQB1*03:03:02	DQA1*01:04:01	DQA1*03:02	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
BC68	DQB1*03:03:02	DQB1*03:03:02	DQA1*03:02	DQA1*03:02	DPB1*05:01:01	DPB1*04:02:01	DPA1*01:03:01	DPA1*02:02:02
BC27	DQB1*06:01:01	DQB1*06:02:01	DQA1*01:02:01	DQA1*01:03:01	DPB1*02:01:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
BC18	DQB1*03:01:01	DQB1*04:08	DQA1*03:03:01	DQA1*06:01:01	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
CC1	DQB1*06:01:01	DQB1*03:03:02	DQA1*03:02	DQA1*01:03:01	DPB1*09:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:01:01
CC2	DQB1*03:02:01	DQB1*04:08	DQA1*03:03:01	DQA1*03:03:01	DPB1*05:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
CC3	DQB1*03:03:02	DQB1*03:03:02	DQA1*03:02	DQA1*03:02	DPB1*02:01:02	DPB1*13:01:01	DPA1*01:03:01	DPA1*02:01:01
CC4	DQB1*03:01:01	DQB1*03:03:02	DQA1*05:05:01	DQA1*03:02	DPB1*05:01:01	DPB1*06:01	DPA1*01:03:01	DPA1*02:02:02
CC5	DQB1*06:01:01	DQB1*03:01:01	DQA1*01:03:01	DQA1*03:03:01	DPB1*05:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
CC6	DQB1*03:01:01	DQB1*03:01:01	DQA1*05:05:01	DQA1*05:05:01	DPB1*02:02	DPB1*06:01	DPA1*01:03:01	DPA1*01:03:01
CC7	DQB1*06:01:01	DQB1*03:03:02	DQA1*03:02	DQA1*01:03:01	DPB1*05:01:01	DPB1*09:01:01	DPA1*02:01:01	DPA1*02:02:02
CC11	DQB1*03:03:02	DQB1*05:03:01	DQA1*01:04:01	DQA1*03:02	DPB1*02:01:02	DPB1*05:01:01	DPA1*01:03:01	DPA1*02:02:02
CC12	DQB1*06:01:01	DQB1*03:03:02	DQA1*01:03:01	DQA1*01:03:01	DPB1*02:01:02	DPB1*09:01:01	DPA1*01:03:01	DPA1*02:01:01
CC59	DQB1*03:01:01	DQB1*05:01:01	DQA1*05:05:01	DQA1*01:01:01	DPB1*04:02:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
CC14	DQB1*03:03:02	DQB1*04:08	DQA1*03:03:01	DQA1*03:02	DPB1*05:01:01	DPB1*04:02:01	DPA1*01:03:01	DPA1*02:02:02
CC15	DQB1*05:01:01	DQB1*06:02:01	DQA1*01:02:01	DQA1*01:01:01	DPB1*04:02:01	DPB1*02:01:02	DPA1*02:02:02	DPA1*01:03:01
CC20	DQB1*04:08	DQB1*04:08	DQA1*03:03:01	DQA1*03:03:01	DPB1*02:01:02	DPB1*05:01:01	DPA1*02:02:02	DPA1*01:03:01
CC21	DQB1*03:02:01	DQB1*05:03:01	DQA1*01:04:01	DQA1*03:01:01	DPB1*02:02	DPB1*02:01:02	DPA1*01:03:01	DPA1*01:03:01
CC23	DQB1*06:01:01	DQB1*06:04:01	DQA1*01:02:01	DQA1*01:03:01	DPB1*05:01:01	DPB1*04:01:01	DPA1*01:03:01	DPA1*02:02:02
CC27	DQB1*03:03:02	DQB1*04:08	DQA1*03:03:01	DQA1*03:02	DPB1*02:01:02	DPB1*14:01	DPA1*02:01:01	DPA1*01:03:01
CC33	DQB1*06:01:01	DQB1*06:01:01	DQA1*01:03:01	DQA1*01:03:01	DPB1*02:01:02	DPB1*02:01:02	DPA1*02:02:02	DPA1*02:02:02
CC36	DQB1*03:03:02	DQB1*04:08	DQA1*03:03:01	DQA1*03:02	DPB1*05:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
CC39	DQB1*03:02:01	DQB1*04:08	DQA1*03:03:01	DQA1*01:03:01	DPB1*03:01:01	DPB1*02:01:02	DPA1*01:03:01	DPA1*02:02:02
CC44	DQB1*06:01:01	DQB1*05:01:01	DQA1*01:01:01	DQA1*01:03:01	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02
CC38	DQB1*03:02:01	DQB1*03:02:01	DQA1*03:02	DQA1*03:02	DPB1*05:01:01	DPB1*05:01:01	DPA1*02:02:02	DPA1*02:02:02

黒字：HLAタイピングプログラムにより決定できたアレル

赤字：HLAタイピングプログラムにより決定できなかったアレル

紫字：DRB3/4/5が同一アレル上に存在しないDRB1タイプ（DRB1の2桁目が01、08、10であるタイプ）

：ホモ接合体サンプルとして判定したアレル

：増幅バイアスが原因でタイプを決定できなかったアレル

：読み取りエラーが原因でタイプを決定できなかったアレル

：HLAタイピングプログラムでは決定できなかったハプロタイプ（補完HLAタイピングシステムで決定できたハプロタイプ）

表7：設定したパラメーターでのHLAタイピング結果（検証セット）

6桁レベル	Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
Program (alleles)	516	58	57	57	59	51	57	59	58	60
Manually (alleles)	532	60	60	60	60	52	60	60	60	60
Accuracy (%)	97.0	96.7	95.0	95.0	98.3	98.1	95.0	98.3	96.7	100

Gグループレベル	Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
Program (alleles)	523	58	57	59	59	51	60	59	60	60
Manually (alleles)	532	60	60	60	60	52	60	60	60	60
Accuracy (%)	98.3	96.7	95.0	98.3	98.3	98.1	100	98.3	100	100

Program : HLAタイピングプログラムにより決定したアレル数

Manually : サブクローニングを要したサンガー法により決定したアレル数

Accuracy : サンガー法により決定したアレルのうちHLAタイピングプログラムにより決定できた割合

表 8 : 開発した HLA タイピングシステムの精度 (検証セット)

検証セットで決定できなかったアレル (6桁レベル HLA タイピングでは 16 アレル、G グループレベル HLA タイピングでは 9 アレル) に関して、HLA タイプの出力パターンを図 18 に示した。

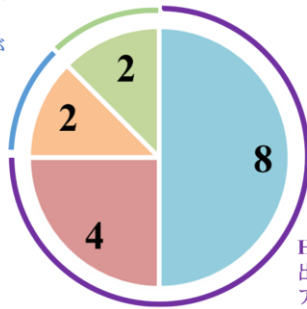
6桁レベル HLA タイピングにおいて、サンガー法で決定した HLA タイプが出力されなかったアレルは 12 アレルであり、そのうち、増幅バイアスが影響し、出力されなかったアレルは 8 アレル、NGS による読み取りエラーが原因で出力されなかったアレルは 4 アレルであった (図 18A)。誤って出力されたアレルは 2 アレルで、NGS の読み取りエラーに起因する。また、BC55 サンプルの HLA-DPB1 遺伝子座の 2 アレルは当システムでは決定不可能なハプロタイプ (補完 HLA タイピングシステムでは決定できたハプロタイプ、項目 2-3 を参照) であった (図 18A、表 7)。BC55 サンプルの HLA-DPB1 遺伝子座以外の 6桁レベル HLA タイピングで、相同染色体の両アレルとも、タイプが決定できなかったのは、CC38 サンプルの HLA-DQB1 遺伝子座のみであった (表 7)。

G グループレベル HLA タイピングにおいて、サンガー法で決定した HLA タイプが出力されなかったアレルは 7 アレルであり、そのうち、増幅バイアスが影響し出力されなかったアレルは 3 アレル、NGS による読み取りエラーが原因で出力されなかったアレルは 4 アレルであった (図 18B)。誤って出力されたアレルは 2 アレルあり、NGS による読み取りエラーが原因であった (図 18B)。

(A) 6桁レベルタイピング

補完HLAタイピングシステムによる解析が必要なアレル

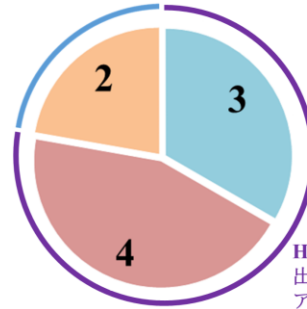
HLAタイプが一部誤って出力されたアレル



HLAタイプが出力されなかったアレル

(B) Gグループレベルタイピング

HLAタイプが一部誤って出力されたアレル



HLAタイプが出力されなかったアレル

- 増幅バイアスが原因でHLAタイプが出力されなかったアレル数
- NGSによる読み取りエラーが原因でHLAタイプが出力されなかったアレル数
- NGSによる読み取りエラーが原因でHLAタイプが一部誤って出力されたアレル数
- 補完HLAタイピングシステムによる解析が必要なアレル数

図 18 : HLA タイピングシステムで決定できなかったアレルについて

(A) 6桁レベルタイピングで決定できなかった 16 アレルの出力パターンの内訳を示す。

(B) Gグループレベルタイピングで決定できなかった 9 アレルの出力パターンの内訳を示す。

2-3. 補完 HLA タイピングシステムの開発

検証実験の結果、多型部位が複数のコンティグにわたって存在するためデータ解析方法の改良では HLA タイピングが不可能な例があることがわかった (図 19)。この検体の場合 DPB1*03:01:01、DPB1*05:01:01、DPB1*104:01、DPB1*135:01 の 4 タイプが出力された。2 つのアレルのタイプ決定のためには 341 番と 726 番の塩基を含む配列データが必要で、鋳型 DNA の調整方法から新たに開発する必要があった。新たに開発したシステムを「補完 HLA タイピングシステム」とする。

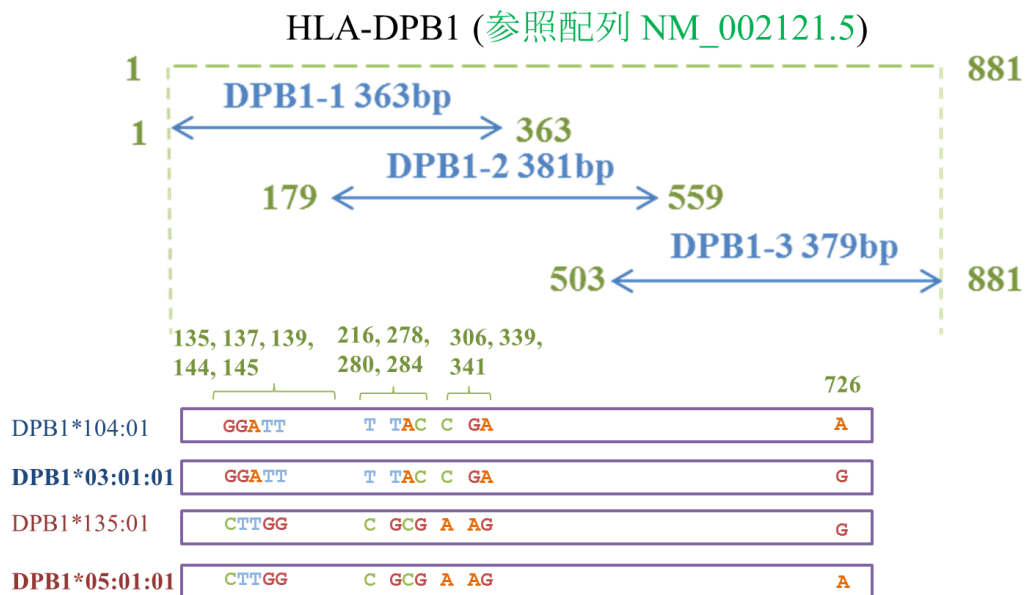


図 19 : HLA タイピングシステムでは同定できなかったハプロタイプ (BC55 サンプル、HLA-DPB1)

HLA タイピングシステムにより出力された 4 タイプの多型情報と位置関係を示す。
数字 (緑) : マッピングに使用する参照配列の最初の塩基を 1 とした場合の各多型の位置。

補完 HLA タイピングシステムのライブラリーを「補完 HLA タイピングプログラム用ライブラリー」、補完 HLA タイピングシステムの解析プログラムを「補完 HLA タイピングプログラム」とした。原理を説明する。そして、各工程の詳細な方法は項目 4-5, 6 で説明する。

2-3-1. 補完 HLA タイピングシステムの全体像

図 20 に補完 HLA タイピングシステムの全体像を示した。下記工程を行うことにより、現在の次世代シーケンサーの読み取り配列長の限界 (~400 塩基) を越えた長さの配列を決定できる。

- 1) 最初の PCR 増幅の際、分子バーコード配列を付加することで、各 DNA 分子を固有の分子バーコード配列で標識する (項目 2-3-2-1 を参照)。
- 2) 増幅産物を細断、環状化する (項目 2-3-2-2 を参照)。断端側から塩基配列を決定することにより解析対象領域のいろいろな領域と分子バーコード配列を含むリードデータが得られる (項目 2-3-3 を参照)。
- 3) 分子バーコード配列を用いることにより、分子ごとに配列をグループ化、全体の配列を決定する (項目 2-3-4 を参照)。

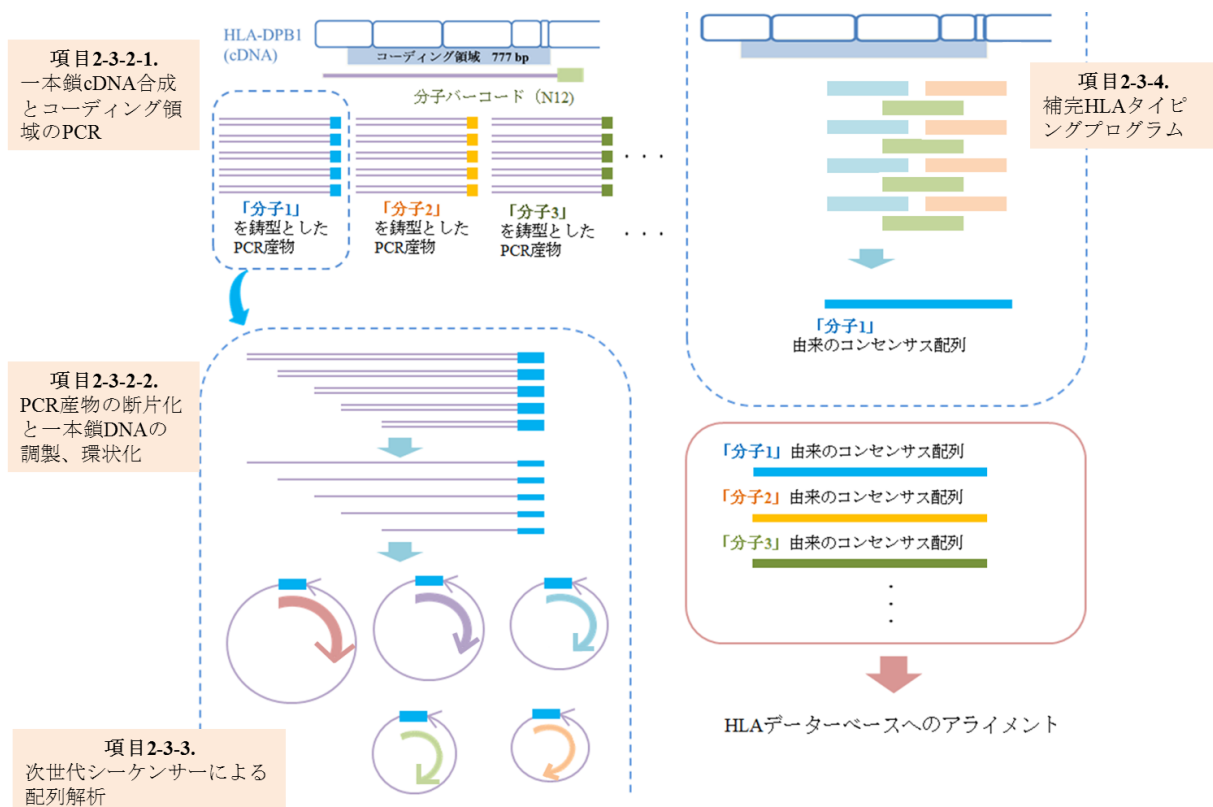


図 20 : 補完 HLA タイピングシステムの全体像

2-3-2. 補完 HLA タイピングプログラム用ライブラリー調製

2-3-2-1. 一本鎖 cDNA 合成とコーディング領域の PCR

図 21 に示したオリゴマー (DPB1-R-adaptor) を用いて逆転写反応を行うことで一本鎖 cDNA を合成した。DPB1-R-adaptor は 5'末端側から、スペーサー配列 (26 塩基)、バーコード配列 (12 塩基)、個人識別配列 (5 塩基)、A プライマーの相補的配列 (30 塩基)、HLA-DPB1 保存領域配列 (23 塩基) から構成され、合計 96 塩基のオリゴマーとなる。スペーサー配列は PCR の際に使用するプライマーのアニーリング配列として使用する。バーコード配列は、各分子を識別し、それぞれでコンセンサス配列を作成する目的で使用する。A プライマーは、Ion Torrent PGM による配列解析で使用するプライマーのひとつである。DPB1-R-adaptor を用いて DPB1 遺伝子座領域の一本鎖 cDNA の合成を行うため、DPB1 遺伝子座特異的にアニーリングできるような配列が DPB1-R-adaptor には必要である。そのため、HLA-DPB1 の HLA DNA データベースに登録されているタイプの非翻訳領域で保存されている配列を DPB1-R-adaptor の 3'末端に配置させた。

合成した一本鎖 cDNA を鋳型とし、DPB1 遺伝子座のコーディング領域外に設計したリン酸化プライマー (P-DPB1-L) とビオチン化プライマー (Biotin-Spacer) を用い、コーディング領域の PCR を行った (図 21)。

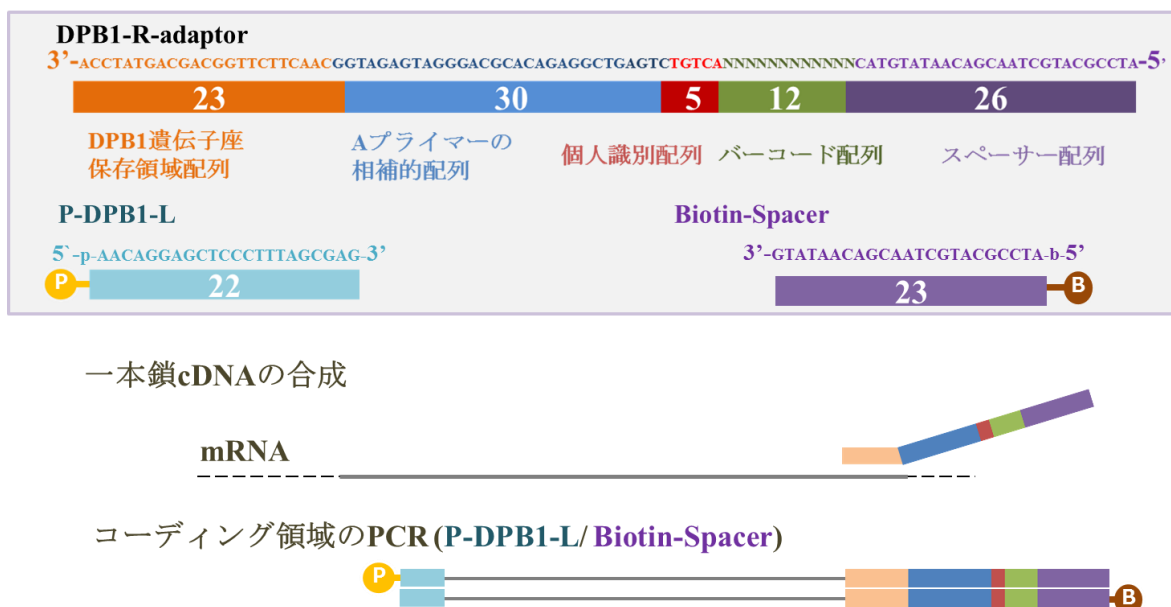


図 21 : 一本鎖 cDNA 合成、コーディング領域の PCR

一本鎖 cDNA 合成の際に使用するオリゴマー (DPB1-R-adaptor) の構成とその塩基配列、そしてコーディング領域の PCR に使用するプライマー (P-DPB1-L、Biotin-Spacer) の構成とその塩基配列を示した。

2-3-2-2. PCR 産物の断片化と一本鎖 DNA の調製、環状化

精製した PCR 産物（項目 2-3-2-1 を参照）の断片化を ds DNA fragmentase（NEB）を用い行った。当酵素による断片化は図 20 のようにランダムに生じる。DPB1-R-adaptor 側の配列が含まれている断片化 PCR 産物を回収するために、ストレプトアビジンの付いた磁気ビーズと磁石を用い、断片化 PCR 産物を回収した。次に、アルカリ変性により一本鎖 DNA を回収し、分子内の環状化を行った（図 20）。

2-3-3. NGS による配列解析

環状化一本鎖 DNA（項目 2-3-2-2 を参照）を鋳型にした PCR（PCR の方法については項目 4-5-7 を参照）を行い、その増幅産物を基に Ion Torrent PGM による配列解析を行なった。Torrent Server から出力された配列データ（BC55 サンプル、リード数 165,205）を用い、補完 HLA タイピングプログラムを開発した。

2-3-4. 補完 HLA タイピングプログラムの開発

まず図 22A のように出力されたリードを分子バーコード配列ごとに分けた。リード数が多い方から 100 種類の分子バーコード配列を持つ配列を使用した。次に、分子バーコード配列ごとでコンセンサス配列を作成する。

図 22B は、HLA タイピングシステムの補完に必要な領域（黄色枠で示す）のリードを DPB1 遺伝子座の参照配列にマッピングさせた時のイメージ図である。まず、解析に使用するリードを参照配列へマッピングし、リードの端のポジションが同じリード集団をリードグループ（PCR による複製配列）とし、リードグループごとでコンセンサス配列を作成した。当処理は、各 PCR 間の増幅バイアスを解消するため、また、リードグループ内での PCR エラーを除去するために行った。

このコンセンサス配列を再度、図 22C のように参照配列へマッピングさせ、出力データから各ポジションの塩基の割合を計算し、各分子バーコード配列におけるコンセンサス配列を作成した。

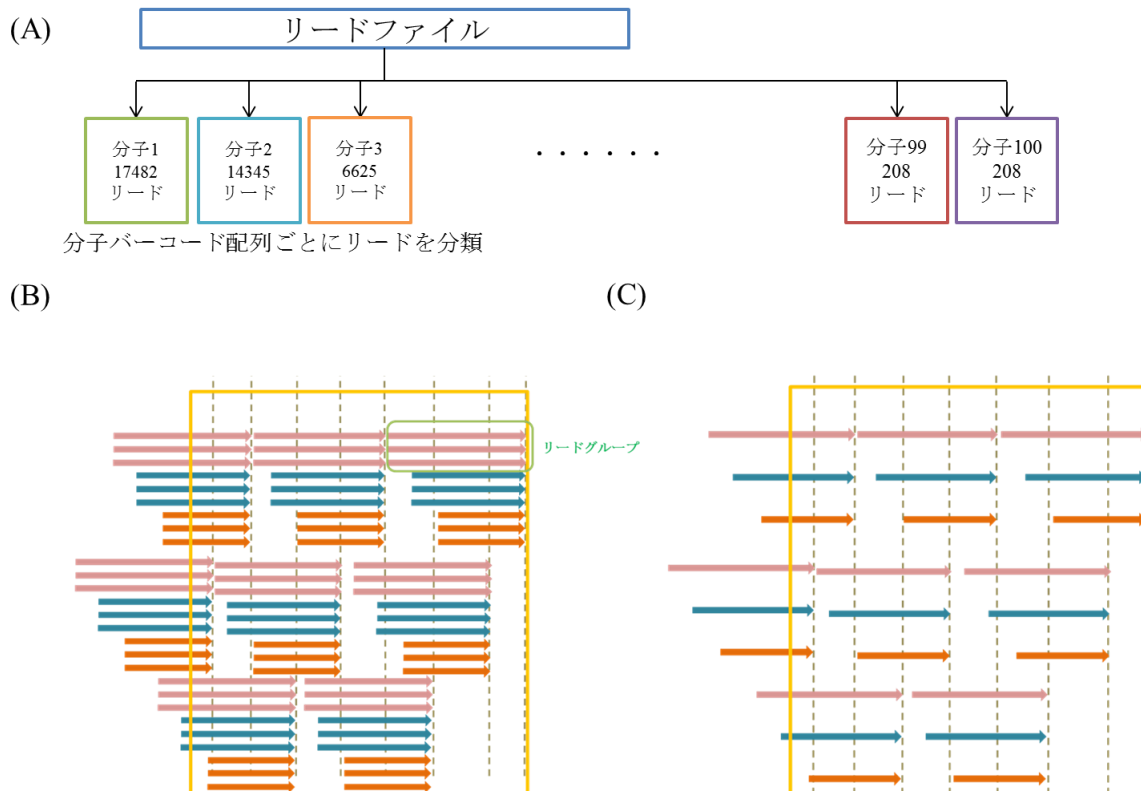


図 22 : コンセンサス配列作成方法の概要

(A) 分子バーコード配列ごとにリードを分類する。

(B) リードの端のポジションが同じリード集団（リードグループ）ごとに色分けし、参照配列へマッピングした時のイメージ

(C) リードグループごとにコンセンサス配列を作成し参照配列へマッピングした時のイメージ
オレンジ色の枠：BC55 の HLA-DPB1 のハプロタイプを決定するために補完すべき領域

2-4. 補完 HLA タイピングシステムの検証

作成した分子バーコード配列ごとコンセンサス配列（項目 2-3-4 を参照）を HLA cDNA データベースへアライメントさせた。項目 2-1-3-3 で説明した Matched Rate (MR) を使用し、HLA タイピングシステムで絞り込んだ 4 タイプ（DPB1*03:01:01、DPB1*05:01:01、DPB1*104:01、DPB1*135:01）（図 19）に完全一致したコンセンサス数を勘定したところ、DPB1*03:01:01、DPB1*05:01:01、DPB1*104:01、DPB1*135:01 それぞれ 20、17、7、2 であった。

結果、勘定したコンセンサス数が高い DPB1*03:01:01、DPB1*05:01:01 を選定し、サンガー法により決定したタイプと一致することを確認した。

以上のように、HLA タイピングシステムの結果を補完することでサンプル BC55 の HLA-DPB1 のハプロタイプを決定できた。

3. 考察

3-1. HLA タイピングシステムについて

3-1-1. HLA タイピングシステムの特徴

開発した HLA タイピングシステムは RNA から 11 遺伝子座 (HLA-A, -B, -C, -DRB1, -DRB3/4/5, -DQB1, -DQA1, -DPB1) のライブラリー調製を行った後に、Ion Torrent PGM による配列解析を行い、その出力データを基に HLA タイピングプログラムによる解析により 6 桁レベル (コーディング領域の塩基配列レベル) で HLA タイプを自動的に決定するものである。

ゲノム DNA からタイピングを行っている論文は多くあるが、RNA から HLA タイピングを行っている論文は少ない^{50), 64)}。ゲノム DNA から個々のエクソンを増幅する方法を用いライブラリー作成を行った場合、エクソン間の cis、trans 情報を得ることができないため、決定できないハプロタイプが存在する⁵⁾。RNA を用いた HLA タイピングでは、各エクソンをまたいだ塩基配列データを用いるため、エクソン間での cis、trans の多型情報を得ることができる利点がある。当手法では、少なくとも 2 エクソンを含む PCR 産物よりライブラリーを作成している。また、6 桁レベルまでの HLA タイピングを行う場合、RNA を用いた方がゲノム DNA よりも解析領域が小さく、ライブラリー調製が簡便である。

我々は HLA タイピングプログラム用ライブラリー調製時に使用するプライマーを 11 遺伝子座に対して 27 セット設計した。IMGT/HLA データベースを調べ、アレル頻度が高いタイプのうち、ほとんどの HLA タイプ (98.9 %以上) において保存された配列領域に設計したため、汎用性の高いプライマーセットである。また、確実に解析領域の配列情報を得るため、そして増幅領域間の cis、trans の多型情報を得るため、それぞれの増幅領域が重複するように設計した。また、27 プライマーセットでのアニーリング温度をすべて同一の温度に設定したため、PCR の実験操作が簡便である。

開発した HLA タイピングプログラムは「配列ファイル作成プログラム」「ハプロタイプ候補配列作成プログラム」「HLA データベース検索プログラム」「HLA タイプ決定プログラム」の 4 段階で構成されており、この 4 つのプログラムを順に実行させることで自動的に個人毎の HLA タイプレポートを出力させることができる。

「配列ファイル作成プログラム」では、両方向のリード (フォワードリード、リバースリード) ごとの配列ファイルを作成することで、各解析領域全体の塩基情報を確実に取得できる手法をとった。当プログラムは、各プライマーの設計位置情報を基に SAM 形式ファイルより各解析領域の配列ファイルを作成するので、他の NGS 機種で

も解析可能であり、汎用性が高い。

「ハプロタイプ候補配列作成プログラム」は、これまでの NGS による HLA タイピングの報告において問題とされていた「不正確なアレルの区別化」や「NGS による塩基読み取りエラー」（項目 1-5-5 を参照）に対処できる手法である。

「HLA データベース検索プログラム」では、BLAT (<http://genome.ucsc.edu/>) の出力ファイルから、完全に配列が一致した HLA cDNA データベース中のタイプを選定することができる数値 (MR) を自動的に計算できる手法を開発した。HLA cDNA データベースの配列の長さに対して、アライメントする配列が長い場合は MR の計算は必要ないが、当手法のように HLA cDNA データベースの配列の長さに対して、アライメントする配列が短い場合は完全に配列が一致した HLA タイプを選定するには MR を計算する必要がある、当手法において特徴的な処理のひとつといえる。

HLA cDNA データベースに登録されている多くのアレルは完全な参照配列ではなく、長い HLA タイプもあれば、その半分以下といった長さの HLA タイプも登録されている。このような HLA cDNA データベースの現状にも「HLA タイプ決定プログラム」は対応している。

3-1-2. HLA タイピングシステムと他研究との比較

3-1-2-1. 他研究の HLA タイプ判定プログラムとの比較

Gabriel ら⁴²⁾、Holcomb ら⁴³⁾ は、454 GS FLX を用いた HLA タイプ判定プログラムを作成している。454 GS FLX は配列決定できる鎖長が長いという長所を持つが、塩基の読み取りエラーが多い、解析時間が長い、データ量が少ない、解析コストがかかるといった短所をもつ (表 1)。Gabriel ら⁴²⁾、Holcomb ら⁴³⁾ のプログラムは、長い鎖長のリードデータが出力可能な 454 GS FLX を活かした HLA タイプ判定プログラムとなっており、複雑な前処理を行うことなく IMGT/HLA データベースとの類似性比較を行い、HLA タイプを判定している。

一方、短い鎖長のリードデータしか出力できない Illumina システムや Ion Torrent PGM を用いた HLA タイピングでは、IMGT/HLA データベースとの類似性比較を行う前後で、正確な HLA タイピングを行うための複雑な処理が必要となる。Illumina システムは、解析時間は他機種より長い、データ量が多いため解析コストが安く、読み取りエラーが他機種に比べて少ないという長所をもつ (表 1)。一方、Ion Torrent PGM は、他機種よりも解析コストが安く、さらに解析時間も短いため、迅速さが求められる HLA 検査には適している機種といえる。しかし、Ion Torrent PGM には読み取りエラーが多いという短所があるため、Ion Torrent PGM からのリードデータから HLA タイピングを行うためには短いリードデータ、塩基の読み取りエラーを克服しなければならない。

Hosomichi ら⁴⁴⁾、Cao ら⁴⁵⁾の方法では、Illumina システムからのリードデータを使用し、アレルごとにつなぎ合わせて解析領域全体のコンセンサス配列を作成する（配列アセンブリ）といったように IMGT/HLA データベースとの類似性比較を行う前の処理に重点を置いている。また、Shiina ら^{46), 47)}は、Ion Torrent PGM を用いた HLA タイピング手法を開発している。彼らが開発した HLA タイプ判定プログラム SeaBass では、IMGT/HLA データベースとの類似性比較を行った後の処理に重点をおき、最終的に、候補アレルに対するマッピングデータから被覆率を参考に HLA タイプを判定している。

我々の HLA タイピングプログラムは、同じ領域にマッピングされたリード集団ごとでハプロタイプの候補となる配列を構成する「ハプロタイプ候補配列作成プログラム」を作成し、IMGT/HLA データベースとの類似性比較を行った後、「HLA タイプ決定プログラム」で各領域のデータベースへの検索結果を統合させるといった方法である。このように、我々の「HLA タイピングプログラム」では、IMGT/HLA データベースとの類似性比較の前後両方に重点を置いた手法となっており、解析領域全体の配列アセンブリを行う事とほぼ同等の意味合いを表すような処理方法といえる。また、我々の手法では、最終的な HLA タイプ判定を、被覆率を参考に決定した Shiina ら^{46), 47)}とは異なり、ハプロタイプスコア（各 HLA タイプ各領域の完全一致リード数）を参考にして行っている。ハプロタイプスコアは、各候補配列の信頼性を表すだけでなく、スコアが同一の場合同じ配列を持つという特徴をもち、その特徴を活かした HLA タイプ判定プログラムとなっている。

3-1-2-2. 他研究との HLA タイピング精度比較

6 桁レベルの NGS を用いた HLA タイピングを行っている報告 (Ozaki ら⁵¹⁾、Nelson ら⁵²⁾、Lank ら⁵⁰⁾) との HLA タイピングの精度比較を行った。使用した NGS の機種、サンプル形態（ゲノム DNA もしくは RNA）、サンプル数、解析した遺伝子座数、と共に HLA タイピング正解率を表 9 に示した。各報告の HLA タイピング正解率は Ozaki ら⁵¹⁾では論文中の「Table S2」、Nelson ら⁵²⁾では論文中の「Table S2」、そして Lank ら⁵⁰⁾では論文中の「Table S6, 7」を参考にして記載した。

なお、それぞれの報告におけるデータセットは異なるサンプルである。

(A)	NGS機種	サンプル形態	サンプル数	遺伝子座数		Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
HLAタイピングシステム	Ion Torrent PGM	RNA	30	11	Program (alleles)	516	58	57	57	59	51	57	59	58	60
					Manually (alleles)	532	60	60	60	60	52	60	60	60	60
					Accuracy (%)	97.0	96.7	95.0	95.0	98.3	98.1	95.0	98.3	96.7	100
Ozaki ⁵¹⁾	Ion Torrent PGM	ゲノム DNA	46	9	Program (alleles)	606	92	92	92	77	71	92	/	90	/
					Manually (alleles)	623	92	92	92	92	71	92	/	92	/
					Accuracy (%)	97.3	100	100	100	83.7	100	100	/	97.8	/
(B)	NGS機種	サンプル形態	サンプル数	遺伝子座数		Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
HLAタイピングシステム	Ion Torrent PGM	RNA	30	11	Program (haplotype)	254	28	27	27	29	27	28	29	29	30
					Manually (haplotype)	268	30	30	30	30	28	30	30	30	30
					Accuracy (%)	94.8	93.3	90.0	90.0	96.7	96.4	93.3	96.7	96.7	100
Nelson ⁵²⁾	MiSeq	ゲノム DNA	336	11	Program (haplotype)	2869	317	326	312	334	306	330	335	274	335
					Manually (haplotype)	3006	335	335	335	335	325	334	336	335	336
					Accuracy (%)	95.4	94.6	97.3	93.1	99.7	94.2	98.8	99.7	81.8	99.7
(C)	NGS機種	サンプル形態	サンプル数	遺伝子座数		Total	A	B	C	DRB1	DRB3/4/5	DQB1	DQA1	DPB1	DPA1
HLAタイピングシステム	Ion Torrent PGM	RNA	30	11	Program (alleles)	516	58	57	57	59	51	57	59	58	60
					Manually (alleles)	532	60	60	60	60	52	60	60	60	60
					Accuracy (%)	97.0	96.7	95.0	95.0	98.3	98.1	95.0	98.3	96.7	100
Lank ⁵⁰⁾	GS Junior (Roche)	RNA	A,B,C: 85 DRB: 78	7	Program (alleles)	720	160	157	151	150	102	/	/	/	/
					Manually (alleles)	805	170	170	170	156	139	/	/	/	/
					Accuracy (%)	89.4	94.1	92.4	88.8	96.2	73.4	/	/	/	/

Program : HLAタイピングプログラムにより決定したアレル数またはハプロタイプ数

Manually : SSOP法、SSP法、SBT法など従来の方法により決定したアレル数またはハプロタイプ数

Accuracy : サンガー法により決定したアレルまたはハプロタイプのうちHLAタイピングプログラムにより決定できた割合

表 9 : 他研究との HLA タイピング精度比較

まず、Ozaki ら ⁵¹⁾ の報告との比較である (表 9A)。当システムの HLA タイピング正解率 (97.0 %) と比べ、Ozaki ら ⁵¹⁾ の報告では 97.3 % であり、ほぼ同等であるが、当システムで行った HLA-DQA1, -DPA1 は解析対象としていない。

次に Nelson ら ⁵²⁾ の報告 (表 9B) と比較した。解析対象としている遺伝子座数は当システムと同一である (11 遺伝子座)。この報告では決定できたアレル数ではなくハプロタイプ数で表示しているため、当システムの結果もハプロタイプ数で表示させて比較した。当システムの HLA タイピング正解率 (94.8 %) と比べ、Nelson ら ⁵²⁾ の報告では 95.4 % であり、ほぼ同等であった。

次に Lank ら ⁵⁰⁾ の報告 (表 9C) と比較した。当報告では当システムと同じ RNA サンプルから 6 桁レベルの HLA タイピングを行っている。HLA タイピング正解率 (89.4 %) も解析対象としている遺伝子座数 (7 遺伝子座) も当システムの方が上回っており、我々の HLA タイピングシステムの方が優れている。

我々が開発した HLA タイピングシステムは他機種より解析時間が短く、解析コストが低い Ion Torrent PGM を用いている。HLA タイピングを行うにあたり、短所といえる短い鎖長のリードデータと NGS による塩基の読み取りエラーを HLA タイピングプログラムで克服し、他の研究と比較しても、11 遺伝子座すべてにおいてバランス良く高い精度で HLA タイピングが行えているといえる。

3-1-3. HLA タイピングシステムの改善点

HLA タイピングシステムの検証（項目 2-2 を参照）において NGS による読み取りエラーが原因で決定できなかったアレル、増幅バイアスが影響し決定できなかったアレル、当システムでは決定不可能なハプロタイプ（補完 HLA タイピングシステムでは決定できたハプロタイプ）が存在した。NGS を用いた HLA タイピングで問題となっている読み取りエラー、増幅バイアス（項目 1-5-5 を参照）の影響が少なからず生じた結果となった。当システムにおいてハプロタイプを決定できる条件は、全解析領域（HLA クラス I：1F~4R、HLA クラス II：1F~3R）において、増幅バイアスがなく、読み取りエラーが少ない必要がある。また、新規 HLA タイプの同定により年々 IMGT/HLA データベースの登録数が増える中、設計したプライマー配列領域に、プライマーのアニーリングが正確にできず増幅が行えない多型をもつタイプが同定されることも考えられる。

これらを踏まえ、当 HLA タイピングシステムの改善点について考察したい。当システムの解析プログラムの中で最も重要な工程は、マルチプルアライメントツールを用い 2 アレルの HLA 配列の候補を作成する「ハプロタイプ候補配列作成プログラム」と考える。マルチプルアライメントツールである MUSCLE を使用し、配列を整列化させるのだが、正確に整列化が行われないと正しい HLA タイピングができない。実際に、読み取りエラーが多い領域から作成した候補配列では正確な HLA タイプを導き出せなかった。このような場合、マルチプルアライメント結果の適当な修正過程の導入、もしくは当解析にさらに合うようなマルチプルアライメントツールの選択など、改善の余地はある。また、当システムの配列解析においては使用していなかったが、読み取りエラーを減少させることができる新しいプロトコール（Ion PGM™ Hi-Q™ Sequencing kit）が Thermo Fisher Scientific 社より市販されている。今後の試薬キット等のバージョンアップにより読み取りエラーはさらに改善されていくことが期待できる。

他に、HLA タイピング検査に要する時間をさらに短縮し、システムの簡略化を図ることもできる。ライブラリー調製段階において、27 プライマーセットによる PCR をマルチプレックスで行うことができれば作業時間の短縮につながる。これは、近年多く開発されている NGS 用ライブラリーの全自動作成装置にも作業工程を組み込みやすくなり、自動化にもつながる。当 HLA タイピングプログラムでは、Intel® Xeno® CPU E5420 @ 2.50 GHz のスペックをもつコンピューターを使用し、1 サンプルずつ解析を行った。解析時間は 1 サンプルあたり約 40 分であった。さらに高スペックのコンピューターの使用、そして複数サンプルの並列解析を行えば、解析時間の短縮につながる。

また、当システムではすべてのプライマーセットでアニーリング温度を同一の 50 °C に設定しているが、増幅バイアスがさらに少なくなるようなアニーリング温度がプライマーセットごとに存在する可能性がある。上記で記したマルチプレックスで

の PCR は行なうことはできないが、プライマーセットごとでアニーリング温度を設定することで増幅バイアスの改善が見込める。

3-2. 補完 HLA タイピングシステムについて

分子バーコード技術、PCR 産物の断片化環状化による解析領域全体の配列を作成する方法は Hong ら⁶⁵⁾ (B 型肝炎ウイルスゲノム)、Stapleton ら⁶⁶⁾ (HIV 外被糖蛋白質遺伝子) によって報告されている。当補完システムでは、この技術を HLA タイピングに応用している。

補完 HLA タイピングシステムは、400 塩基以上のコンセンサス配列を作成できるだけでなく、分子バーコード配列を利用して分子ごとの配列を作成できるので、HLA タイピングシステムの問題として残った増幅バイアスの影響を軽減することが期待できる。また、補完 HLA タイピングシステムのライブラリー調製は、工程数が多く煩雑であるため、調製方法の改良を行うことでさらに有用な HLA タイピング手法につながることを期待できる。

4. 材料と方法

4-1. 材料

健康な日本人男性 3 人から血液サンプルを採取し、Ficoll-Paque PLUS (GE Healthcare) によりヒト末梢血単核球の分離をプロトコールに従って行った。TRIzol RNA Isolation Reagent (Thermo Fisher Scientific) を用い、ヒト末梢血単核球より全 RNA の抽出をプロトコールに従って行った。また、大阪府立病院機構大阪府立成人病センターにより外科切除が行われ、マイナス 80°C で凍結保存されていた乳癌組織 19 症例、大腸癌組織 20 症例を用い、全 RNA の抽出を行った。凍結ミクロトーム CM1900 (Leica) を用い、20 μm の厚さで凍結組織の切り出しを行った後、TRIzol RNA Isolation Reagent を用い、切り出した約 10 枚の組織切片からの全 RNA の抽出をプロトコールに従って行った。

4-2. HLA タイピングプログラム用ライブラリー調製手法

4-2-1. 一本鎖 cDNA 合成

100 ng の全 RNA (項目 4-1 を参照) を用いて一本鎖 cDNA を合成した。5 μl の全 RNA、3 μl の Nuclease-Free water、4 μl の 2.5 mM dNTP (Takara bio)、1 μl の 10 μM ランダムヘキサマーを混合し、65 °C で 5 分間インキュベートした後、2 分間氷上で静置した。続いて、4 μl の 5 x 1st strand buffer (Invitrogen)、1 μl の 0.1 M DTT (Invitrogen)、1 μl の RNase Out ribonuclease Inhibitor (Invitrogen)、1 μl の Super Script III Reverse Transcriptase (Invitrogen) を混合し、13 μl のランダムヘキサマーとアニーリングした全 RNA に加え、25 °C で 5 分間、50 °C で 60 分間、70°C で 15 分間インキュベートした後、氷上で冷却した。以上操作を 1 サンプルあたり 2 チューブで実施し計 40 μl の一本鎖 cDNA を合成した。

4-2-2. HLA プライマーの設計

各遺伝子座に対して HLA cDNA データベースに登録されている各 HLA タイプの配列をマルチプルアライメントツール MUSCLE (<http://www.drive5.com/muscle>) を用いてマルチプルアライメントを行い、各 HLA タイプで保存されている配列領域を探し、その領域に Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>) を用いプライマー設計を行った (表 2)。

4-2-3. HLA プライマーを用いた PCR

1 μ l の一本鎖 cDNA 反応液 (項目 4-2-1 を参照)、4 μ l の 5 x Q5 buffer (NEB)、1.6 μ l の 2.5 mM dNTP、2 μ l の設計したそれぞれ 5 μ M であるプライマーペア (項目 4-2-2 を参照)、11.2 μ l の Nuclease-Free water、0.2 μ l の Q5 Hot Start High-Fidelity DNA Polymerase (NEB) を混合し、20 μ l の反応系で PCR を行った。Veriti[®] サーマルサイクラー (Thermo Fisher Scientific) を使い、98 $^{\circ}$ C で 30 秒間インキュベートした後、98 $^{\circ}$ C で 10 秒間、50 $^{\circ}$ C で 10 秒間、72 $^{\circ}$ C で 30 秒間というサイクルを 30 回行った後、72 $^{\circ}$ C で 2 分間インキュベートした。

そして、27 種類の増幅産物の濃度を Quanti-iT[™] dsDNA Assay Kit, Broad range (Invitrogen) を用いてプロトコールに従い測定した。次に、測定した濃度を基に 27 種類の増幅産物を等量混合し、AMPure XP kit (Beckman Coulter) を用いて、プロトコールに従い精製を行った。そして、Quanti-iT[™] dsDNA Assay Kit, または Qubit dsDNA HS Assay Kit (Invitrogen) を用いて、精製した増幅産物混合液の濃度をプロトコールに従い測定した。

4-2-4. NGS テンプレート調製

精製した増幅産物混合液のうち 20 ng もしくは 50 ng を使用し、Ion Xpress Plus Fragment Library Kit (Thermo Fisher Scientific) と Ion Xpress barcode Adaptors 1-16 Kit (Thermo Fisher Scientific) によりシークエンス解析用のライブラリー調製をプロトコールに従い、行った。調製したライブラリーの濃度を Agilent High Sensitivity DNA Assay kit (Agilent Technologies) を用いて、プロトコールに従い測定した。

4-2-5. NGS による配列解析

調製した各サンプルのライブラリーを等量ずつ混合した後、そのライブラリーの混合液 2.5 μ l を使用し、Ion OneTouch[™] 2 (Thermo Fisher Scientific) と Ion OneTouch[™] ES (Thermo Fisher Scientific) を使いプロトコールに従いシークエンステンプレートを調整した。Ion PGM[™] Sequencing 400 Kit (Thermo Fisher Scientific) を用いて 318v2 chip 解析用のプロトコールに従い Ion Torrent PGM による配列解析を行った。そして、Torrent Server (Ion PGM[™] Sequencer による解析用サーバー) 上でデータ品質のフィルタリングを行い、BAM 形式ファイルとしてデータを出力した。

4-3. HLA タイピングプログラムによる HLA タイプ判定方法

4-3-1. 領域ごとの配列ファイルの作成

マッピングの結果情報が格納されている BAM 形式ファイルをマッピングファイル処理ツールである SAMtools⁶⁷⁾ を用いて SAM 形式ファイルに変換後、各遺伝子座の配列ファイルを作成した。設計領域の 55% もしくは 60% の長さ以上の配列のうち、プライマー設計領域の塩基配列を除去した配列を配列ファイルとした。なお、リード数が 10,000 を超える領域の場合、ランダムに 10,000 リードを選択した。

4-3-2. ハプロタイプ候補配列ファイルの作成

4-3-2-1. MUSCLE によるマルチプルアライメント

マルチプルアライメントツールである MUSCLE により、マルチプルアライメントを各配列ファイル（項目 4-3-1 を参照）に対して行い、配列の整列化を行う。解析時間を短縮させるため、アライメントの反復操作を最小に設定した。図 24A は配列 1 と配列 2 由来の 100 リードがマルチプルアライメントされた結果を示している。

4-3-2-2. 各ポジションの塩基文字情報の取得

当項目では、アレル間において違いがある塩基のポジションを「ヘテロポジション」、アレル間において同じ塩基であるポジションを「ホモポジション」とする。

まず、MUSCLE により整列化（項目 4-3-2-1 を参照）させた各ポジションの塩基文字（A、T、C、G、ハイフン）の割合を図 24B のように計算した。各ポジションで計算された塩基文字の割合が 1% 以下であれば、NGS による読み取りエラーとみなし、その塩基文字を除外した。図 24B の例では、ポジション 3 の T、ポジション 9 のハイフン、ポジション 17 の C の存在割合が 1% であるので、それら塩基文字を除く。

1% 以下である塩基文字を除去した後に、各ポジションの残りの塩基文字数を調べ、塩基文字が 2 種類以上であれば「ヘテロポジション」、1 種類であれば「ホモポジション」とみなした。図 24B の例ではポジション 7 が 3 種類、ポジション 14 が 2 種類、ポジション 23 が 2 種類となるため、これらポジションは「ヘテロポジション」、他のポジションは「ホモポジション」となる。

4-3-2-3. ハプロタイプパターン候補配列とハプロタイプスコアの取得

取得した各ポジションの塩基文字情報を元に、考えられる組み合わせのハプロタイプパターン（候補 1~12）を図 24C のように作成した。

「ホモポジション」では、取得した塩基文字をそのポジションに配置させる。次に、ヘテロポジションでの塩基文字の種類数をもとに、考えうるすべてのハプロタイプの組み合わせを羅列する。図 24C の場合、12 パターン (=3×2×2) のハプロタイプとなる。次に、各ハプロタイプパターンを有するリード数 (ハプロタイプスコア) をマルチプルアライメントデータの「ヘテロポジション」を調べることで勘定した。図 24C の場合、ハプロタイプスコアは候補 1 と候補 8 でそれぞれ 50、48 となる。候補 12 を除いた残りの候補配列ではハプロタイプスコアは 0 となる。図 24A のポジション 7 に塩基置換エラーが生じているリードが 2 つ存在するため、候補 12 のハプロタイプスコアは 2 となる。次に、項目 2-2 において設定したフィルタリングパラメーター 1 を用いてハプロタイプスコアが高いハプロタイプ (候補 1 と候補 8) を HLA データベース検索プログラム (項目 2-1-3-3 を参照) で使用する。

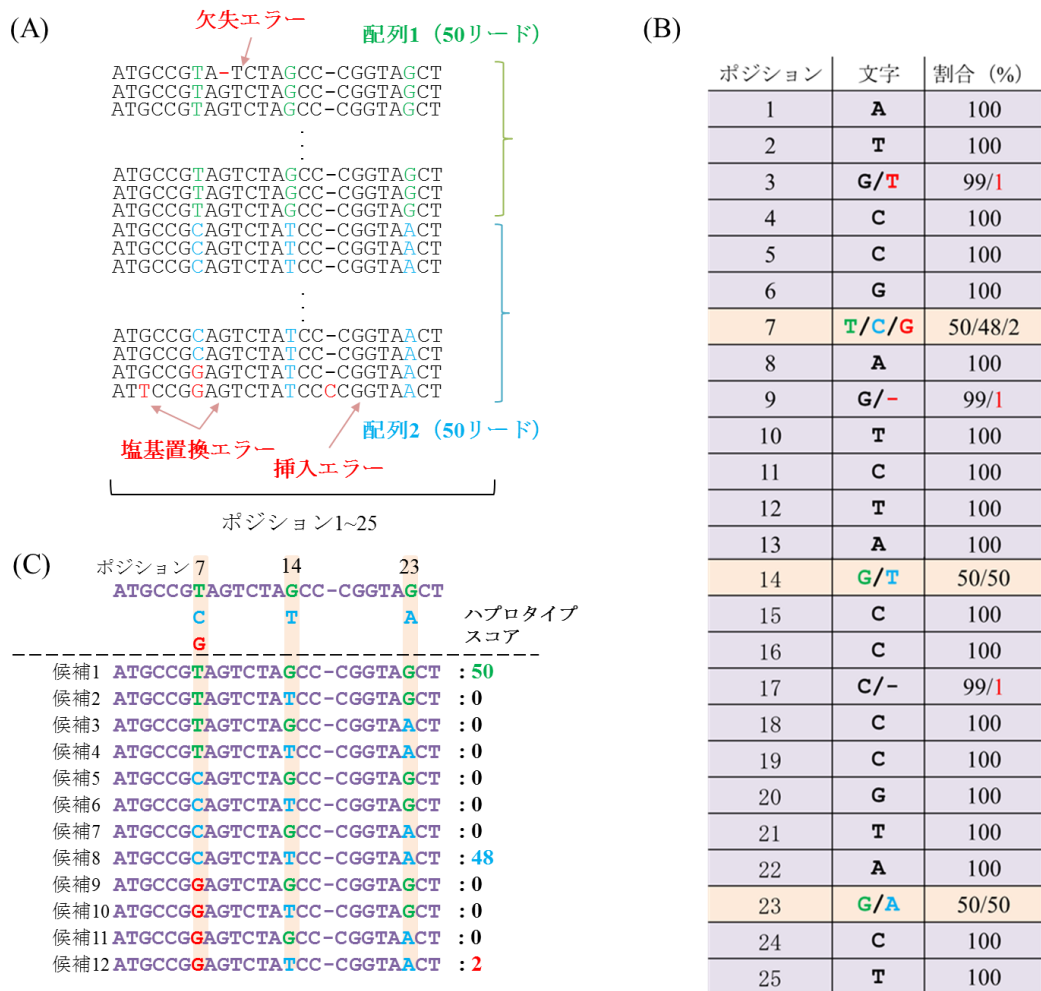


図 24 : ハプロタイプ候補配列ファイルの作成

- (A) 2 アレル (配列 1, 2) 由来の 100 リード (25bp) をマルチプルアライメントさせた例を示す。
(B) 各ポジションにおける塩基文字の種類と割合を示す。オレンジで示したポジションは塩基文字の存在割合が 1% 以下である文字を除去しても、2 つ以上の文字が残存するポジション。
(C) ハプロタイプ候補配列の 12 パターンの配列とそれぞれのハプロタイプスコア。

4-3-3. HLA データベース検索ファイルの作成

当項目では、項目 4-3-2 において作成した各領域のハプロタイプ候補配列が完全に一致する HLA タイプを検索する方法を以下に示す。

作成した各領域のハプロタイプ候補配列をアライメントツールである BLAT (BLAST-like alignment tool) により、HLA cDNA データベース (version 3.26.0、2016 年 10 月更新) に対してアライメントを行った。そのうち、不一致塩基数、リードの挿入エラー数が 0 であるアライメントデータのみを選択した。HLA cDNA データベースとハプロタイプ候補配列が重複している領域において完全に一致しているタイプを正確に選定するために MR を計算した。

MR を計算する際の分母として使用する重複領域の長さの計算方法を説明する。重複領域の長さを計算するための情報 (①~⑨) を、① ハプロタイプ候補配列の参照配列上でのスタートポジション (Mapped Start Position of Query ; MSPQ)、② ハプロタイプ候補配列の参照配列上でのエンドポジション (Mapped End Position of Query ; MEPQ)、③ ハプロタイプ候補配列の長さ (Length of query; Lq)、④ ハプロタイプ候補配列のアライメントスタートポジション (Query Start ; QS)、⑤ ハプロタイプ候補配列の欠失エラー塩基数 (Tgap)、⑥ HLA cDNA データベースの参照配列上でのスタートポジション (Mapped Start Position of Target ; MSPT)、⑦ HLA cDNA データベースの参照配列上でのエンドポジション (Mapped End Position of Target ; MEPT)、⑧ HLA cDNA データベースの長さ (Length of target; Lt)、⑨ HLA cDNA データベースの被アライメントスタートポジション (Target Start ; TS) とした (図 25A)。①、③、④、⑤、⑧、⑨は BLAT 出力ファイルの中に存在するデータであるが、②、⑥、⑦に関しては BLAT 出力ファイルに含まれないため、それぞれ以下のように計算した。

ハプロタイプ候補配列の参照配列上でのエンドポジション (②、MEPQ) :

ハプロタイプ候補配列が参照配列にマッピングさせた際のエンドポジションを MSPQ (①)、Lq (③)、Tgap (⑤) のデータを用い、以下のように計算した。

$$\text{MEPQ (②)} = \text{MSPQ (①)} + \text{Lq (③)} + \text{Tgap (⑤)} - 1$$

HLA cDNA データベースの参照配列上でのスタートポジション (⑥、MSPT) :

HLA cDNA データベースとハプロタイプ候補配列の重複領域の長さを算出するためには、HLA cDNA データベースとハプロタイプ候補配列の位置関係を一致させる必要がある。仮に HLA cDNA データベースを参照配列にマッピングさせた場合のスタートポジション (⑥、MSPT) を取得できれば、HLA cDNA データベースとハプロタイプ候補配列の位置関係を一致させることができる。MSPT (⑥) は、MSPQ (①)、QS (④)、TS (⑨) のデータを用い、以下のように計算した。

$$\text{MSPT (⑥)} = \text{MSPQ (①)} + \text{QS (④)} - \text{TS (⑨)}$$

「MSPQ (①) + QS (④)」で、ハプロタイプ候補配列における HLA cDNA データベースへのアライメントがスタートする参照配列上でのポジションが求められる。TS (⑨) は HLA cDNA データベースの配列上において、ハプロタイプ候補配列のアライメントがスタートしたポジションを表しているため、「MSPQ (①) + QS (④)」で求めたポジションから TS (⑨) を引けば、HLA cDNA データベースの参照配列上におけるスタートポジションを算出することができる。

HLA cDNA データベースの参照配列上でのエンドポジション (⑦、MEPT) :

MSPT (⑥)、Lt (⑧) のデータを用い、以下のように計算した。

$$\underline{\text{MEPT (⑦)} = \text{MSPT (⑥)} + \text{Lt (⑧)} - 1}$$

以上の計算方法により得られたデータ (①~⑨) を用い、HLA cDNA データベースとハプロタイプ候補配列の重複領域を計算する。ハプロタイプ候補配列は各領域 (1F~4R) において、HLA cDNA データベースとの位置関係は様々であり重複領域の計算方法も異なる。したがって、HLA cDNA データベースとハプロタイプ候補配列の位置関係の場合分けを以下のように行い、重複領域を計算した (図 25B)。

I : MSPQ (①) \leq MSPT (⑥) かつ MEPT (⑦) \leq MEPQ (②)

(ハプロタイプ候補配列領域に HLA cDNA データベース登録配列領域が完全に含まれる場合)

II : MSPQ (①) \leq MSPT (⑥) かつ MEPT (⑦) $>$ MEPQ (②)

(HLA cDNA データベース登録配列領域にハプロタイプ候補配列領域の右側だけが含まれる場合)

III : MSPQ (①) $>$ MSPT (⑥) かつ MEPT (⑦) \leq MEPQ (②)

(HLA cDNA データベース登録配列領域にハプロタイプ候補配列領域の左側だけが含まれる場合)

IV : MSPQ (①) $>$ MSPT (⑥) かつ MEPT (⑦) $>$ MEPQ (②)

(HLA cDNA データベース登録配列領域にハプロタイプ候補配列領域が完全に含まれる場合)

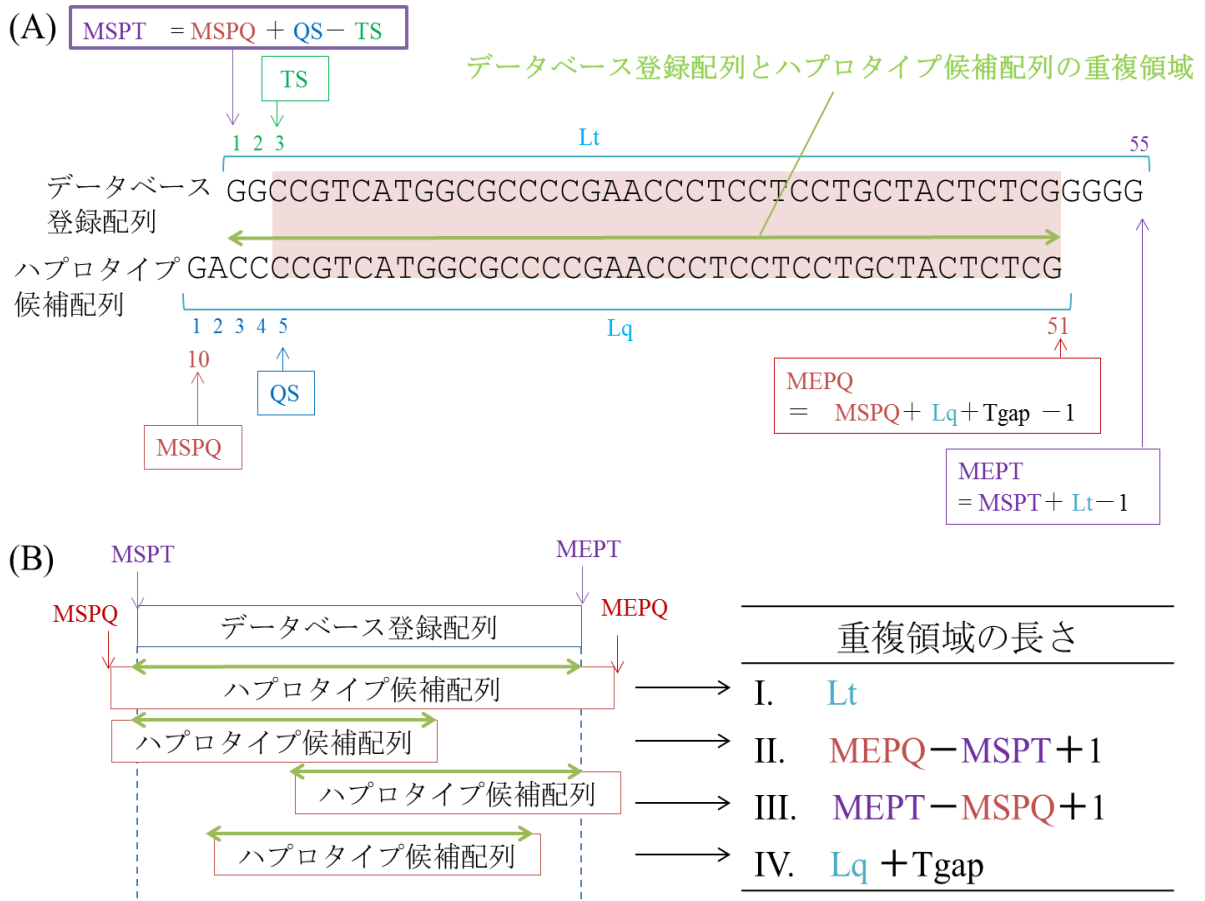


図 25 : HLA cDNA データベース登録配列とハプロタイプ候補配列の重複領域の長さの計算方法

(A) HLA cDNA データベース登録配列とハプロタイプ候補配列がアライメントしたときのイメージと重複領域の長さを求めるための情報を示す。

重複領域の長さを求めるための情報。

MSPQ : ハプロタイプ候補配列の参照配列上でのスタートポジション。

MEPQ : ハプロタイプ候補配列の参照配列上でのエンドポジション。

Lq : ハプロタイプ候補配列の長さ。

QS : ハプロタイプ候補配列のアライメントスタートポジション。

Tgap : ハプロタイプ候補配列の欠失エラー塩基数。

MSPT : HLA cDNA データベースの参照配列上でのスタートポジション。

MEPT : HLA cDNA データベースの参照配列上でのエンドポジション。

Lt : HLA cDNA データベースの長さ。

TS : HLA cDNA データベースの被アライメントスタートポジション。

(B) 重複領域の長さの計算方法の場合分け

I : $MSPQ \leq MSPT$ かつ $MEPT \leq MEPQ$

II : $MSPQ \leq MSPT$ かつ $MEPT > MEPQ$

III : $MSPQ > MSPT$ かつ $MEPT \leq MEPQ$

IV : $MSPQ > MSPT$ かつ $MEPT > MEPQ$

4-3-4. HLA タイプ決定プログラムの解析フローチャート

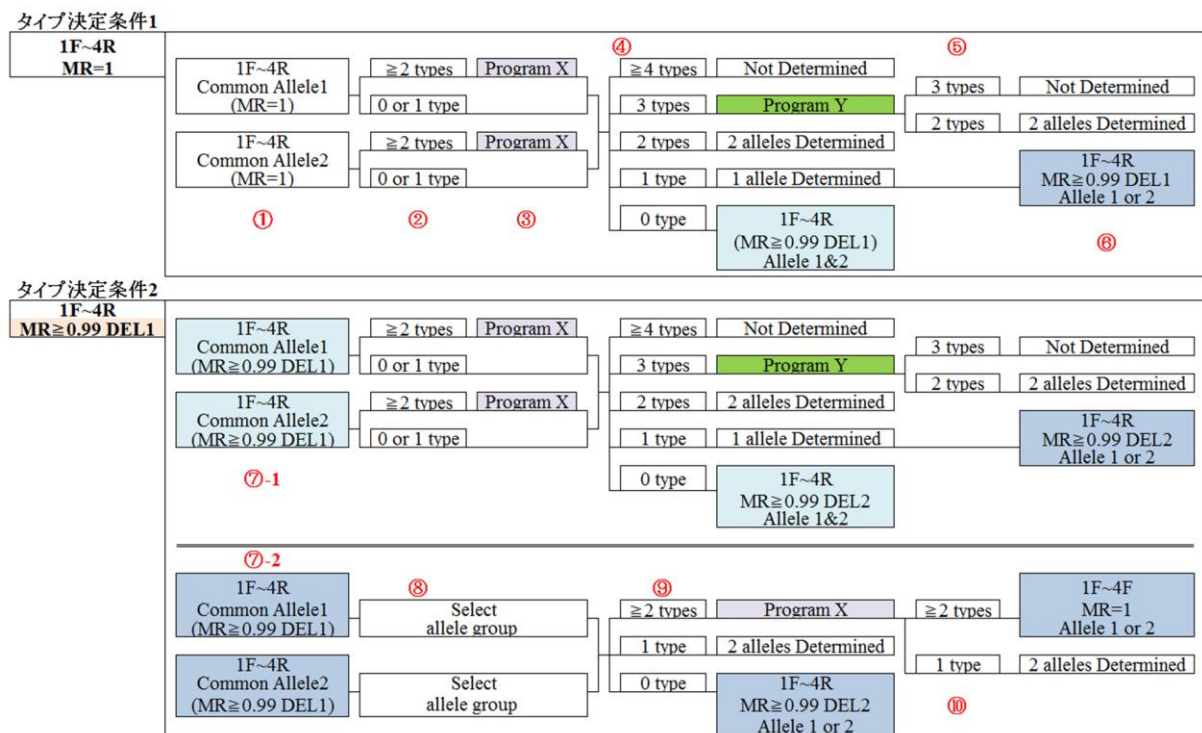
当項目では、HLA タイプ決定プログラムの具体的な解析の流れを図 26 の解析フローチャートを用いて説明する。図 26 は「NGS の読み取りエラーのため正解となる配列を絞り込めない場合（項目 2-1-3-4-1 を参照、「Program X」とする）」、「NGS の読み取りエラーの中で特に欠失エラーのため MR ≠ 1 になる場合（項目 2-1-3-4-2 を参照）」、「別コンティグに多型がある場合（項目 2-1-3-4-3 を参照、「Program Y」とする）」、「HLA cDNA データベースに登録された HLA 配列が不完全な場合（項目 2-1-3-4-4 を参照）」を組み込んだ HLA タイプ決定プログラムの解析フローチャート（「タイプ決定条件 1」～「タイプ決定条件 12」）である。

「タイプ決定条件 1」と「タイプ決定条件 2」の各ステップ（①～⑩）について以下に説明する。

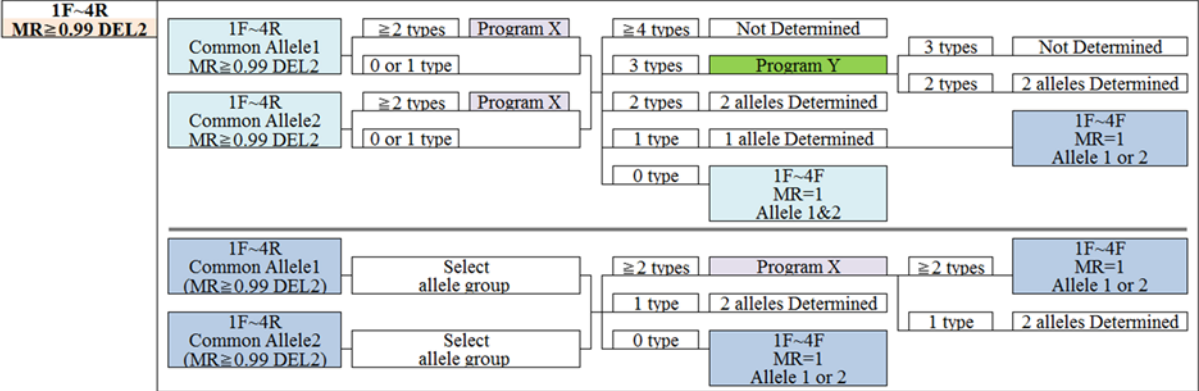
- ①：項目 2-1-3-4-1 で説明した方法により、2 アレルグループを区別する。「Allele 1」はアレルグループ 1、「Allele 2」はアレルグループ 2 を表す。
- ②：各アレルグループに含まれるタイプ数を表す。
- ③：各アレルグループに含まれるタイプ数が 2 タイプ以上である場合、Program X を実行する。
- ④：③までの処理により、候補として残ったタイプ数とその場合分けを表す。4 タイプ以上の場合、HLA タイプ決定プログラムでは 2 アレルの HLA タイプを決定できなかったサンプルとなる。3 タイプである場合、Program Y を実行させる。2 タイプである場合、その 2 タイプを決定する。1 タイプである場合、まずそのタイプを 1 アレルのタイプと決定する。0 タイプである場合、MR = 1 では決定できないサンプルと判断し、⑦-1 の処理へ移行する。
- ⑤：Program Y により出力されたタイプ数を表す。そのタイプ数が 3 タイプである場合、HLA タイプ決定プログラムでは 2 アレルの HLA タイプを決定できなかったサンプルとし、2 タイプである場合、その 2 タイプを決定とする。
- ⑥：1 アレルのタイプは決定し、もう片アレルのタイプを決定するために調べる次の条件 $MR \geq 0.99$ 、DEL 1 を表す。この場合、もう片アレルのタイプが追加処理を必要とするタイプである可能性を考慮して⑦-2 の処理へ移行する。
- ⑦-1： $MR \geq 0.99$ 、DEL 1 の 1 段目は $MR = 1$ において判定したタイプ数が 0 である場合（④の 0 タイプの場合）のルートを表す。そのため、最高 2 タイプを決定するためのルートになる。
- ⑦-2： $MR \geq 0.99$ 、DEL 1 の 2 段目は $MR = 1$ において判定したタイプ数が 1 であり、1 タイプ既に決定できた場合（⑥）からのルートになる。そのため、最高 1 タイプを決定するためのルートになる。
- ⑧：既に決定したタイプが存在していない方のアレルグループを選択する。

- ⑨：⑧で選択したアレルグループに存在しているタイプ数とその場合分けを表す。⑧で選択したアレルグループに存在しているタイプ数が 2 タイプ以上である場合、Program X を実行する。1 タイプである場合、そのタイプをもうアレルのタイプとして決定する。0 タイプである場合、 $MR \geq 0.99/DEL2$ を調べる。
- ⑩：Program X により出力されたタイプ数とその場合分けを表す。2 タイプ以上である場合、欠失を許すと決定できないタイプだと判断し、残りの追加処理である「HLA タイプの長さがアレル間で異なる場合の追加処理」を実行する。1 タイプである場合、そのタイプをもうアレルのタイプとして決定する。

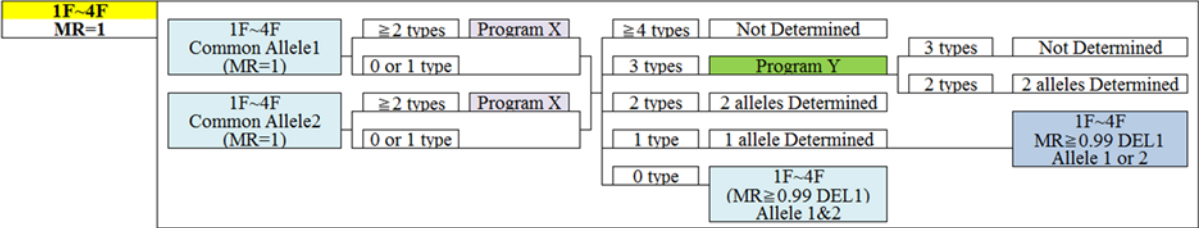
以上①～⑩と同様の処理をタイプ決定条件 3～12 においても図 26 に示すチャートに従い実行する。最後まで調べて 1 タイプしか決定できなかった場合、ホモ接合型のサンプルとして判定した。



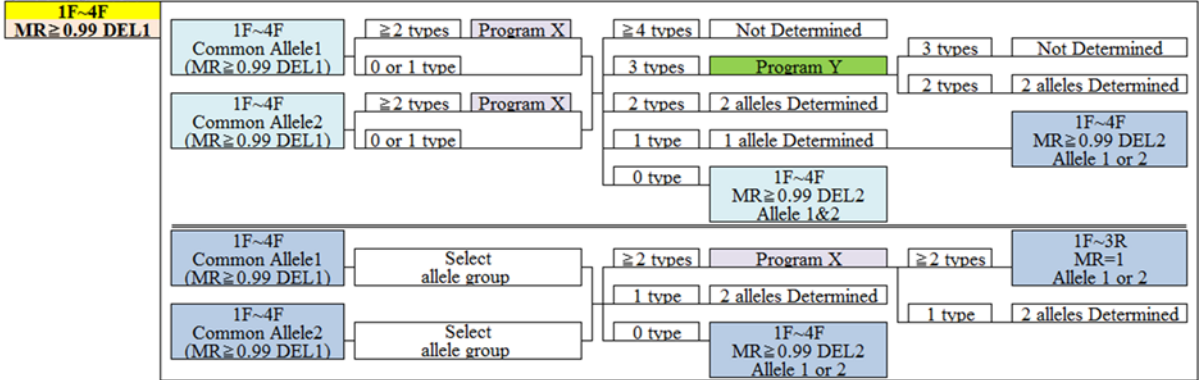
タイプ決定条件3



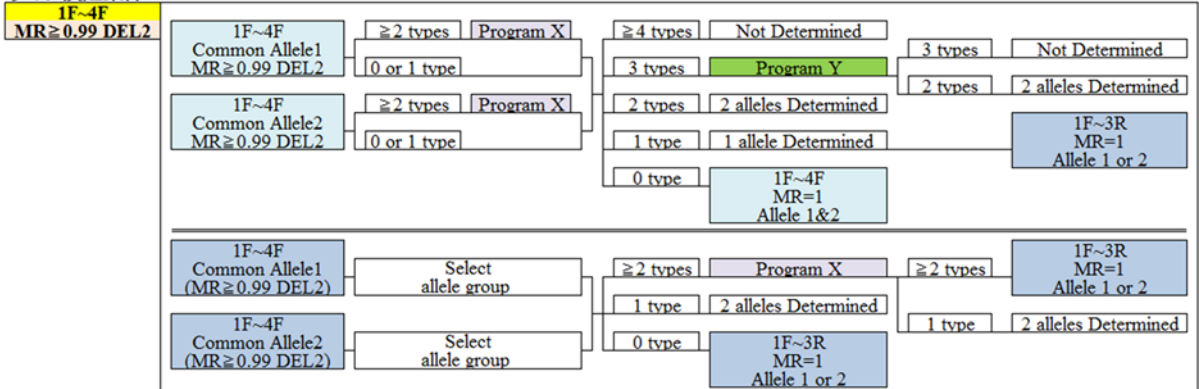
タイプ決定条件4



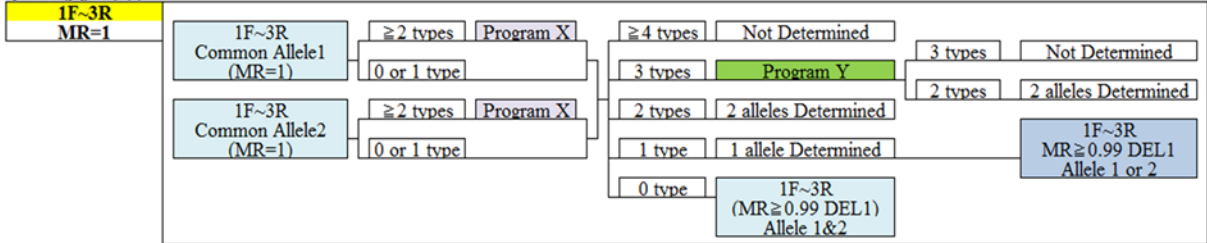
タイプ決定条件5



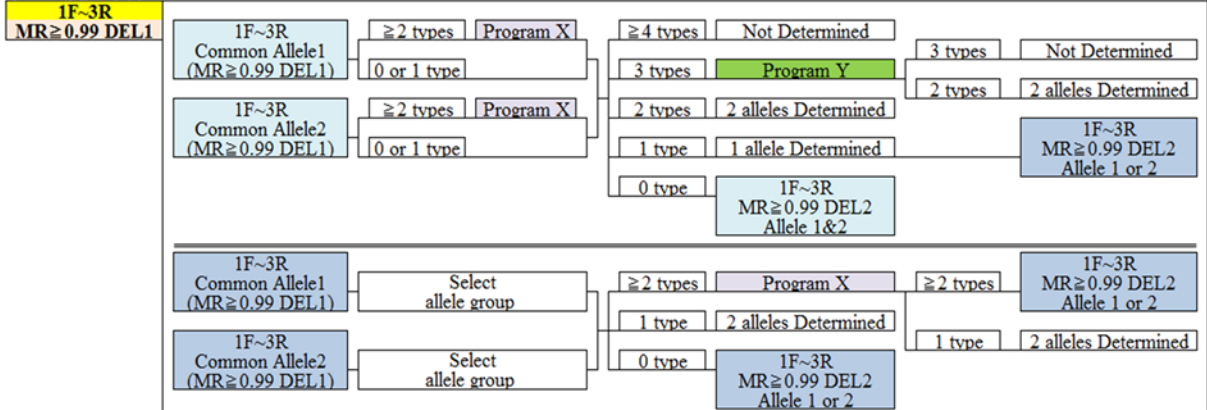
タイプ決定条件6



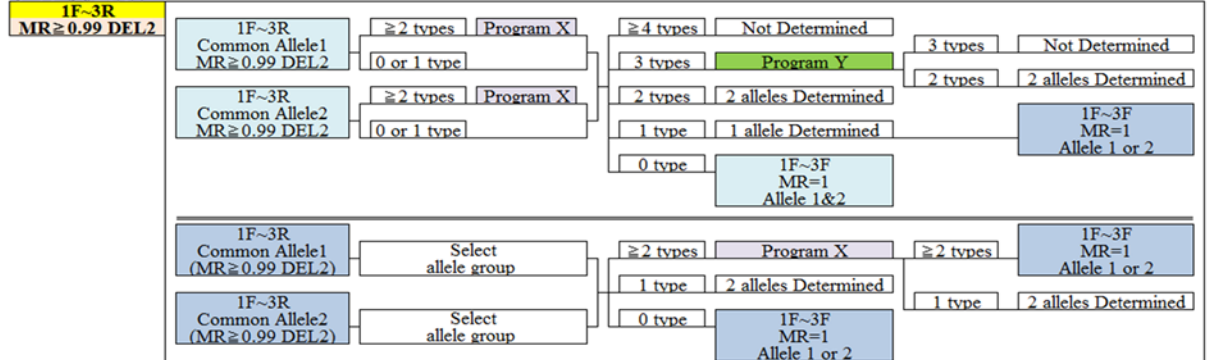
タイプ決定条件7



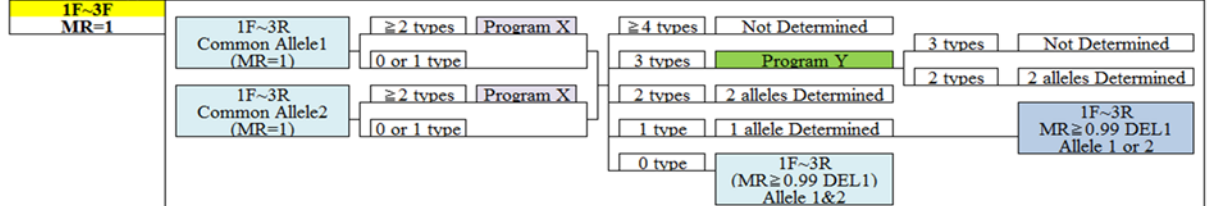
タイプ決定条件8



タイプ決定条件9



タイプ決定条件10



4-4. サンガー法による HLA タイプの決定

4-4-1. コーディング領域の PCR

設計した 27 ペアの HLA プライマー (表 2) のうち、最も外側同士のプライマーペア (HLA クラス I : 1F と 4R、HLA クラス II : 1F と 3R) を用いてコーディング領域の PCR を行った。1 μ l の一本鎖 cDNA 反応液 (項目 4-2-1 を参照)、4 μ l の 5x Q5 buffer、1.6 μ l の 2.5mM dNTP、2 μ l の 10 μ M HLA プライマーペア、11.2 μ l の Nuclease-Free water、0.2 μ l の Q5 Hot Start High-Fidelity DNA Polymerase を混合し 20 μ l の反応系で PCR を行った。Veriti[®] サーマルサイクラーを用い、98 $^{\circ}$ C で 30 秒間インキュベートした後、98 $^{\circ}$ C で 10 秒間、50 $^{\circ}$ C で 10 秒間、72 $^{\circ}$ C で 30 秒間というサイクルを 30 回行った後、72 $^{\circ}$ C で 2 分間インキュベートした。

4-4-2. PCR 産物のリン酸化

作成した PCR 産物 (項目 4-4-1 を参照) とサブクローニング用のベクター DNA をライゲーションさせるために、PCR 産物のリン酸化を行った。6.8 μ l の PCR 産物、1 μ l の 10 x T4 PNK Reaction Buffer (NEB)、1 μ l の 10 mM ATP (NEB)、0.2 μ l の T4 PNK (NEB) を混合し、10 μ l の反応液量において 37 $^{\circ}$ C で 30 秒間インキュベートし、65 $^{\circ}$ C で 20 分間、酵素を不活化させた。

4-4-3. サブクローニング

4-4-3-1. T4 DNA ligase によるライゲーション

10 fmol の pCR-Blunt (Invitrogen)、30 fmol のリン酸化 PCR 産物 (項目 4-4-2 を参照)、1 μ l の 10x T4 DNA Ligase Buffer (NEB)、0.5 μ l の 400 units/ μ l T4 DNA Ligase (NEB)、Nuclease-Free water を合計 10 μ l の液量となるように混合し、室温で 2 時間インキュベートした後、65 $^{\circ}$ C で 10 分間、酵素を不活化させた。

4-4-3-2. 形質転換

形質転換には、MAX Efficiency DH5 α Competent Cells (Invitrogen) を用いた。0.1 x TE を用いて 5 倍希釈したライゲーション産物 (項目 4-4-3-1 を参照) 1 μ l を 50 μ l のコンピテントセル溶液へ加え、混合した後、氷上で 30 分間静置させ、42 $^{\circ}$ C で 45 秒間ヒートショックさせた。氷上で 2 分間静置させた後、900 μ l の SOC 培地を加え、15 ml チューブへ移し、温度 37 $^{\circ}$ C の条件下において 1 時間攪拌させた。その後、カナマイシン含有 LB プレート上で 37 $^{\circ}$ C 16 時間インキュベートさせた。

4-4-3-3. 各コロニーを鋳型とした PCR

項目 4-4-3-2 で培養した大腸菌のそれぞれのコロニーを鋳型とし、インサートの外側に位置するプライマーを用いて PCR を行った。

1 μ l の 10x High Fidelity PCR Buffer (Invitrogen)、0.8 μ l の 2.5 mM dNTP、0.1 μ l の 10 μ M seqF primer (5'-GCCAGTGTGCTGGAATTCA-3')、0.1 μ l の 10 μ M seqR primer (5'-CCAGTGTGATGGATATCTGCAGA-3')、7.36 μ l の Nuclease-Free water、0.04 μ l の Platinum Taq DNA Polymerase, High Fidelity (Invitrogen) の混合液 10 μ l に、各コロニーを溶解させ、PCR を行った。Veriti[®] サーマルサイクラーを用い、98 $^{\circ}$ C で 30 秒間インキュベートした後、94 $^{\circ}$ C で 15 秒間、60 $^{\circ}$ C で 30 秒間、68 $^{\circ}$ C で 60 秒間というサイクルを 40 回行った。

4-4-3-4. Exo I / SAP 処理

1 μ l の 10x SAP Buffer (Takara bio)、0.32 μ l の 1 unit/ μ l SAP (Takara bio)、0.16 μ l の 5 unit/ μ l Exo I (NEB)、作成した 10 μ l の PCR 産物 (項目 4-4-3-3 を参照)、8.52 μ l の Nuclease-Free water を混合し、20 μ l の反応液量で 37 $^{\circ}$ C で 30 分間インキュベートした後、90 $^{\circ}$ C で 10 分間酵素を失活させた。

4-4-3-5. シーケンス反応

設計した 27 の HLA プライマーペア (項目 4-2-2 を参照) のうち、最も外側のプライマー (HLA クラス I では 1F と 4R、HLA クラス II では 1F と 3R) を用いてシーケンス反応を行なった。

1 μ l の Big Dye v3.1 (Applied Biosystems)、0.16 μ l の 5 x sequencing buffer (applied biosystems)、0.32 μ l の 10 μ M HLA プライマー、10 μ l の Exo I /SAP 処理 PCR 産物 (項目 4-4-3-4 を参照)、8.52 μ l の Nuclease-Free water を混合し、96 $^{\circ}$ C で 30 分間インキュベートした後、96 $^{\circ}$ C で 10 秒間、55 $^{\circ}$ C で 5 秒間、60 $^{\circ}$ C で 2 分間というサイクルを 30 回行い、反応させた。

4-4-3-6. ゲル濾過

シーケンス反応液 (項目 4-4-3-5 を参照) を精製するために、Sephadex G50 溶液を用いたゲル濾過を行った。Sephadex G50 溶液は 400 ml の Nuclease-Free water と 20 g Sephadex G50 (GE healthcare) を混合し、121 $^{\circ}$ C で 20 分間オートクレーブしたものである。96-microplate の上に置いた Multiscreen HV plate (Merck) のウェルに 300 μ l の Sephadex G50 溶液を分注し、2500 rpm の回転速度で 1 秒間遠心した後、さらに 100 μ l の Sephadex G50 溶液を加え、2500 rpm の回転速度で 5 分間遠心し、ゲル濾過用のブ

プレートを作成した。シーケンス反応液に 10 μ l の Nuclease-Free water を加えた合計 30 μ l の反応液をゲル濾過用のプレートに添加し、2500 rpm の回転速度で 5 分間遠心し、精製されたシーケンス反応液（濾液）を回収した。

4-4-3-7. キャピラリー電気泳動

精製したシーケンス反応液（項目 4-4-3-6 を参照）を 90 $^{\circ}$ C で 30 秒間インキュベートした後、ABI PRISM 3100 Genetic Analyzer（Applied Biosystems）によりキャピラリー電気泳動を行い、Bio Edit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) を用い、塩基配列の決定を行った。

4-5. 補完 HLA タイピングプログラム用ライブラリー調製

4-5-1. 分子バーコード配列を含むオリゴマーによる一本鎖 cDNA 合成

100 ng の全 RNA を用いて一本鎖 cDNA を合成した。1 μ l の 100 ng/ μ l 全 RNA (10 ng)、7 μ l の Nuclease-Free water、4 μ l の 2.5 mM dNTP、1 μ l の 10 μ M DPB1-R-adaptor プライマー（項目 2-3-2-1 を参照）を混合し、65 $^{\circ}$ C で 5 分間インキュベートした後、2 分間氷上で静置した。

次に、4 μ l の 5 x 1st strand buffer、1 μ l の 0.1M DTT、1 μ l の RNase Out ribonuclease Inhibitor、1 μ l の Super Script III Reverse Transcriptase を混合し、13 μ l の DPB1-R-adaptor プライマーとアニーリングした全 RNA に加え、25 $^{\circ}$ C で 5 分間、50 $^{\circ}$ C で 60 分間、70 $^{\circ}$ C で 15 分間インキュベートした後、氷上で冷却した。

4-5-2. HLA-DPB1 遺伝子座領域の PCR

当項目では 2 回の PCR を行った。1 μ l の一本鎖 cDNA（項目 4-5-1 を参照）、16.99 μ l の Nuclease-Free water、2.5 μ l の 10x High Fidelity PCR Buffer、2.2 μ l の 2.5 mM dNTP、1.2 μ l の 50 mM MgSO₄ (Invitrogen)、0.11 μ l の 5 unit/ μ l Platinum Taq、0.5 μ l の 10 μ M の DPB1-F プライマー（項目 2-3-2-1 を参照）、0.5 μ l の 10 μ M の biotin-Spacer プライマー（項目 2-3-2-1 を参照）、合計 25 μ l を 94 $^{\circ}$ C で 30 秒間インキュベートさせた後、94 $^{\circ}$ C で 30 秒間、65 $^{\circ}$ C で 30 秒間、68 $^{\circ}$ C で 1 分間という反応を 30 サイクル行い、Veriti[®] サーマルサイクラーを用い、PCR を行った。PCR 反応液は AMPure XP kit を用いて、プロトコールに従い精製を行った。2 回目の PCR は、1 ウェルあたり精製した PCR 産物 1 μ l を鋳型とし、リン酸化したプライマー（p-DPB1-F プライマー）を用いて、8 ウェルにおいて PCR を行った。PCR 反応液は AMPure XP kit を用い、8 ウェル分（200 μ l）まとめて精製し、Qubit dsDNA HS Assay Kit を用いて濃度測定を行った。

4-5-3. ビオチン化 PCR 産物の断片化

項目 4-5-2 で精製した PCR 産物 (84.2 ng/ μ l) 200 ng の断片化を行った。2.4 μ l の PCR 産物、2 μ l の 10x Fragmentase Reaction Buffer v2 (NEB)、13.6 μ l の Nuclease-Free water の合計 18 μ l を混合した後、氷上に静置していた 2 μ l の ds DNA fragmentase (NEB) を加え、合計 20 μ l とした後、37 °C で 5 分間インキュベートさせた。酵素の不活化を行う為、20 μ l の反応液に、0.5 M EDTA を 7.5 μ l 加えた。反応液は AMPure XP kit を用いて、プロトコールに従い精製を行った (20 μ l の Nuclease-Free water により断片化されたビオチン化 PCR 産物を回収)。

4-5-4. 一本鎖 DNA の調製

10 mg/ml の Dynabeads® MyOne™ Streptavidin C1 (Thermo fisher Scientific) 溶液を用い、断片化 PCR 産物 (項目 4-5-3 を参照) から一本鎖 DNA を調製した。

30 秒間攪拌させた Dynabeads® MyOne™ Streptavidin C1 溶液 10 μ l を 1.5 ml 低吸着チューブへ分注した。分注した Dynabeads® MyOne™ Streptavidin C1 溶液を 1 x B&W Buffer (5 mM Tris-HCL pH 7.5、0.5 mM EDTA、1 M NaCl) 10 μ l を用い 3 回洗浄した。

洗浄後、20 μ l の 2 x B&W Buffer と 20 μ l の断片化されたビオチン化 PCR 産物を加えた。攪拌を 3 分おきに行いながら 15 分間室温でインキュベートした。DynaL Invitrogen Bead Separator (Invitrogen) に 3 分間静置させ、上清を除去した。その後、100 μ l の 1 x B&W Buffer を用い 3 回洗浄した。

次に、ビオチンが付いていない一本鎖 DNA を回収するためにアルカリ変性を行った。100 μ l のアルカリ溶液 (0.1 M NaOH, 1 mM EDTA) と 1 μ l の 10% Triton X-100 を Dynabeads® MyOne™ Streptavidin C1 により固定化されたビオチン化 PCR 産物に加え、2 分おきに攪拌を行いながら室温で 10 分間インキュベートした。上清 (ビオチンが付いていない一本鎖 DNA が含まれている) を 2 ml 低吸着チューブへ分注した。1 回の回収では回収しきれなかった一本鎖 DNA を回収する目的で、再度アルカリ溶液、10% Triton X-100 を加える工程から繰り返した。合計で 200 μ l のビオチンが付いていない一本鎖 DNA が含まれているアルカリ溶液を回収した。

次に、そのアルカリ溶液を中性化させるために 20 μ l の 1 M HCL を加え、攪拌させ、NEB Monorch PCR DNA clean up kit (NEB) を用いて、一本鎖 DNA を精製するプロトコールに従い一本鎖 DNA の精製を行った。

4-5-5. リアルタイム PCR による一本鎖 DNA の濃度測定

精製した一本鎖 DNA (項目 4-5-4 を参照) の濃度をリアルタイム PCR により測定した。リアルタイム PCR で使用したプライマー (DPB1-RT-F, DPB1-RT-R) について 図 27 に示した。DPB1-R-adaptor の配列がすべて含まれているような断片化一本鎖 DNA を測定するために、リアルタイム PCR の解析領域は DPB1-R-adaptor の配列に近い領域とした。DPB1-RT-F, DPB1-RT-R はそれぞれ DPB1 遺伝子座で保存されている領域に設計した。

濃度測定用の検量線は、項目 4-5-2 で作成した PCR 産物 (断片化する前の二本鎖 DNA) を用いて作成し、一本鎖 DNA の濃度を測定した。1 ウェルあたり 10 μ l の SYBR[®] Premix Ex Taq II (Tli RNaseH Plus) (Takara bio)、1.6 μ l の 10 μ M の一本鎖 DNA 濃度測定用プライマーセット (DPB1-RT-F と DPB1-RT-R)、0.4 μ l の Rox Reference Dye II (Takara bio)、1 μ l の鋳型 DNA、7 μ l の Nuclease-Free water を混合し、合計 20 μ l の液量とした。Applied Biosystems 7500 Real-Time PCR System を用い 95 $^{\circ}$ C で 30 秒間行った後、95 $^{\circ}$ C で 5 秒間、60 $^{\circ}$ C で 34 秒間というサイクルを 40 回繰り返し、産物量の定量を行った。

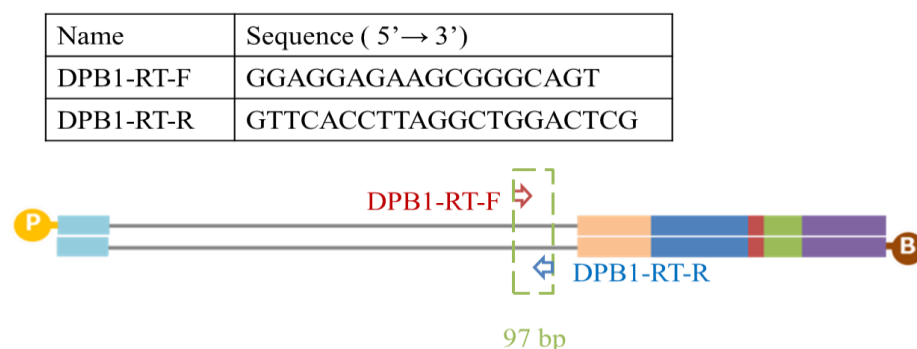


図 27 : リアルタイム PCR による一本鎖 DNA の濃度測定

リアルタイム PCR による濃度測定に使用するプライマーの配列情報と解析領域 (緑の点線) 位置

4-5-6. 一本鎖 DNA の環状化

T4 RNA Ligase 1 (NEB) を用いて、一本鎖 DNA (反応液中の DNA 濃度 : 10 amol/ μ l) を環状化させた。1 本の 0.2 ml PCR tube につき、2 μ l の 10 x T4 RNA Ligase Buffer (NEB)、0.4 μ l の 10 mM ATP、4 μ l の 50 % PEG8000 (NEB)、0.2 μ l の T4 RNA Ligase 1、3.7 μ l の Nuclease-Free water、10 μ l の 20 amol/ μ l 一本鎖 DNA (項目 4-5-4 を参照) を混合し、37 $^{\circ}$ C で 16 時間インキュベートした後、65 $^{\circ}$ C で 15 分間、酵素の不活化を行った。

次に、環状化しなかった一本鎖 DNA を除去するため、20 μ l の反応液に 3 μ l の 10 x

Exonuclease I Reaction Buffer (NEB)、1 μl の 20 units/ μl Exonuclease I (NEB)、6 μl の Nuclease-Free water を加え、37°C で 60 分間インキュベートさせた後、酵素の不活化のために 80°C で 20 分間反応させた。当反応を 8 つの 0.2 ml PCR tube で行い、反応液量は合計で 240 μl となった。

NEB Monorch PCR DNA clean up kit を用いて、環状化一本鎖 DNA を精製するプロトコールに従い、240 μl の環状化一本鎖 DNA の精製を行った。

4-5-7. 環状化 DNA を鋳型にした PCR

12 種類の DPB1 遺伝子座特異的プライマーと A プライマー (表 10, 図 28) を用い、精製した環状化 DNA (項目 4-5-6 を参照) を鋳型とした PCR を行った。

1 μl の精製した環状化 DNA (項目 4-5-6 を参照)、16.99 μl の Nuclease-Free water、2.5 μl の 10x High Fidelity PCR Buffer、2.2 μl の 2.5 mM dNTP、1.2 μl の 50 mM MgSO_4 、0.11 μl の 5 unit/ μl Platinum Taq、1 μl の 5 μM ずつのプライマーペアの合計 25 μl を 94 °C で 30 秒間インキュベートさせた後、94 °C で 30 秒間、65 °C で 30 秒間、68 °C で 1 分間という反応を 40 サイクル行い Veriti® サーマルサイクラーにより PCR を行った。

各 PCR 産物を混合し、200~500 bp の範囲でゲル切り出しを行った。MinElute Gel Extraction Kit (Qiagen) により切り出したゲル切片中 DNA の精製を行った後、Quanti-iT™ dsDNA Assay Kit, Broad range を用いて、精製した PCR 産物混合液の濃度をプロトコールに従い測定した。Ion PGM™ Sequencing Hi-Q kit (Thermo Fisher Scientific) を用い項目 4-2-5 で説明した同様の方法で Ion Torrent PGM による配列解析を行った。

プライマー名	配列 (5'→3') P1プライマー配列 DPB1遺伝子座特異的配列	アニーリング温度	マッピングポジション
P1-DPB1-No1	CCTCTCTATGGGCAGTCGGTGATGTGAGCAGCACCATCAGTAAC	53	97
P1-DPB1-No2	CCTCTCTATGGGCAGTCGGTGATGTAATCTCTGGAGTGGCCCTG	55	134
P1-DPB1-No3	CCTCTCTATGGGCAGTCGGTGATTGTAGATGTATCTCTCCAGGAAGC	52	204
P1-DPB1-No4	CCTCTCTATGGGCAGTCGGTGATGTCCTCTGGCTGTTCCAG	53	306
P1-DPB1-No5	CCTCTCTATGGGCAGTCGGTGATCCAGCTCGTAGTTGTGICTG	53	363
P1-DPB1-No6	CCTCTCTATGGGCAGTCGGTGATCCAGCTCGTAGTTGTGICTG	56	427
P1-DPB1-No7	CCTCTCTATGGGCAGTCGGTGATGGACTTGAATGCTGCCTGG	54	495
P1-DPB1-No8	CCTCTCTATGGGCAGTCGGTGATCCAGTCTCCATTACGGATCAGG	54	566
P1-DPB1-No9	CCTCTCTATGGGCAGTCGGTGATGTGTCTCCACTTGGCAG	53	639
P1-DPB1-No10	CCTCTCTATGGGCAGTCGGTGATCAGTCCCCTCAATGTCTTACTC	55	717
P1-DPB1-No11	CCTCTCTATGGGCAGTCGGTGATGCAGATCCTCGTTGAACCTTCTTG	53	799
P1-DPB1-No12	CCTCTCTATGGGCAGTCGGTGATCACAGCAAATGCTTTTCCTAAGG	52	861
Aプライマー	CCATCTCATCCCTGCGTGTC		

Aプライマー、P1プライマー：Ion Torrent PGMによる配列解析で使用するプライマー

表 10 : 環状化 DNA を鋳型とした PCR で使用するプライマーセット

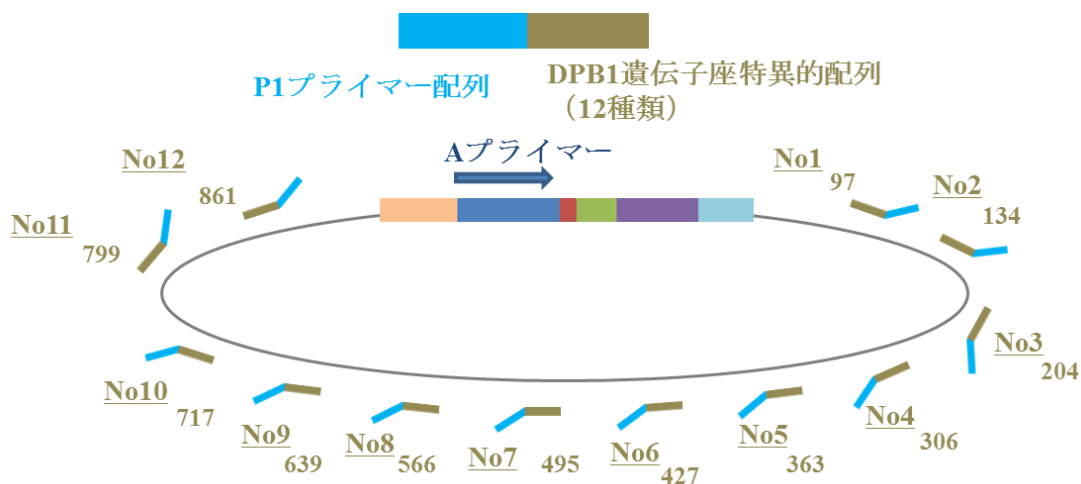


図 28：環状化 DNA を鋳型にした PCR

P1-DPB1-No1~12 それぞれのプライマーが断片化していない時の環状化 DNA へ、アニーリングする様子を示した。各数字は、アニーリングした時の A プライマーからの塩基長を表す。

4-6. 補完 HLA タイピングプログラムによる解析方法

配列解析が行われたリードデータに対し、各リードの分子バーコード配列情報を基に、分子バーコード配列あたりのリード数を取得した。そして、解析領域において、分子バーコード配列あたりのリード数が大きい 100 種類の分子バーコード配列を選択した。その後、以下のような方法で各分子バーコード配列グループのコンセンサス配列を作成した。

4-6-1. リードグループごとのコンセンサス配列の取得

DPB1 遺伝子座の参照配列に対し、マッピングツールである BWA-ver 0.7.15 (<http://bio-bwa.sourceforge.net/>) を用いてマッピングを行ない、選択した分子バーコード配列を持つリードの SAM 形式ファイルを出力した。

マッピングポジションごとの塩基の数をカウントするためのツールである IGVtools (<https://software.broadinstitute.org/software/igv/igvtools>) を用い、リードの端のポジションが同じリード集団をリードグループとし、そのコンセンサス配列を作成した。当処理は、各 PCR 間の増幅バイアスを解消するため、また、リードグループ内での PCR エラーを除去するために行う。IGVtools の出力データを用い、ポジションごとの塩基の割合を計算し、最も割合の高い塩基をそのポジションの塩基と決定していくことで解析領域全体のコンセンサス配列を作成した。

なお、各リードグループを作成するためのスタートポジション、マッピングエンドポジションは項目 4-3-1 と同様にして取得した。

4-6-2. 解析領域のコンセンサス配列作成

リードグループごとのコンセンサス配列ファイル（項目 4-7-1 を参照）を、DPB1 遺伝子座の参照配列に対し、BWA-ver 0.7.15 を用いてマッピングを行ない、SAM 形式ファイルとして出力した。そして、IGVtools を用い、解析に使用する領域のコンセンサス配列を作成した。ここまでの処理を分子バーコード配列ごとに行い、100 分子バーコード配列分のコンセンサス配列を作成した。

4-6-3. HLA cDNA データベース検索

作成した各コンセンサス配列（項目 4-7-2 を参照）がどの HLA タイプに完全一致するかを調べた。解析の流れは HLA タイピングプログラムで開発した方法（項目 4-3-3 を参照）と同様であり、HLA-DPB1 の HLA cDNA データベースに対するアライメントを行なった後に、MR を計算し、MR = 1 であるものを完全一致タイプとして出力した。そして、各タイプと一致したコンセンサス配列数を勘定し、HLA タイピングシステムのアライメントデータと統合し、HLA タイプを決定した。

5. 謝辞

本研究に際し、御指導、御鞭撻を賜りました株式会社 DNA チップ研究所 加藤菊也部長（元奈良先端科学技術大学院大学客員教授／大阪府立成人病センター研究所長）、奈良先端科学技術大学院大学 疾患分子遺伝学講座 客員准教授 兼 大阪府立成人病センター研究所主任研究員 久木田洋児先生に心から感謝いたします。そして、本研究の論文審査の主査を務めて頂きました奈良先端科学技術大学院大学 腫瘍細胞生物学講座 加藤順也教授、また、論文審査の副査を務めて頂きました奈良先端科学技術大学院大学 分子情報薬理学講座 伊東広教授、計算システムズ生物学講座 金谷重彦教授に深く感謝致します。

最後に、本学での五年間に及ぶ研究生生活を支えてくださった大阪府立成人病センター研究所、奈良先端科学技術大学院大学の皆様、そして両親、友人に心から感謝致します。

6. 参考文献

- 1) Gorer, P. A. (1937). The genetic and antigenic basis of tumor transplantation. *J. Pathol. Bacteriol.* 44, 691-697.
- 2) Snell, G. D. (1948). Methods for the study of histocompatibility genes. *J. Genet.* 49, 87-108.
- 3) Dausset, J. (1958). Iso-leuco-anticorps. *Acta. Haematol.* 20, 156-166.
- 4) Thorsby, E. (2009). A short history of HLA. *Tissue Antigens.* 74, 101-116.
- 5) Shiina, T., Suzuki, S., Kulski, J. K. (2016). MHC Genotyping in Human and Nonhuman Species by PCR-based Next-Generation Sequencing. *Next Generation Sequencing - Advances, Applications and Challenges.* Chapter 3.
- 6) Shiina, T., Hosomichi, K., Inoko, H., Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* 54, 15-39.
- 7) Choo, S. Y. (2007). The HLA System: Genetics, Immunology, Clinical Testing and Clinical Implication. *Yonsei Med. J.* 48, 11-23.
- 8) Tiercy, J. M. (2002). Molecular basis of HLA polymorphism: implications in clinical transplantation. *Transpl. Immunol.* 9, 173-180.
- 9) Shankarkumar, U., Pawar, A., Ghosh, K. (2008). Implications of HLA sequence-based typing in transplantation. *J Postgrad. Med.* 54, 41-44.
- 10) The MHC sequencing consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature.* 401, 921-923.
- 11) Marsh, S. G. E., Albert E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Fernandez-Vina, M., Geraghty, D. E., Holdsworth, R., Hurley, C. K., Lau, M., Lee, K. W., Mach, B., Maiers, M., Mayr, W. R., Muller, C. R., Parham, P., Petersdorf, E. W., Sasazuki, T., Strominger, J. L. et al. (2010). Nomenclature for factors of the HLA system, 2010. *Tissue Antigens.* 75, 291-455.
- 12) Nunes, E., Heslop, H., Fernandez-Vina, M., Taves, C., Wagenknecht, D. R., Eisenbrey, A. B., Fischer, G., Poulton, K., Wacker K., Hurley, C. K., Noreen, H., Sacchi, H. (2011). Definitions of histocompatibility typing terms: Harmonization of Histocompatibility Typing Terms Working Group. *Hum. Immunol.* 72, 1214-1216.
- 13) He, J., Li, Y., Bao, X., Qiu, Q., Yuan, X., Xu, C., Shen, J., Hou, J. (2012). Common and well-documented (CWD) alleles of human leukocyte antigen-A, -B, -C, -DRB1, and -DQB1 loci for the Chinese Han population do not quite correlate with the ASHI CWD alleles. *Hum. Immunol.* 73, 61-66.
- 14) Sasazuki, T., Juji, T., Morishima, Y., Kinukawa, N., Kashiwabara, H., Inoko, H., Yoshida, T., Kimura, A., Akaza, T., Kamikawaji, N., Koder, Y., Takaku, F. (1998). Effect of matching of class I HLA alleles on clinical outcome after transplantation of hematopoietic

- stem cells from an unrelated donor. Japan Marrow Donor Program. *N. Engl. J. Med.* 339, 1177-1185.
- 15) Morishima, Y., Kashiwase, K., Matsuo, K., Azuma, F., Morishima, S., Onizuka, M., Yabe, T., Murata, M., Doki, N., Eto, T., Mori, T., Miyamura, K., Sao, H., Ichinohe, T., Saji, H., Kato, S., Atsuta, Y., Kawa, K., Kodera, Y., Sasazuki, T. (2015). Biological significance of HLA locus matching in unrelated donor bone marrow transplantation. *Blood.* 125, 1189-1197.
- 16) Eapen, M., Rubinstein, P., Zhang, M. J., Stevens, Kurtzberg, J., Scaradavou, A., Loberiza, F. R., Champlin, R. E., Klein, J. P., Horowitz, M. M., Wagner, J. E. (2007). Outcomes of transplantation of unrelated donor umbilical cord blood and bone marrow in children with acute leukaemia: a comparison study. *Lancet.* 369, 1947-1954.
- 17) Spellman, S. R., Eapen, M., Logan, B. R., Mueller, C., Rubinstein, P., Setterholm, M. I., Woolfrey, A. E., Horowitz, M. M., Confer, D. L., Hurley, C. K. (2012). A perspective on the selection of unrelated donors and cord blood units for transplantation. *Blood.* 120, 259-265.
- 18) Gao, X., Bashirova, A., Iversen, A. K. N., Phair, J., Goedert, J. J., Buchbinder, S., Hoots, K., Vlahov, D., Altfeld, M., O'Brien, S. J., Carrington, M. (2005). AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat. Med.* 11, 1290-1292.
- 19) Lie, B. A., Thorsby, E. (2005). Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr. Opin. Immunol.* 17, 526-531.
- 20) Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A. M., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., James, I., Christiansen, F. T. (2002). Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet.* 359, 727-732.
- 21) Gabriel, C., Stabentheiner, S., Danzer, M., Proll, J. (2011). What Next? The Next Transit from Biology to Diagnostics: Next Generation Sequencing for Immunogenetics. *Transfus. Med. Hemother.* 38, 308-317.
- 22) Ottinger, H. D., Ferencik, S., Beelen, D. W., Lindemann, M., Peceny, R., Elmaagacli, A. H., Husing, J., Grosse-Wilde, H. (2003). Hematopoietic stem cell transplantation: contrasting the outcome of transplantations from HLA-identical siblings, partially HLA-mismatched related donors, and HLA-matched unrelated donors. *Blood.* 102, 1131-1137.
- 23) Kassar, N. E., Legouvello, S., Joseph, C. M., P Salesses, P., Rieux, C., Cordonnier, C., Vernant, J. P., Farcet, J. P., Bierling, P., Kuentz, M. (2001). High resolution HLA class I and II typing and CTLp frequency in unrelated donor transplantation: a single-institution retrospective study of 69 BMTs. *Bone Marrow Transplant.* 27, 35-43.
- 24) Carapito, R., Radosavljevic, M., Bahram, S. (2016). Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease

- association studies. *Hum. Immunol.* 77, 1016-1023.
- 25) Wagner, J. E., Barker, J. N., DeFor, T. E., Baker, K. S., Blazar, B. R., Eide, C., Goldman, A., Kersey, J., Krivit, W., MacMillan, Orchard, P. J., Peters, C., Weisdorf, D. J., Ramsay, N. K. C., Davies, S. M. (2002). Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival. *Blood.* 100, 1611-1618.
 - 26) Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., Cherry, J. M., Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790-1797.
 - 27) Vandiedonck, C., Taylor, M. S., Lockstone, H. M., Plant, K., Taylor, J. M., Durrant, C., Broxholme, J., Fairfax, B. P., Knight, J. C. (2011). Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. *Genome Res.* 21, 1042-1054.
 - 28) Horton, R., Niblett, D., Milne, S., Palmer, S., Tubby, B., Trowsdale, J., Beck, S. (1998). Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. *J. Mol. Biol.* 282, 71-97.
 - 29) Vandiedonck, C., Knight, J. C. (2009). The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief. Funct. Genomic Proteomic.* 8, 379-394.
 - 30) Bunce, M., O'Neill, C. M., Barnardo, M. C., Krausa, P., Browning, M. J., Morris, P. J. Welsh, K. I. (1995). Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens.* 46, 355-367.
 - 31) Oleruo, O., Zetterquist, H. (1992). HLA-DR typing by PCR amplification with sequencespecific primers (PCR-SSP) in 2 hours: An alternative to serological DR typing in clinical practice including donor-recipient matchmg in cadaveric transplantation. *Tissue Antigens.* 39, 225-235.
 - 32) Dunbar, S. A. (2006). Applications of LuminexR xMAPi technology for rapid, high-throughput multiplexed nucleic acid detection. *Clin. Chim. Acta.* 363, 71-82.
 - 33) Saiki, R. K., Walsh, P. S., Levenson, C. H., Erlich, H. A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Sci. USA.* 86, 6230-6234.
 - 34) Santamaria, P., Lindstrom, A. L., Boyce-Jacino, M. T., Myster, S. H., Barbosa, J. J., Faras, A.J., Rich, S. S. (1993). HLA class I sequence-based typing. *Hum. Immunol.* 37, 39-50.
 - 35) Adams, S. D., Barracchini, K. C., Chen, D., Robbins, F., Wang, L., Larsen, P., Luhm, R., Stroncek, D. F. (2004). Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification. *J. Transl. Med.* 2, 30.

- 36) Hosomichi, K., Shiina, T., Tajima, A., Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *J. Hum. Genet.* 60, 665-673.
- 37) Erlich, H. A. (2015). HLA typing using next generation sequencing: An overview. *Hum. Immunol.* 76, 887-890.
- 38) Cox, S. T., McWhinnie, A. J., Robinson, J., Marsh, S. G. E., Parham, P., Madrigal, J. A., Little, A. M. (2010). Cloning and sequencing full-length HLA-B and -C genes. *Yi. Chuan.* 32, 685-693.
- 39) Erlich, H. (2012). HLA DNA typing: past, present, and future. *Tissue Antigens.* 80, 1-11.
- 40) Meral, B. (2014). Bone Marrow and Stem Cell Transplantation. Second edition, 73-138.
- 41) Dunn, P. P. (2011). Human leucocyte antigen typing: techniques and technology, a critical appraisal. *Int. J. Immunogenet.* 38, 463-73.
- 42) Gabriel, C., Danzer, M., Hackl, C., Kopal, G., Hufnagl, P., Hofer, K., Polin, H., Stabentheiner, S., Proll, J. (2009). Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. *Hum. Immunol.* 70, 960-964.
- 43) Holcomb, C. L., Hoggland, B., Anderson, M. W., Blake, L. A., Bohme, I., Egholm, M., Ferriola, D., Gabriel, C., Gelber, S. E., Goodridge, D., Hawbecker, S., Klein, R., Ladner, M., Lind, C., Monos, D., Pando, M. J., Proll, J., Sayer, D. C., Schmitz-Agheguian, G., Simen, B. B. et al. (2011). A multi-site study using high-resolution HLA genotyping by next generation sequencing. *Tissue Antigens.* 77, 206-217.
- 44) Hosomichi, K., Jinam, T. A., Mitsunaga, S., Nakaoka, H., Inoue, I. (2013). Phase-defined complete sequencing of the HLA genes by next-generation sequencing. *BMC Genomics.* 14, 355
- 45) Cao, H., Wang, Y., Zhang, W., Chai, X., Zhang, X., Chen, S., Yang, F., Zhang, C., Guo, Y., Liu, Y., Tang, Z., Chen, C., Xue, Y., Zhen, H., Xu, Y., Rao, B., Liu, T., Zhao M., Zhang, W., Li, Y. et al. (2013). A Short-Read Multiplex Sequencing Method for Reliable, Cost-Effective and High-Throughput Genotyping in Large-Scale Studies. *Hum. Mutat.* 34, 1715-1720.
- 46) Shiina, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari, A., Oka, A., Umemura, T., Joshita, S., Takahashi, O., Hayashi, Y., Paumen, M., Katsuyama, Y., Mitsunaga, S., Ota, M., Kulski, J. K., Inoko, H. (2012). Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens.* 80, 305-316.
- 47) Suzuki, S., Ozaki, Y., Masuya, A., Shigenari, A., Mitsunaga, S., Inoko, H., Shiina, T. (2015). Development of a new HLA allele assignment program, Sequence Alignment Based Assigning Software (SeaBass). *Major Histocompatibility Complex* 22, 102-107.
- 48) Erlich, R. L., Jia, X., Anderson, S., Banks, E., Gao, X., Carrington, M., Gupta, N., DePristo, M. A., Henn, M. R., Lennon, N. J., de Bakker, P. I. (2011). Next-generation

- sequencing for HLA typing of class I loci. *BMC Genomics*. 12, 42.
- 49) Danzer, M., Niklas, N., Stabentheiner, S., Hofer, K., Pröll, J., Stücker, C., Raml, E., Polin, H., Gabriel, C. (2013). Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. *BMC Genomics*. 14, 221.
- 50) Lank. S. M., Golbach, B. A., Creager, H. M., Wiseman, R. W., Keskin, D. B., Reinherz, E. L., Brusic, V., O'Connor, D. H. (2012). Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing. *BMC Genomics*. 13, 378.
- 51) Ozaki, Y., Suzuki, S., Kashiwase, K., Shigenari, A., Okudaira, Y., Ito, S., Masuya, A., Azuma, F., Yabe, T., Morishima, S., Mitsunaga, S., Satake, M., Ota, M., Morishima, Y., Kulski, J. K., Saito, K., Inoko, H., Shiina, T (2015). Cost-efficient multiplex PCR for routine genotyping of up to nine classical HLA loci in a single analytical run of multiple samples by next generation sequencing. *BMC Genomics*. 16, 318.
- 52) Nelson, W. C., Pyo, C. W., Vogan, D., Wang, R., Pyon, Y. S., Hennessey, C., Smith, A., Pereira, S., Ishitani, A., Geraghty, D. E. (2015). An integrated genotyping approach for HLA and other complex genetic systems. *Hum. Immunol.* 76, 928-38.
- 53) Lind, C., Ferriola, D., Mackiewicz, K., Heron, S., Rogers, M., Slavich, L., Walker, R., Hsiao, T., McLaughlin, L., D'Arcy, M., Gai, X., Goodridge, D., Sayer, D., Monos, D. (2010). Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Hum. Immunol.* 71, 1033-42.
- 54) Ozaki, Y., Suzuki, S., Shigenari, A., Okudaira, Y., Kikkawa, E., Oka, A., Ota, M., Mitsunaga, S., Kulski, J. K., Inoko, H., Shiina, T. (2013). HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput sequencing. *Tissue Antigens*. 83, 10-16.
- 55) Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biology*. 14, R51.
- 56) Barnard, R., Futo, V., Pecheniuk, N., Slattery, M., Walsh, T. (1998). PCR bias toward the wild-type k-ras and p53 sequences: implications for PCR detection of mutations and cancer diagnosis. *Biotechniques*. 25, 684-691.
- 57) Drake, C. G., Lipson, E. J., Brahmer, J. R. (2014). Breathing new life into immunotherapy: review of melanoma, lung and kidney cancer. *Nat. Rev. Clin. Oncol.* 11, 24-37.
- 58) Beatty, G. L., Gladney, W. L. (2015). Immune escape mechanisms as a guide for cancer immunotherapy. *Clin. Cancer Res.* 21, 687-692.
- 59) Ott, P. A., Hodi, F. S., Robert, C. (2013). CTLA-4 and PD-1/PD-L1 blockade: new immunotherapeutic modalities with durable clinical benefit in melanoma patients. *Clin.*

- Cancer Res. 19, 5300-5309.
- 60) Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer.* 12, 252-264.
 - 61) Johnson, D. B., Estrada, M. V., Salgado, R., Sanchez, V., Doxie, D. B., Opalenik, S. R., Vilgelm, A. E., Feld, E., Johnson, A. S., Greenplate, A. R., Sanders, M.E., Lovly, C. M., Frederick, D. T., Kelley, M. C., Richmond, A., Irish, J. M., Shyr, Y., Sullivan, R. J., Puzanov, I., Sosman, J. A., Balko, J. M. (2016). Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat. Commun.* 7, 10582.
 - 62) Thor, S. P., Garrido, F. (2016). Targetless T cells in cancer immunotherapy. *J. Immunother. Cancer.* 4, 23.
 - 63) Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797.
 - 64) Lank, S. M., Wiseman, R. W., Dudley, D. M., O'Connor, D. H. (2010). A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping. *Hum. Immunol.* 71, 1011-1017.
 - 65) Hong, L. Z., Hong, S., Wong, H. T., Pauline, P. A., Cheng, Y., Wilm, A., de Sessions, P. F., Lim, S. G., Nagarajan, N., Hibberd, M. L., Quake, S. R., Burkholder, W. F. (2014). BAsE-Seq: a method for obtaining long viral haplotypes from short sequence reads. *Genome Biol.* 15, 517.
 - 66) Stapleton, J. A., Kim, J., Hamilton, J. P., Wu, M., Irber, L. C., Maddamsetti, R., Briney, B., Newton, L., Burton, D. R., Brown, C. T., Chan, C., Buell, C. R., Whitehead, T. A. (2016). Haplotype-Phased Synthetic Long Reads from Short-Read Sequencing. *PLoS One.* 11, e0147229.
 - 67) Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25, 2078-2079.