

HLA遺伝子完全配列決定 パイプラインの構築

国立遺伝学研究所 総合遺伝研究系
人類遺伝研究部門
細道一善

2015年3月28日(土)
平成26年度 総合データ解析トライアル 研究終了報告会

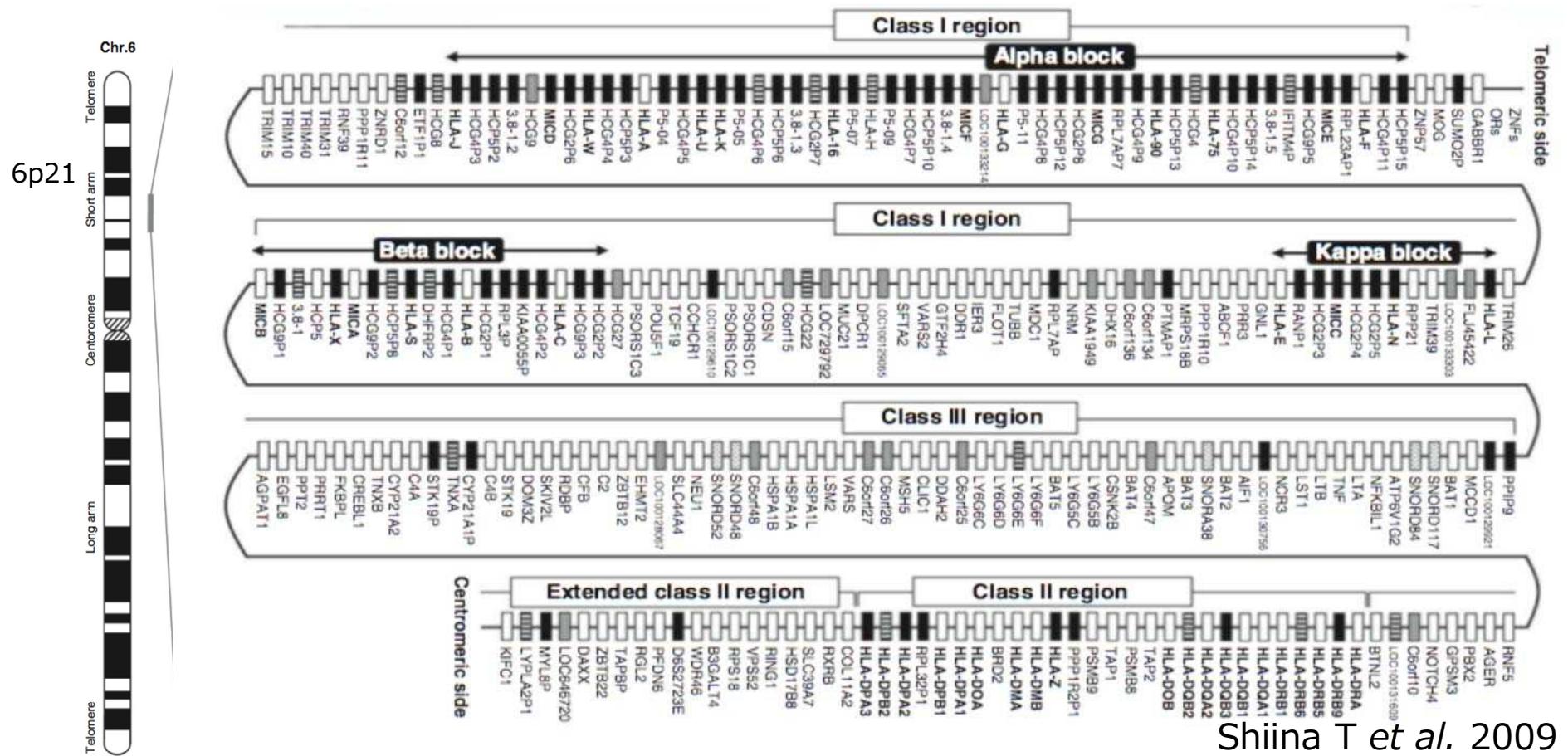


©2015 細道一善(国立遺伝学研究所) licensed under CC表示2.1日本

Topics

1. 研究開発の目的・新たな活用法
2. 研究開発の説明
3. Webツールとしての公開
4. 研究開発を活用した有用な知識の発見
5. 今後の本研究開発の将来性

HLA領域のゲノム配列



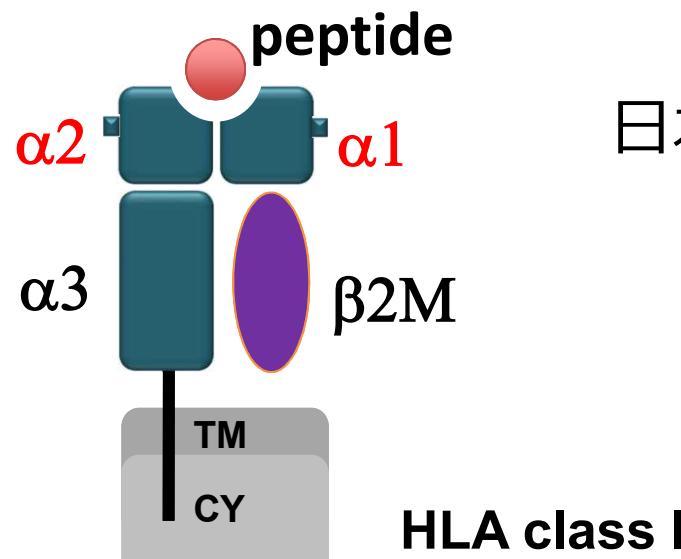
HLA領域の特徴

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

極めて高度な多型性を示す

移植のみならず100以上の疾患と関連する

HLA クラスI α1およびα2ドメインの 多様性



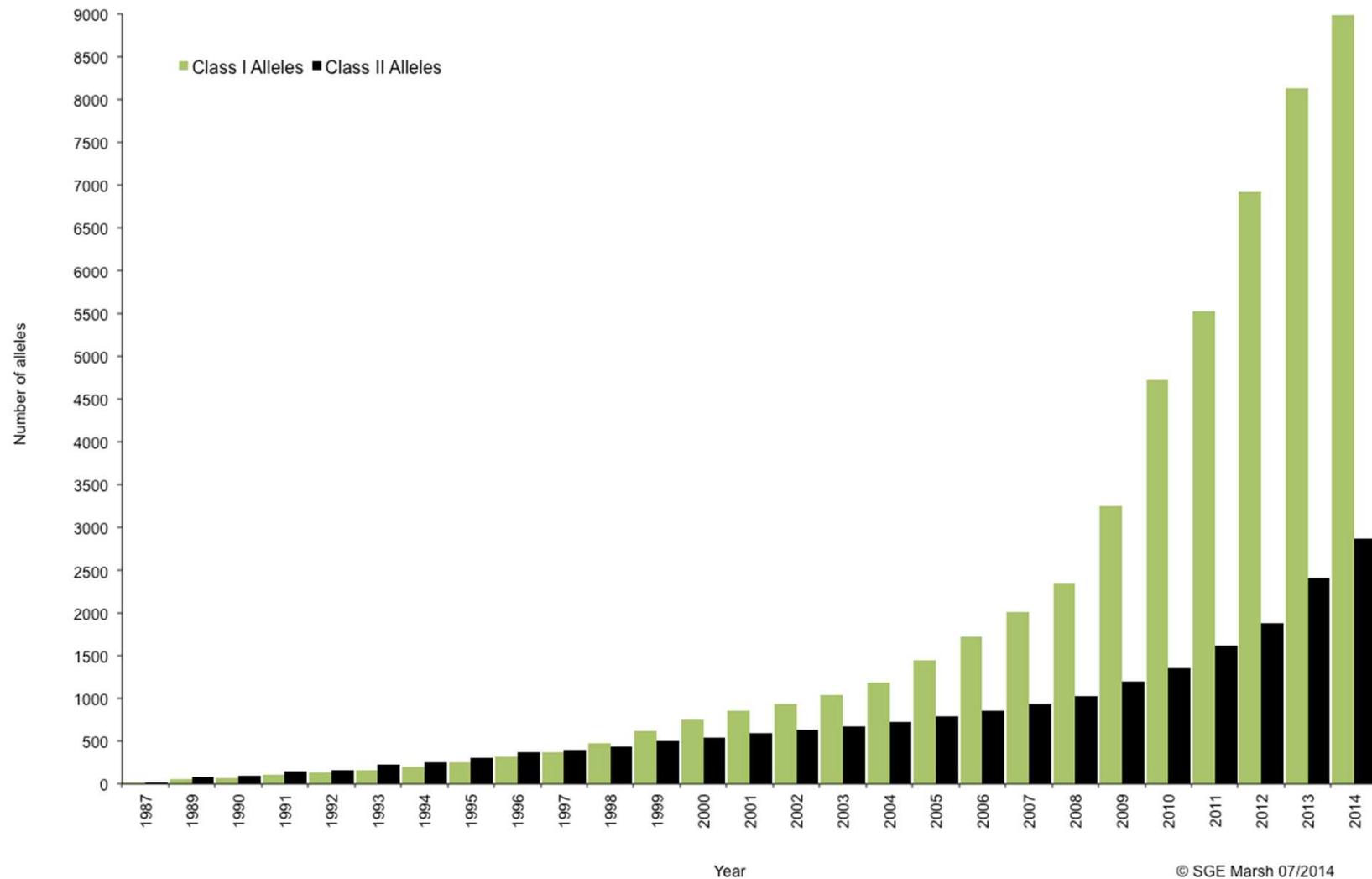
日本人で頻度の高いHLA-B アレル5種間

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

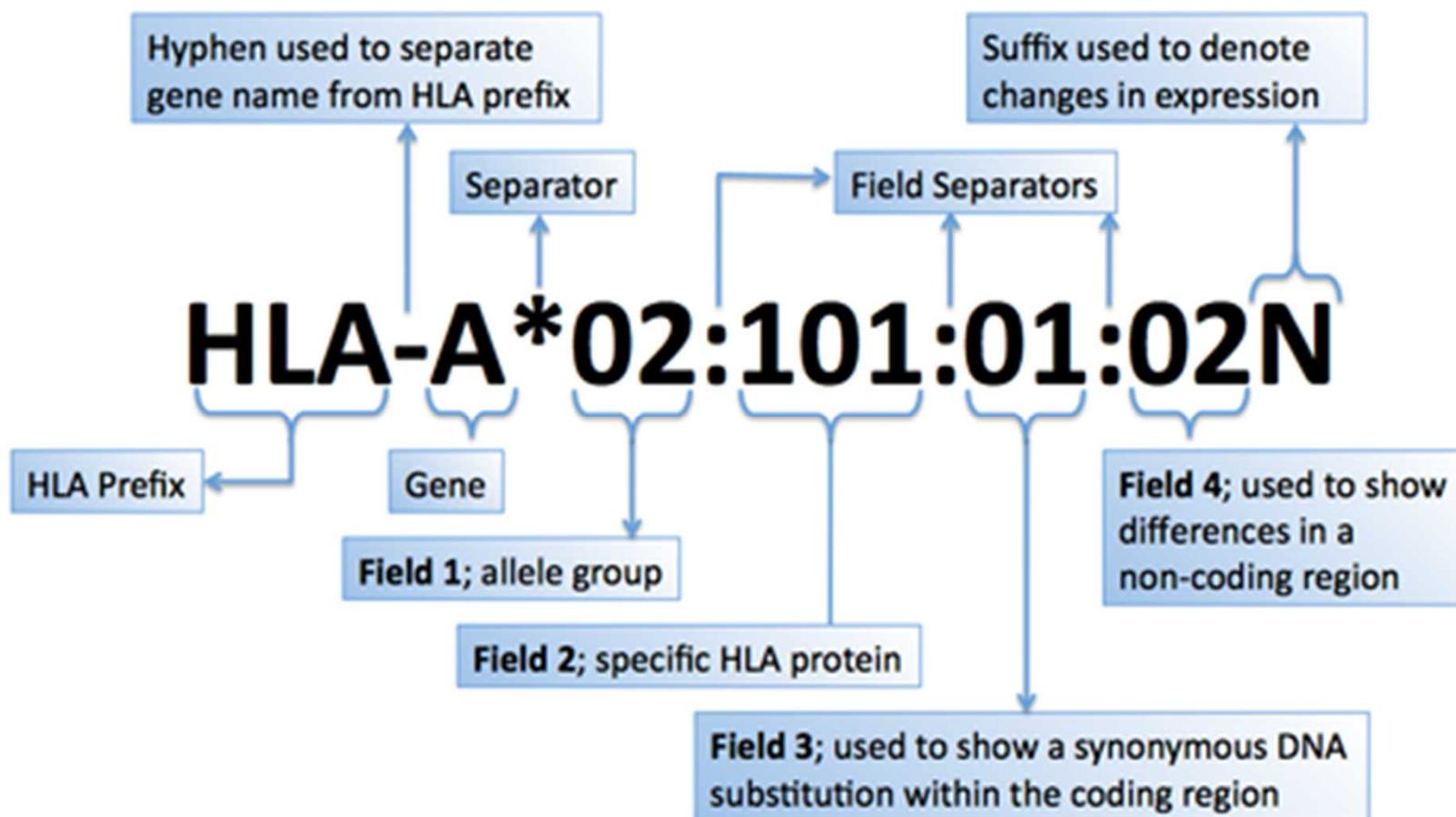
IMGT/HLAデータベースに登録されて いるHLAアレル数

Numbers of HLA Alleles			
HLA Class I Alleles		8,976	
HLA Class II Alleles		2,870	
HLA Alleles		11,846	
HLA Class I			
Gene	A	B	C
Alleles	2,884	3,589	2,375
Proteins	2,041	2,668	1,677
Nulls	133	119	71
HLA Class II			
Gene	DRB	DQB1	DPB1
Alleles	1,642	664	422
Proteins	1,211	435	351
Nulls	37	16	10

IMGT/HLAデータベースに登録されている HLAアレル数



HLAアレルの命名法

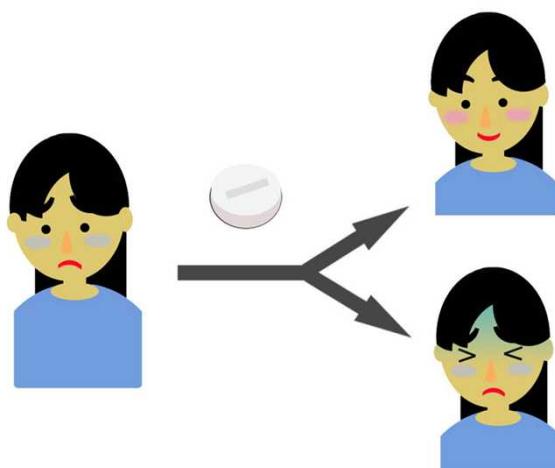


HLAと薬剤副作用との関連性

薬剤副作用	関連を示すHLA型	陽性率(%)	オッズ比
Tiopronin(重金属)と肝内胆汁うつ滞	<i>HLA-A*33:03</i>	93	41.5
Carbamazepine(抗痙攣剤)と Stevens-Johnson 症候群	<i>HLA-B*15:02</i>	100	895.5
Abacavir(抗HIV剤)と胃腸障害、嗜眠、低血圧による致死副作用	<i>HLA-B*57:01</i>	78	117.5
Allopurinol(抗痛風、抗尿酸血症剤)と 薬疹	<i>HLA-B*58:01</i>	100	393.5
Ticlopidine(抗血小板剤)と肝障害	<i>HLA-A*33:03</i>	86	36.5
Amoxicillin-clavulanate potassium (抗生素質)と肝障害	<i>HLA-DRB1*15:01</i>	57	35.6
Flucloxacillin(抗生素質)と肝障害	<i>HLA-B*57:01</i>	84	80.6

原因がHLA遺伝子そのものであるか、連鎖不平衡によりHLA遺伝子と関連しているようにみえるかはAbacavirを除き不明

HLAタイピングによる予防的診断



	HLA-B*57:01 positive	HLA-B*57:01 negative
ADR negative	25	794
ADR positive	23	0
ADR Positive %	48%	0%

Mallal S et al. N Engl J Med. 2008

対象データおよび対象データベース

The screenshot shows the homepage of the HLA Database. The header features the "HLA DATABASE" logo with a green leafy background. Below the header is a navigation bar with links: Human Variation DB, HLA Database (which is highlighted in blue), SNP Control, Case Control GWAS, CNV Database, and CNV Association.

The main content area has a green header bar labeled "About HLA Database". The text explains that the HLA database is a repository system for HLA data, containing haplotype data and disease-related information. It expresses appreciation for HLA data submission and notes support from the Ministry of Education, Culture, Sports, Science and Technology.

Below this is a "SEARCH & BROWSE" section with a green header bar labeled "Study info". It includes a "Multiple alignment Japanese Haplotype" section with dropdown menus for "Select multiple studies" and "Select reference study", and a "Base level view" button set to "HLA-A". There are "Submit" and "overview" buttons. A "Pair-alignment" section follows, with dropdown menus for "First" (set to "cox") and "Second" (set to "cox"), both with "Submit" buttons. A "Multi-alignments" section is also present.

The left sidebar contains a vertical menu with sections: "About This Database", "HELP | FAQ", "DATABASE", "LINK", and "CONTACT". Under "DATABASE", there are links to dbGAP, GeMDBJ, JSNP, HAPMAP, dbSNP, and HGVbase. Under "LINK", there are links to HGVRD top-page, HGVRD data sharing policy, NBDC, DBCLS, University of Tokyo, National Institute of Genetics, University of Tokyo Hospital, and CRL, Hitachi, LTD.

統合化推進プログラムのデータベースの 新たな活用法

HLA DATABASE

Human Variation DB HLA Database SNP Control Case Control GWAS CNV Database CNV Association

About HLA Database

This HLA database (HLA DB) is a repository system and has been constructed to achieve permanent data management and information sharing of HLA data. HLA-DB contains HLA haplotype data and disease-related information. We greatly appreciate your HLA data submission.

This work has been supported by the Ministry of Education, Culture, Sports, Science and Technology.

SEARCH & BROWSE

Study info

Sequences

- Multiple alignment Japanese Haplotype
 - Select multiple studies
 - Select reference study
 - Base level view: HLA-A overview Submit
- Pair-alignment
 - First: cox Second: cox Submit
- Multiple alignment Japanese Haplotype
 - Select multiple studies

日本人HLA遺伝子情報の追加
データベースの拡充

HLA database 登録配列

Immunogenetics (2008) 60:1–18
DOI 10.1007/s00251-007-0262-2

 OpenAccess

ORIGINAL PAPER

Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project

Table 1 Haplotype sequence contig length, number of gaps and HLA allele types

Haplotype	Length (bp)	Gaps	HLA-A	HLA-B	HLA-C	HLA-DQA1	HLA-DQB1	HLA-DRB1
PGF	4754829	0	A*03010101	B*070201	Cw*07020103	DQA1*010201	DQB1*0602	DRB1*150101
COX	4731878	0	A*01010101	B*080101	Cw*070101	DQA1*050101	DQB1*020101	DRB1*030101
QBL	4249272	5	A*260101	B*180101	Cw*050101	DQA1*050101	DQB1*020101	DRB1*030101
APD	4160965	16	A*01010101	—	—	—	—	—
DBB	2330101	28	A*02010101	—	Cw*06020101	DQA1*0201	DQB1*030302	DRB1*070101
MANN	4191014	10	A*290201	B*440301	Cw*160101	DQA1*0201	DQB1*0202	DRB1*070101
MCF	4087413	15	[A*020101]	B*15010101	Cw*030401	DQA1*0303	DQB1*030101	—
SSTO	3704249	22	A*320101	B*44020101	Cw*050101	DQA1*030101	DQB1*030501	DRB1*040301

MHC Haplotype Consortium Web Resource



Home Research Scientific resources Work & study About us

What we do History How we work People Press Public engagement Campus Contact

20th March 2002 - updated 12 August 2010

MHC Haplotype Consortium Web Resource

A Cambridge-based Consortium today announced a new web resource designed to help medical researchers in the fight against diseases like diabetes, rheumatoid arthritis and multiple sclerosis.

The [MHC Haplotype Consortium](#) - formed by researchers at the [Wellcome Trust Sanger Institute](#) and the [University of Cambridge](#) - will define the most common differences in DNA code that can lead to these diseases. An accurate molecular understanding should lead to improved diagnosis and treatment.

Sir Alec Jeffreys, Professor of Genetics at Leicester University, said, *"This exciting new development will for the first time allow us to see the true patterns of genetic diversity in an extremely important gene complex that lies at the heart of the immune system. It will not only impact substantially on our understanding of the role of the MHC in disorders such as autoimmune disease and in combating infectious disease, but will reveal fundamental new information on processes such as natural selection that shape human DNA diversity."*

The warriors in our molecular defence against attack by microbes - the components of our immune system - are usually a highly organized and well-disciplined force. Mobile and versatile, new and improved defences are deployed to respond to each invasion.

But sometimes this force receives the wrong orders or misinterprets the correct orders. In a misguided move, as well as fighting off the threats from outside, they attack our own cells, destroying as they go. This treasonous act is autoimmune disease, which underlies a spectrum of disorders from multiple sclerosis to diabetes.

The new website will catalogue the changes in DNA code that lead to confusion and rebellion in the immune system. The DNA differences will be crucial in fighting autoimmune disease, some of which are increasing in the UK and around the world.

The key region of the genome is called the MHC - major histocompatibility complex - a cluster of more than 200 genes on chromosome 6. A principal function of the MHC is to discriminate between 'self' - our own cells and tissues - and 'non-self' - the invaders we would hope to destroy.

The MHC is the most variable region of our genome, driven by the need to ward off the ever-changing pathogens that invade us and so cataloguing the changes related to autoimmune disease is especially difficult. The consortium will analyse the DNA sequence from eight individuals specially chosen for the project.

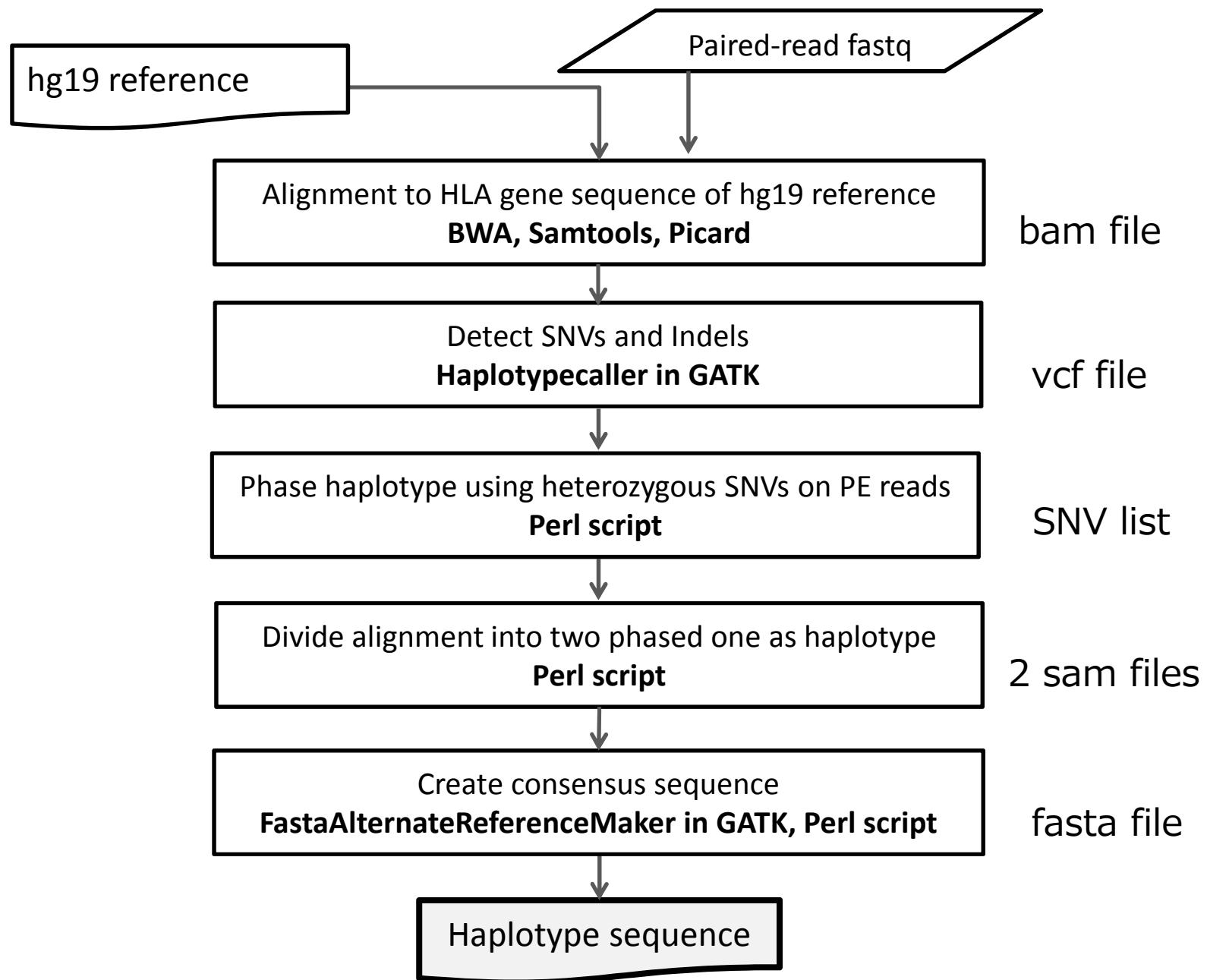
"As well as being an ethnic mix, these individuals represent the most common combinations of genes associated with certain diseases and so we will identify the most important sequence variants that predispose to these conditions," said Stephan Beck, Head of Human Sequencing at the Wellcome Trust Sanger Institute.

Already two genomes have been finished and are displayed on the site. Data from the others is being added as soon as it is obtained. The project, which has received £ 2.2 million in funding from the Wellcome Trust, is projected to continue until 2004. Members of the Haplotype Consortium share the view that prepublication release of these data will help speed up the progress of research into autoimmune disease.

Topics

1. 研究開発の目的・新たな活用法
2. 研究開発の説明
3. Webツールとしての公開
4. 研究開発を活用した有用な知識の発見
5. 今後の本研究開発の将来性

Workflow of phase-defined sequencing



NBDC Research ID: hum0003.v1

研究内容の概要

目的： HLA遺伝子(HLA-A, -C, -B, -DRB1, -DQB1および-DPB1)におけるHLAアレル完全長塩基配列(約3kb～14kb)を決定する

方法： Long range PCRにより目的領域を增幅し、 Illumina社のMiSeqにより塩基配列を解析する

対象： 複数集団由来セルライン：33名

データID	内容	制限	公開日
DRA000908	NGS (HLAアレルの塩基配列決定)	オープン	2013/07/01

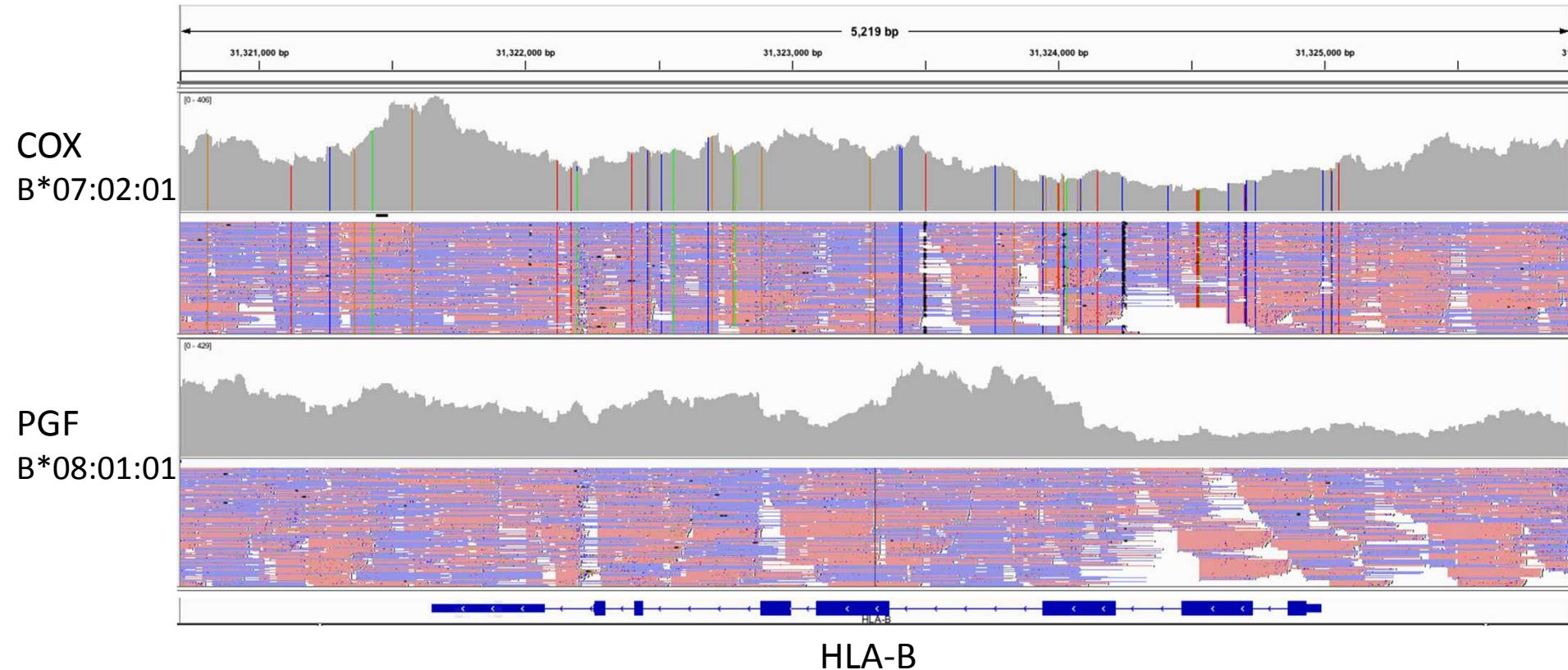
*リリース情報は [こちら](#)

分子データ

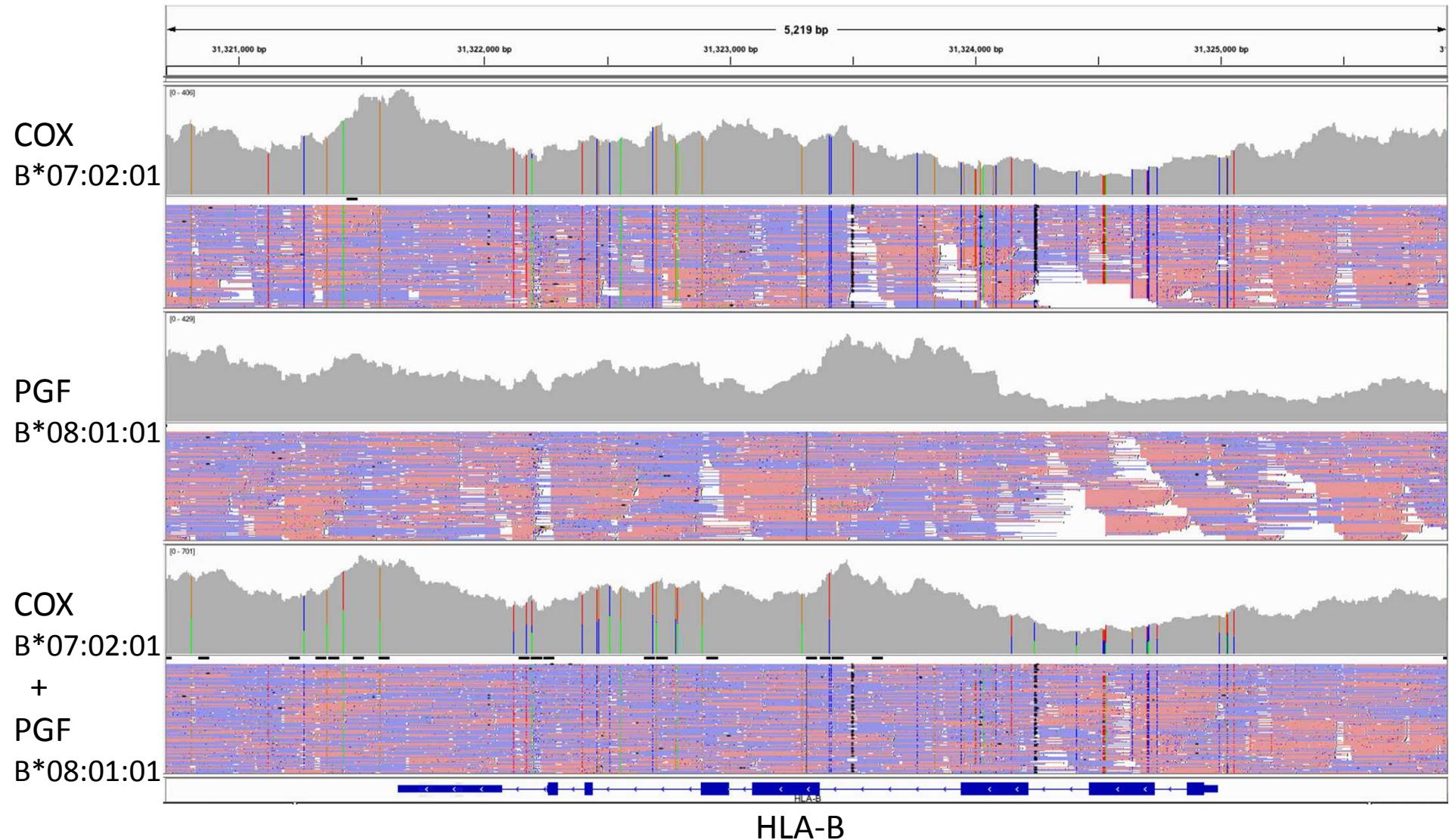
DRA000908

対象	セルライン33名分
規模	Target Capture
対象領域 (Target Captureの場合)	HLA領域

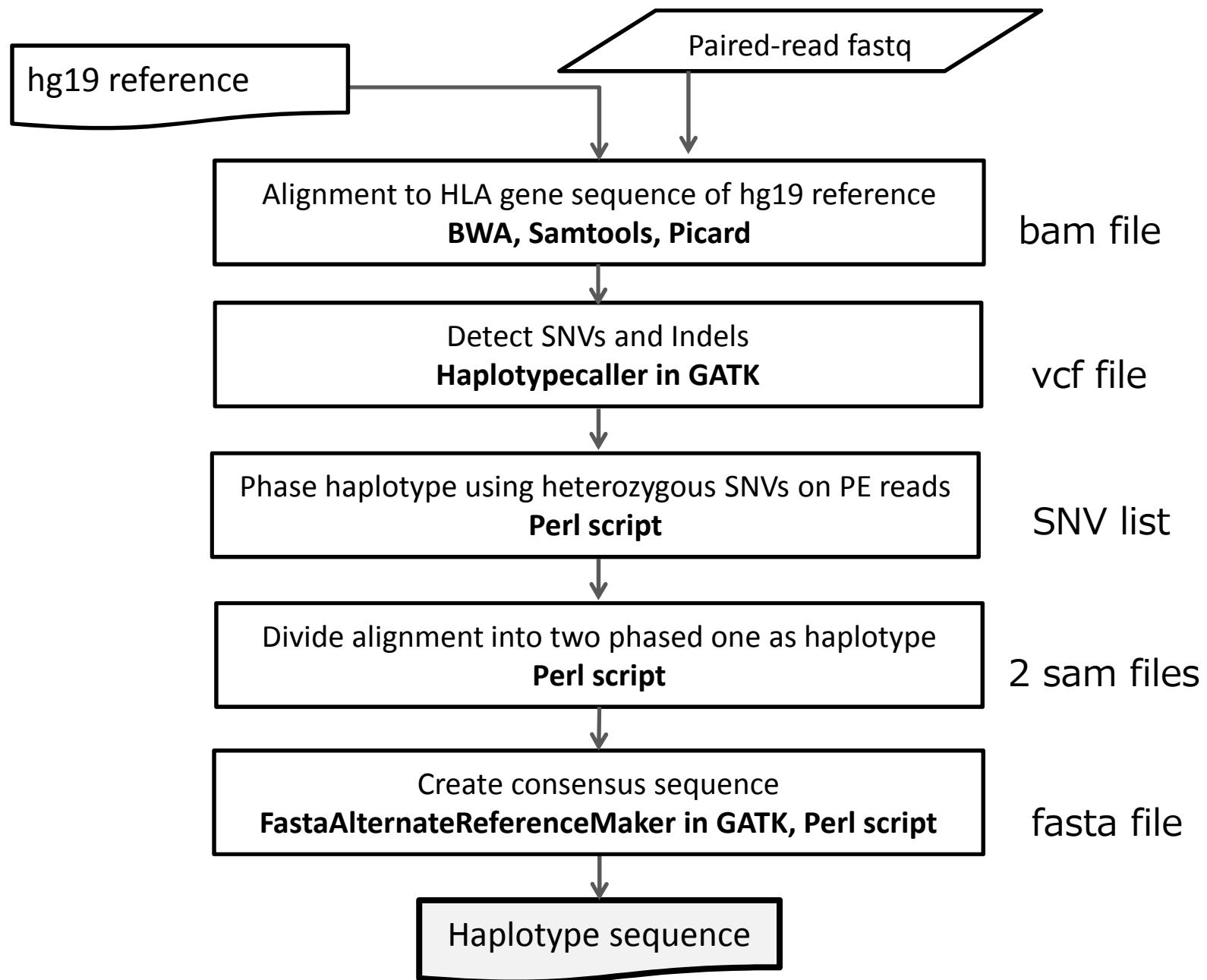
セルラインにおけるHLA-BのNGSデータ



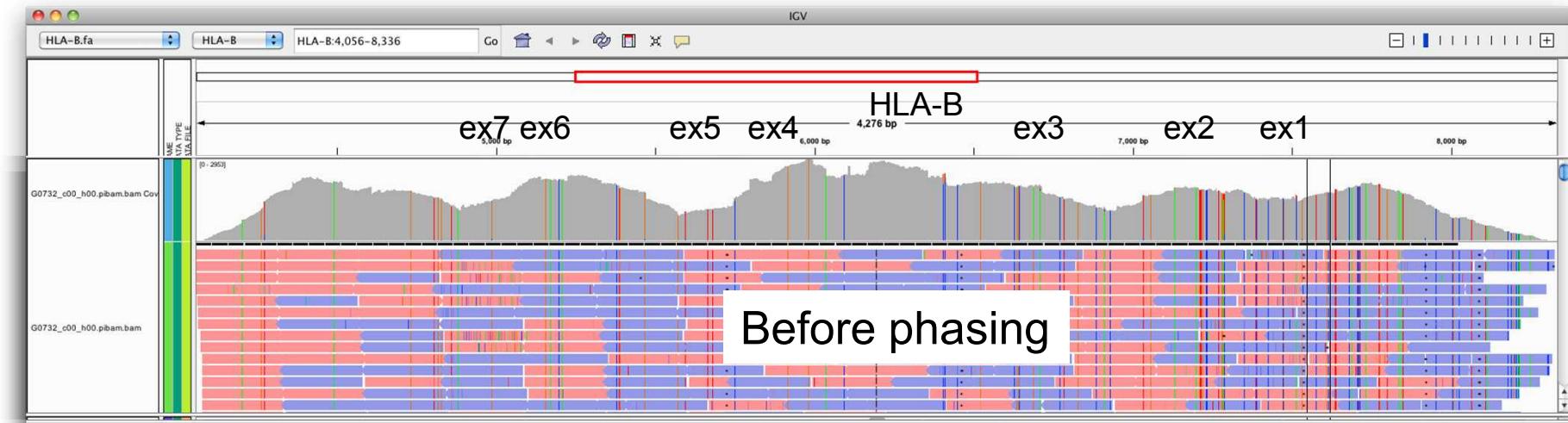
セルラインにおけるHLA-BのNGSデータ



Workflow of phase-defined sequencing



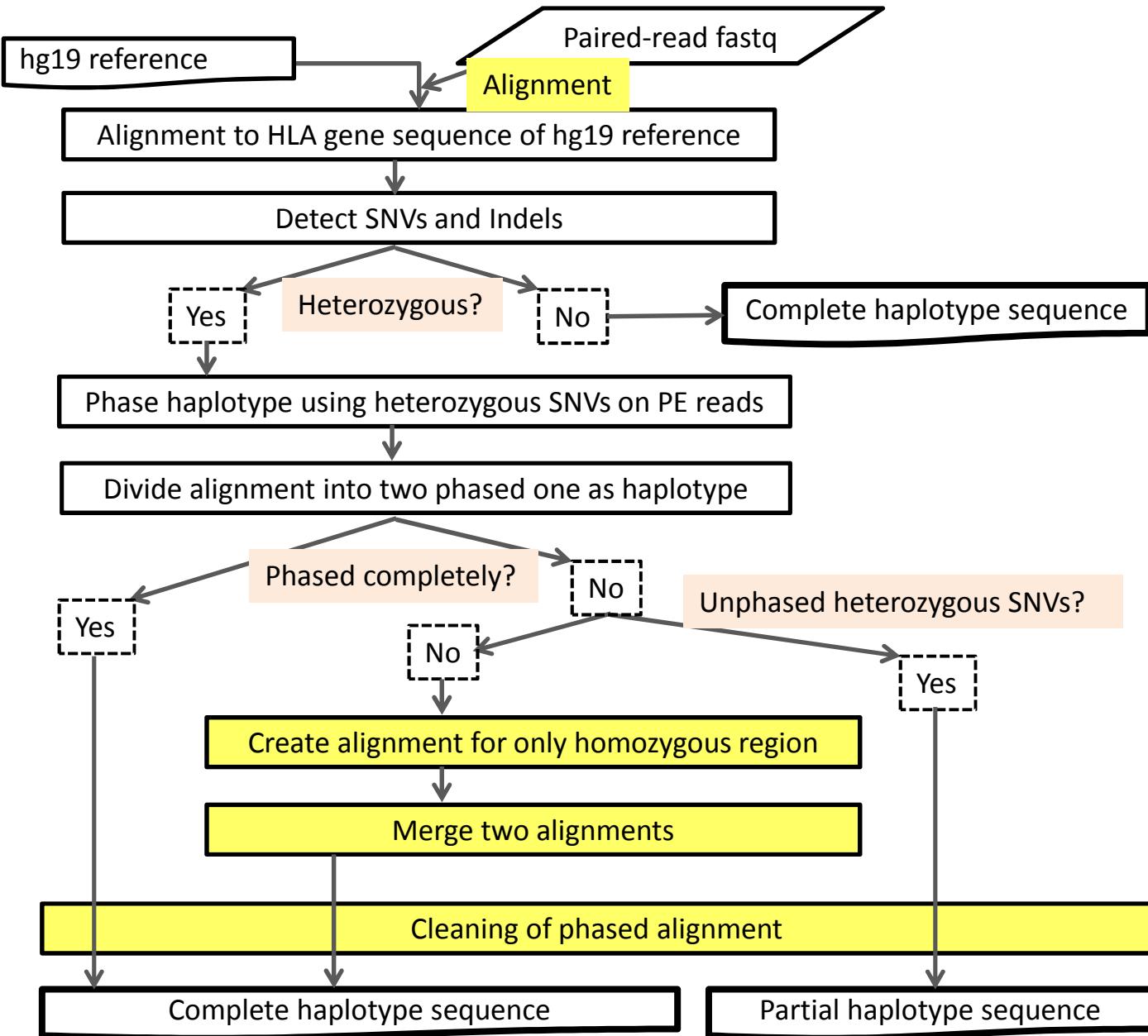
HLA遺伝子配列完全決定のワークフロー

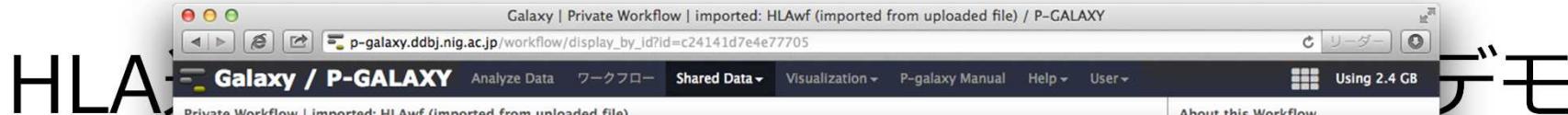


Phased haplotype 1 → $B^*51:01:01$

Phased haplotype 2 → $B^*15:02:01$

実データ解析へのパイプライン最適化





Galaxy / P-GALaxy

Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User

ツール search tools

Work Flow

Get Data Send Data ENCODE Tools Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Multiple Statistics

Metagenomics

FASTA manipulation

NGS: QC and Analysis

NGS: Mapping

NGS: Indel Analysis

psLayout version 3

match	mis-	rep.	N's	Q gap	Q gap	T gap	T gap	strand	Q	Q	Q	T	T	T	block	blockSizes	qStarts	tStarts		
1082	7	0	0	0	0	0	0	-	B*40:02:01_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
814	8	0	0	0	0	0	0	-	B*40:02:02_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
814	8	0	0	0	0	0	0	-	B*40:02:03_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
882	8	0	0	0	0	0	0	-	B*40:02:05_890_bp	890	0	890	dataset_11339.dat	1089	158	1048	1	890,	0,	158,
816	6	0	0	0	0	0	0	-	B*40:02:09_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
887	8	0	0	0	0	0	0	-	B*40:02:11_895_bp	895	0	895	dataset_11339.dat	1089	194	1089	1	895,	0,	194,
814	8	0	0	0	0	0	0	-	B*40:02:12_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
1080	9	0	0	0	0	0	0	-	B*40:03_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
1082	7	0	0	0	0	0	0	-	B*40:04_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
1079	10	0	0	0	0	0	0	-	B*40:05_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
1089	0	0	0	0	0	0	0	-	B*40:06:01_01_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
1089	0	0	0	0	0	0	0	-	B*40:06:01:02_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
821	1	0	0	0	0	0	0	-	B*40:06:02_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
894	1	0	0	0	0	0	0	-	B*40:06:03_895_bp	895	0	895	dataset_11339.dat	1089	194	1089	1	895,	0,	194,
894	1	0	0	0	0	0	0	-	B*40:06:04_895_bp	895	0	895	dataset_11339.dat	1089	194	1089	1	895,	0,	194,
544	2	0	0	0	0	0	0	-	B*40:06:05_546_bp	546	0	546	dataset_11339.dat	1089	470	1016	1	546,	0,	470,
821	1	0	0	0	0	0	0	-	B*40:06:06_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
821	1	0	0	0	0	0	0	-	B*40:06:07_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
1079	10	0	0	0	0	0	0	-	B*40:08_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
541	5	0	0	0	0	0	0	-	B*40:109_546_bp	546	0	546	dataset_11339.dat	1089	470	1016	1	546,	0,	470,
816	6	0	0	0	0	0	0	-	B*40:11:01_822_bp	822	0	822	dataset_11339.dat	1089	194	1016	1	822,	0,	194,
541	5	0	0	0	0	0	0	-	B*40:11:02_546_bp	546	0	546	dataset_11339.dat	1089	470	1016	1	546,	0,	470,

Galaxy / P-GALaxy

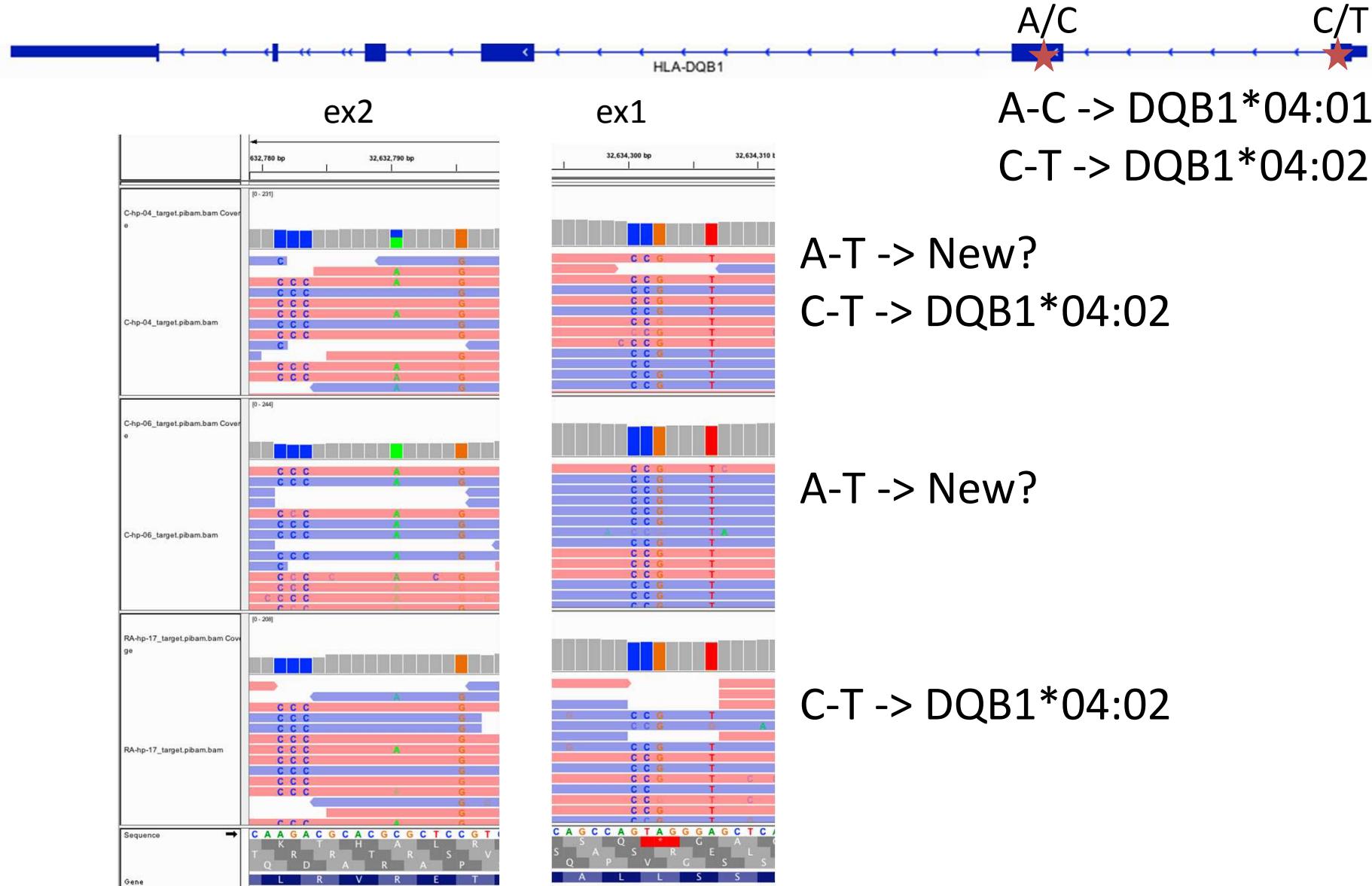
Analyze Data ワークフロー Shared Data Visualization P-galaxy Manual Help User

ツール search tools

Work Flow Get Data Send Data ENCODE Tools Lift-Over Text Manipulation

match	mis-	rep.	N's	Q gap	Q gap	T gap	T gap	strand	Q	Q	Q	T	T	T	block	blockSizes	qStarts	tStarts		
1089	0	0	0	0	0	0	0	-	B*40:06:01:01_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,
1089	0	0	0	0	0	0	0	-	B*40:06:01:02_1089_bp	1089	0	1089	dataset_11339.dat	1089	0	1089	1	1089,	0,	0,

研究開発を行ったパイプライン を活用した有用な知識の発見



研究開発を行ったパイプライン を活用した有用な知識の発見

EMBL-EBI

Services Research Training About us

IPD - IMGT/HLA

KIR

Overview IMGT/HLA KIR MHC HPA ESTDAB Contact Support

IPD > IMGT/HLA

IMGT/HLA

Release 3.19.0, 2015-01-19

The IMGT/HLA Database provides a specialist database for sequences of the human major histocompatibility complex (HLA) and includes the official sequences for the WHO Nomenclature Committee For Factors of the HLA System. The IMGT/HLA Database is part of the international ImMunoGeneTics project (IMGT).

The database uses the 2010 nomenclature designations in all tools. To aid in the adoption of the new nomenclature, all search tools can be used with both the current and pre-2010 allele designations. The pre-2010 nomenclature designations are only used where older reports or outputs have been made available to download.

Latest Developments

- [HLA-DPB1 T-Cell Epitope Algorithm](#)
- [What's new in the latest release](#)

Latest Publications

- Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P, Marsh SGE
The IPD and IMGT/HLA database: allele variant databases
Nucleic Acids Research (2015) **43**:D423-431

Sponsors

The IMGT/HLA Database is sponsored by a number of institutes and companies, for further details of all our supporters and how you can help please see the [funding page](#).

IMGT/HLA

- [About IMGT/HLA](#)
- [Access](#)
 - [Alignments](#)
 - [Alleles](#)
 - [SBT - Ambiguous Alleles](#)
 - [BLAST Searches](#)
 - [Cells](#)
 - [FTP Directory](#)
 - [HLA Dictionary](#)
 - [More Tools](#)
 - [DPB TCE Tool](#)
 - [Search Determinants](#)
 - [Statistics](#)
- [FAQ](#)
- [Links](#)
- [Publications](#)
- [Nomenclature](#)
- [Release Information](#)
- [Submissions](#)

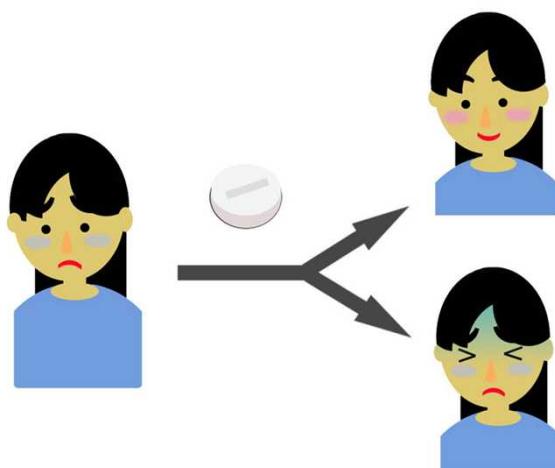
DQB1*04:01 の完全長塩基配列

1 GCAATTCT CTCCCCTGAA ATATGATCCC CACTTAATT GCCCTACTGA AAGAATCCCA
61 AGTATAAAAA CAACCAGTT TTAATCAATA TTACAAAGAT GTTTACTGTT GAATCGCATT
121 TTTCTTGCG TTCTTAAAAT CCCTTAGGCA TTCAATCTTC AGCTCTTCTA TAATTGAGAG
181 GAAGTTTCA CCTCAAATGT TCATCCAGTG CAATTGAAG ACGTCACAGT GCCAGGGACT
241 GAATTGAGAA CCTTCACAAA AAAAAATGTC TGCCTAGAGA CAGATTAGGT CCTTCAGCTC
301 CAGTGCTGAT TGGTTCCCTT CCAAAGGACC ATCCAATCCT GCCACGCAGG GAAACATCCA
361 CAGGTTTTA TTCTTCTGC CAGGTACATC AGATCCATCA GGTCCAAGCT GTGTTGACTA
421 CCACTACTTT TCCCTTCGTC TCAATTATGT CTTGGAAGAA GGCTTGCGG ATCCCTGGAG
481 GCCTTCGGGT AGCAACTGTG ACCTTGATGC TGGCGATGCT GAGCACCCCCG GTGGCTGAGG
541 GCAGAGACTC TCCCAGTAAG TGCAGGGCCA CTGCTCTCCA GAGCCGCCAC TCTGGGAACA
601 GGCTCTCCTT GGGCTGGGGT AGGGGGATGG TGATCTCCAT GATCTCGGAC ACAATCTTC

6301 CAATGGTCTC TGTTCATGGT ATATTTGCTG CTATGAGGAT CAAGACTTAG GGTCGAAGTT
6361 TGCCAGTTTC TAGGAATCTC CAGAGGTTGT TCCCCAGAAC CAAGCCTTAA CTTTGGTGGT
6421 ATCTTCTTGT GAAATGTGAA GCCAGAACCA CAGCTAAAT GTTAGACAAG AGGATGATGC
6481 CCACTTGTG CCACATGTTG GTGGCTACTG CCTGTAGGCA TTTCCAGTG ACTGAAAGAG
6541 GCTGCTAGTG GTAGGGATGA GGTATCATCC AATTTCTAA AAAGATTGAA CCCTTCATAT
6601 TCCCCAGAAC AGTAACAGCT GTTCCGCCAC TTCCCACATA TCTGCATCAA GCTGAAGTTC
6661 TGTGTCCTCA CGAGCTGATT TCACCTTGC ACAGATCTG CGGGAGGTGA CAATAATACA
6721 TTCTGGACCT CAGCTTCTC TGTCTGAAGC TGCAGGGGGC CCCTGAGGGG TGGGGGAGAT
6781 TGCAGGCCCA CCAGCGTACC CTGTGCTGAT CATCCCTCTT CTCTCTTCTT CAGGGCTCCT
6841 GCACTGACTC CTGAGACTAT TTTAACTGGG ATTGGTTATC ACTTTCTGT AACGCCTGCT

IMGT/HLAに登録されているDQB1*04:01 の配列とは
エクソン1に1塩基の違いが認められた

今後の本研究開発の将来性



	HLA-B*57:01 positive	HLA-B*57:01 negative
ADR negative	25	794
ADR positive	23	0
ADR Positive %	48%	0%

Mallal S et al. N Engl J Med. 2008

個の医療