HLA遺伝子完全配列決定パイプラインの構築

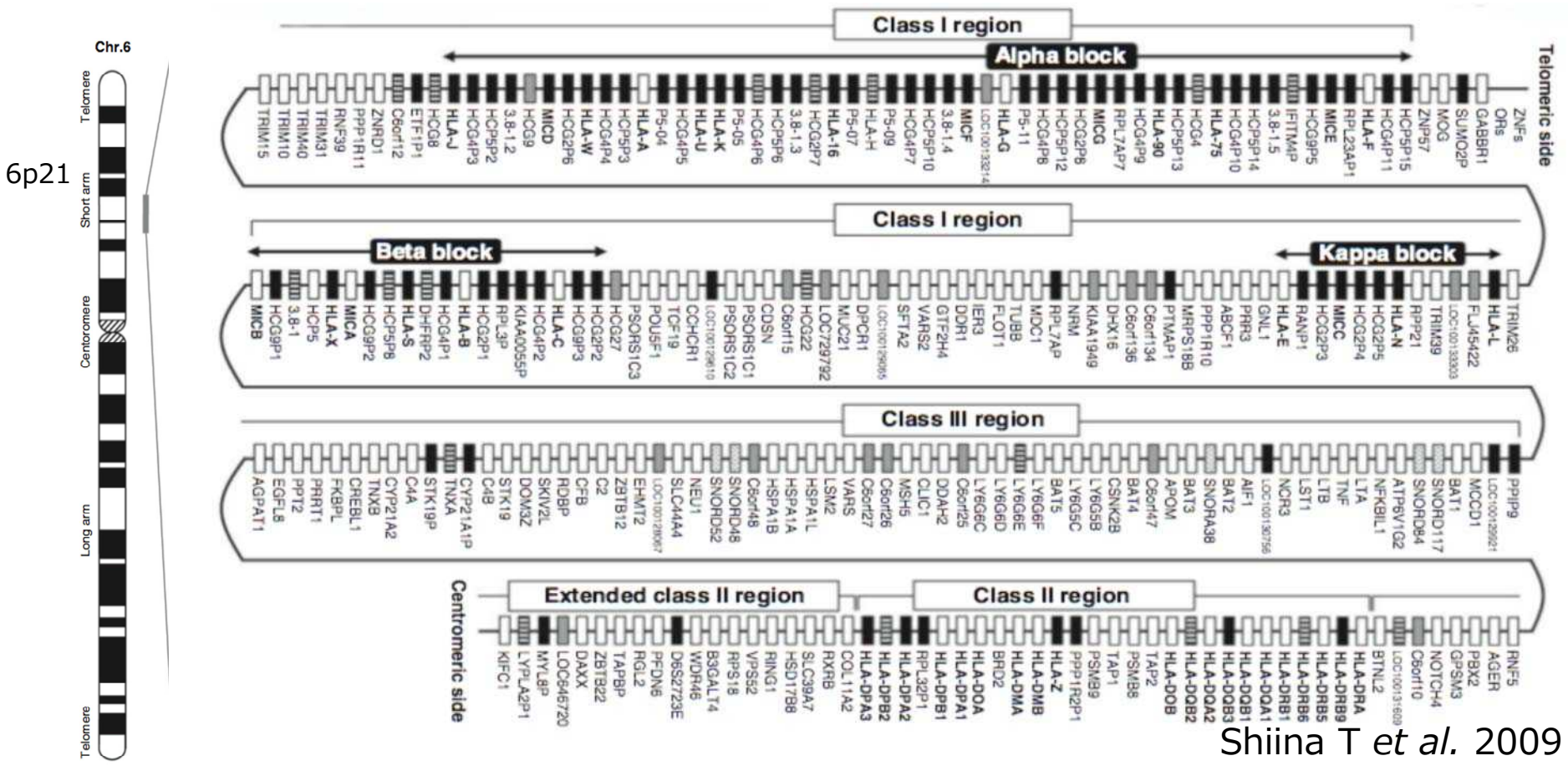
国立遺伝学研究所 総合遺伝研究系
人類遺伝研究部門
細道一善

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

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



Shiina T *et al.* 2009

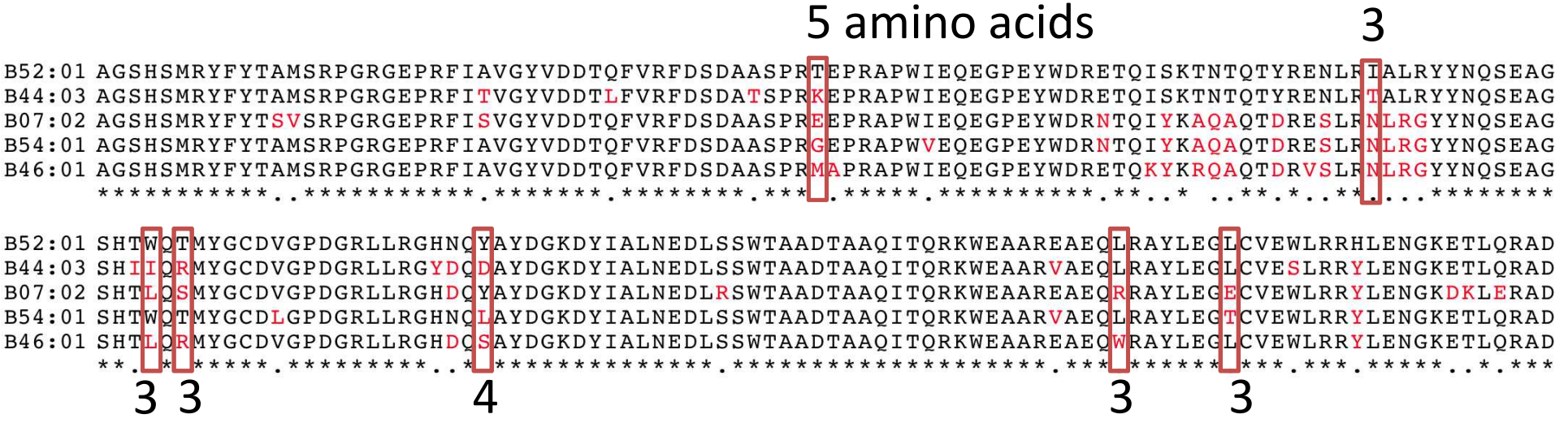
HLA領域の特徴

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

極めて高度な多型性を示す

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

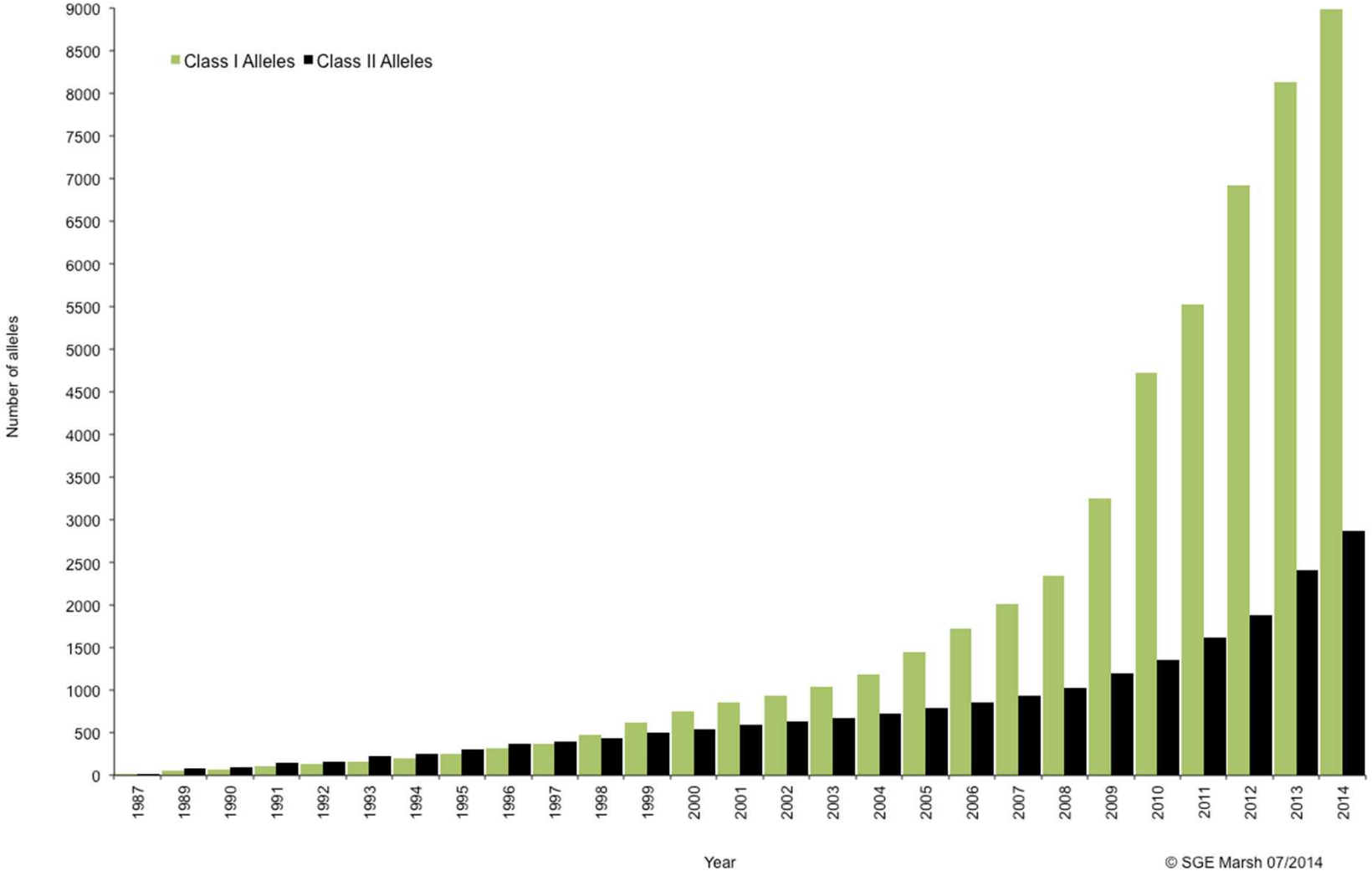
アミノ酸配列での類似性: **88.9%**
塩基配列での類似性: **93.9%**

IMGT/HLAデータベースに登録されているHLAアレル数

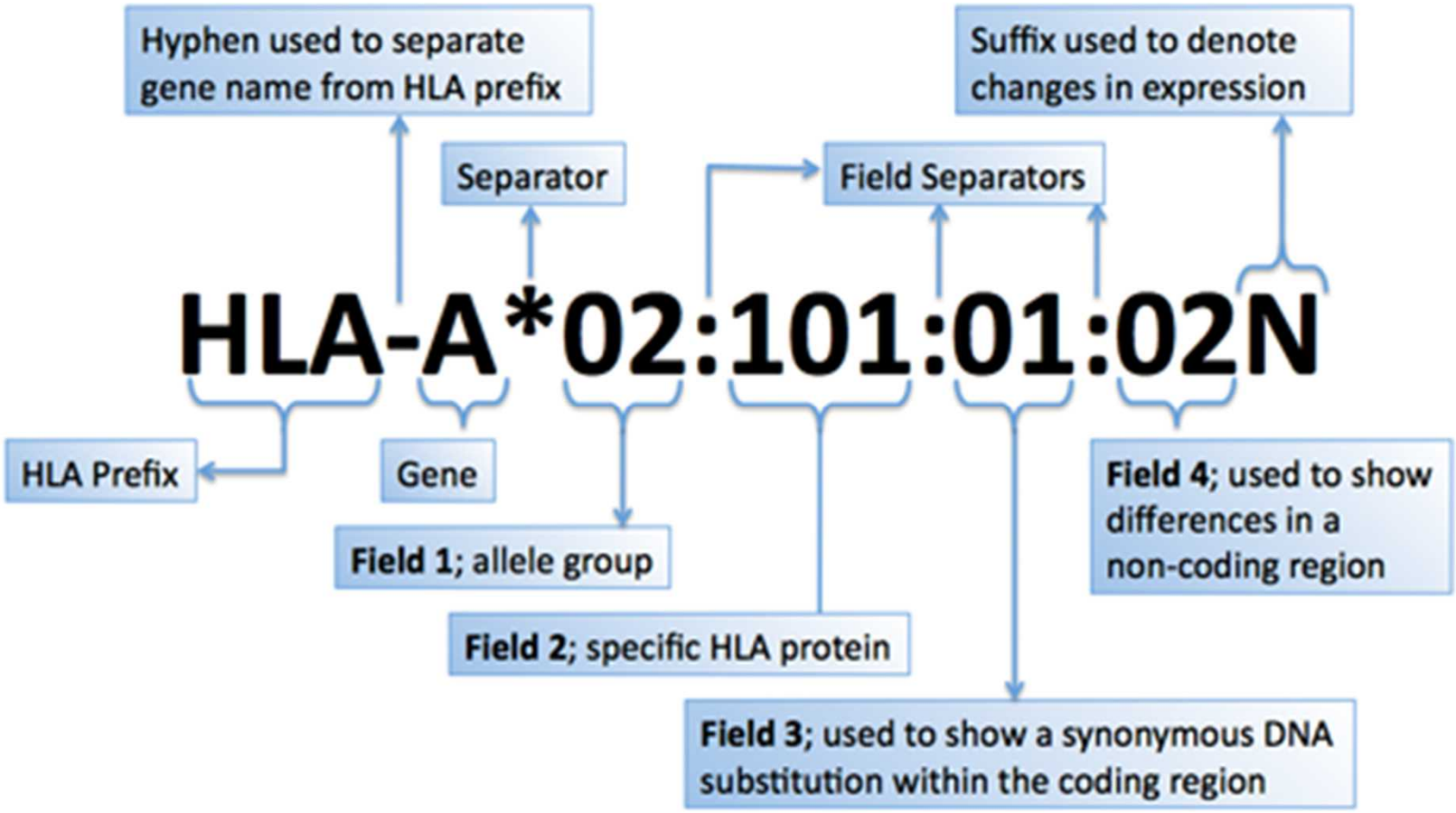
Numbers of HLA Alleles			
HLA Class I Alleles			8,976
HLA Class II Alleles			2,870
HLA Alleles			11,846
HLA Class I			
Gene	**A**	**B**	**C**
Alleles	2,884	3,589	2,375
Proteins	2,041	2,668	1,677
Nulls	133	119	71
HLA Class II			
Gene	**DRB**	**DQB1**	**DPB1**
Alleles	1,642	664	422
Proteins	1,211	435	351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

IMGT/HLA database Version Report - 3.17 (2014-07)

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

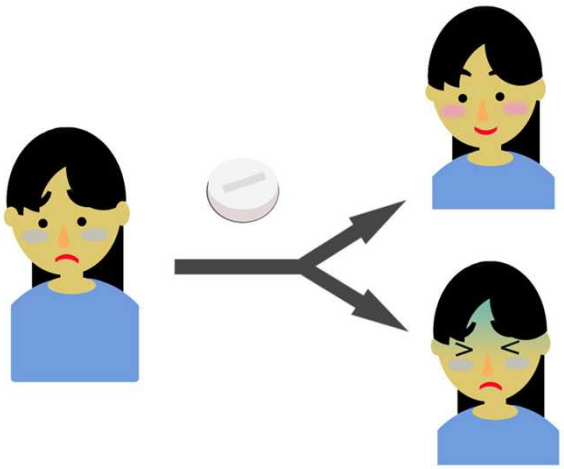

Hyphen used to separate gene name from HLA prefix
Separator
Field Separators
Suffix used to denote changes in expression
HLA-A*02:101:01:02N
HLA Prefix
Gene
Field 1; allele group
Field 2; specific HLA protein
Field 3; used to show a synonymous DNA substitution within the coding region
Field 4; used to show differences in a non-coding region

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)と肝内胆汁うっ滞	*HLA-A*33:03*	93	41.5
Carbamazepine(抗痙攣剤)と Stevens-Johnson 症候群	*HLA-B*15:02*	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	*HLA-B*57:01*	78	117.5
Allopurinol(抗痛風、抗尿酸血症剤)と薬疹	*HLA-B*58:01*	100	393.5
Ticlopidine(抗血小板剤)と肝障害	*HLA-A*33:03*	86	36.5
Amoxicillin-clavulanate potassium (抗生物質)と肝障害	*HLA-DRB1*15:01*	57	35.6
Flucloxacillin(抗生物質)と肝障害	*HLA-B*57:01*	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

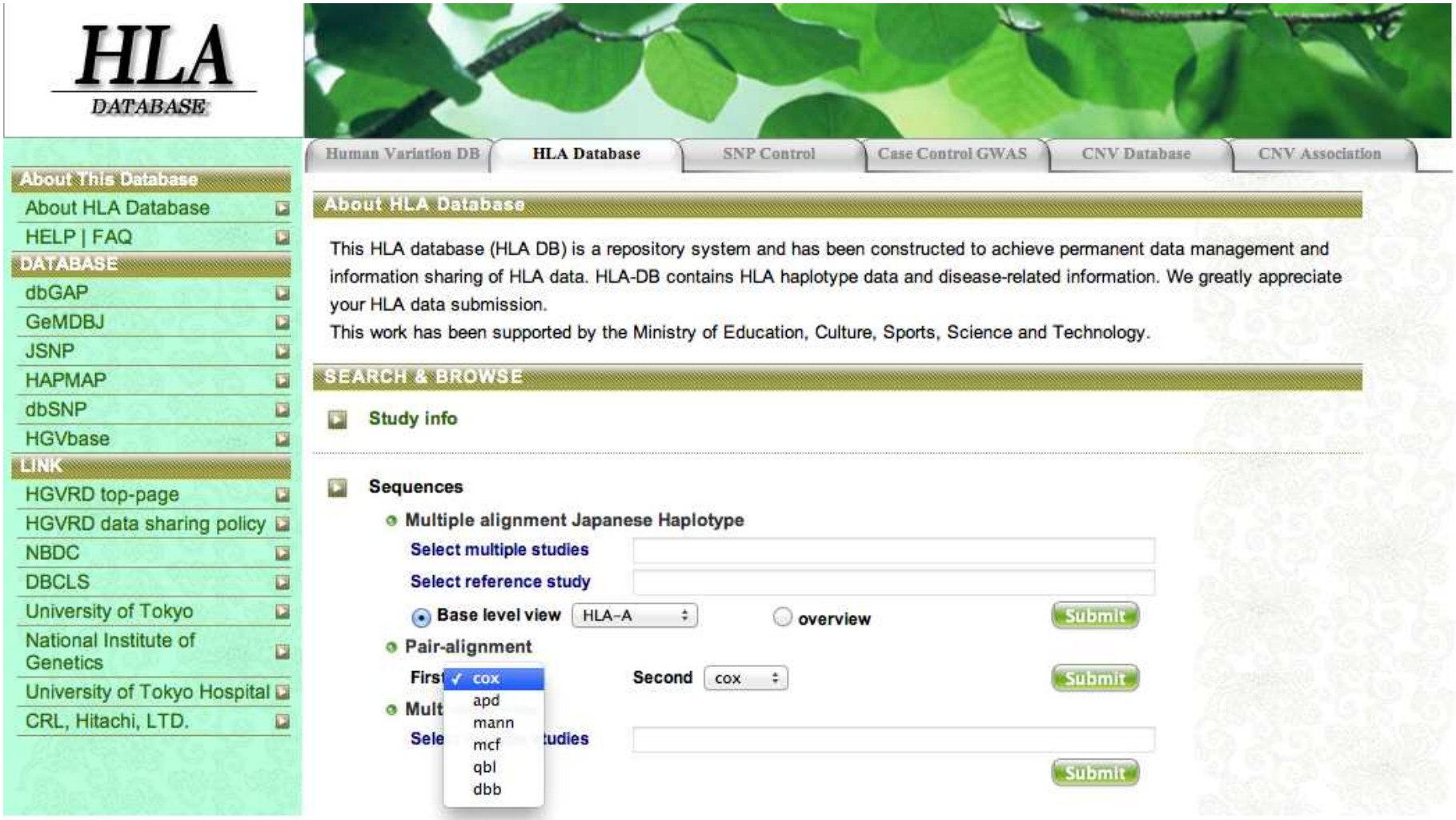
HLAタイピングによる予防的診断



	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



統合化推進プログラムのデータベースの新たな活用法

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

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.

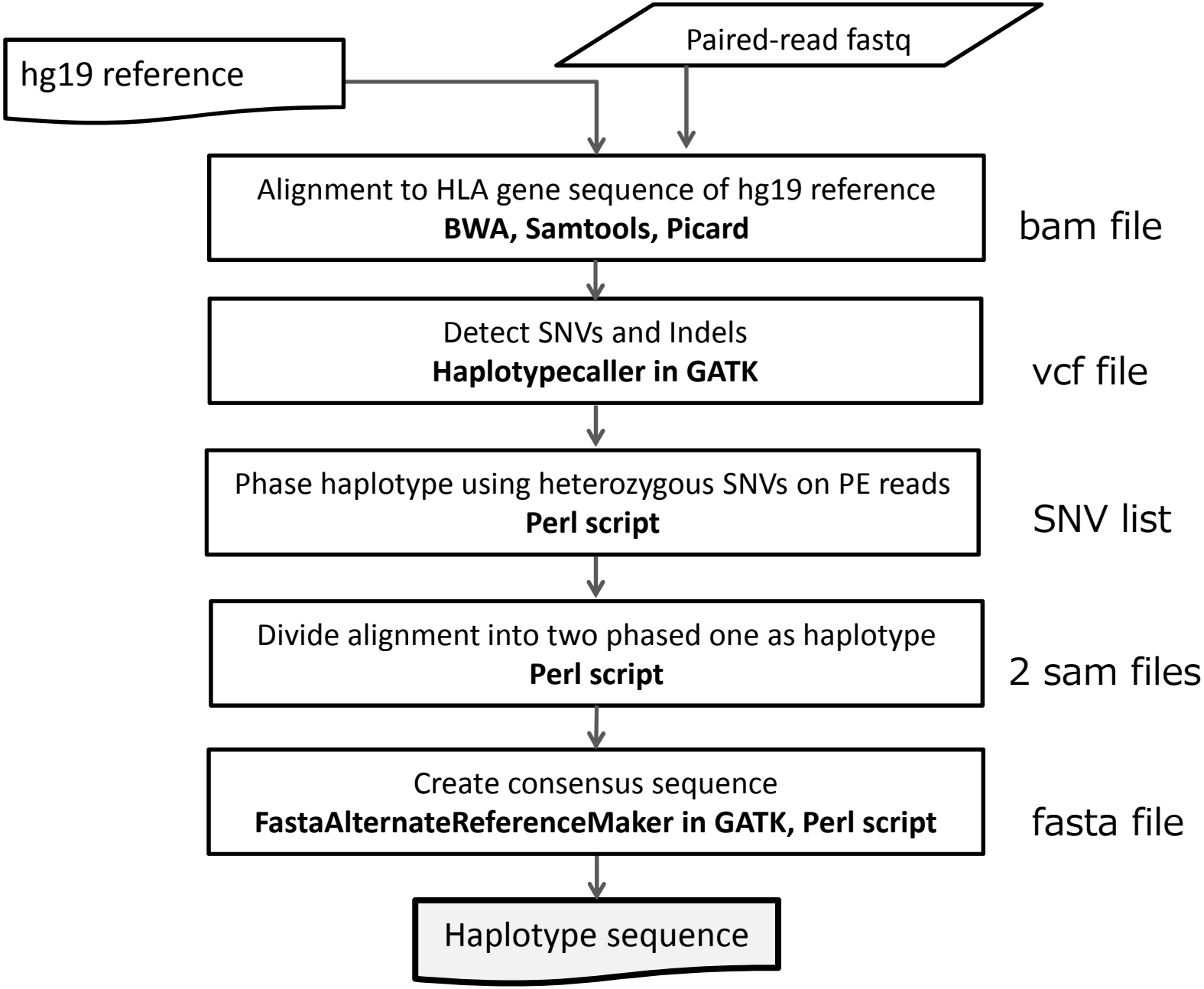
"As well as being an ethnic mix, these individuals represent the most common combinations of genes associated with certain diseases and so we will identify the most important sequence variants that predispose to these conditions," said Stephan Beck, Head of Human Sequencing at the Wellcome Trust Sanger 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により目的領域を増幅し、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

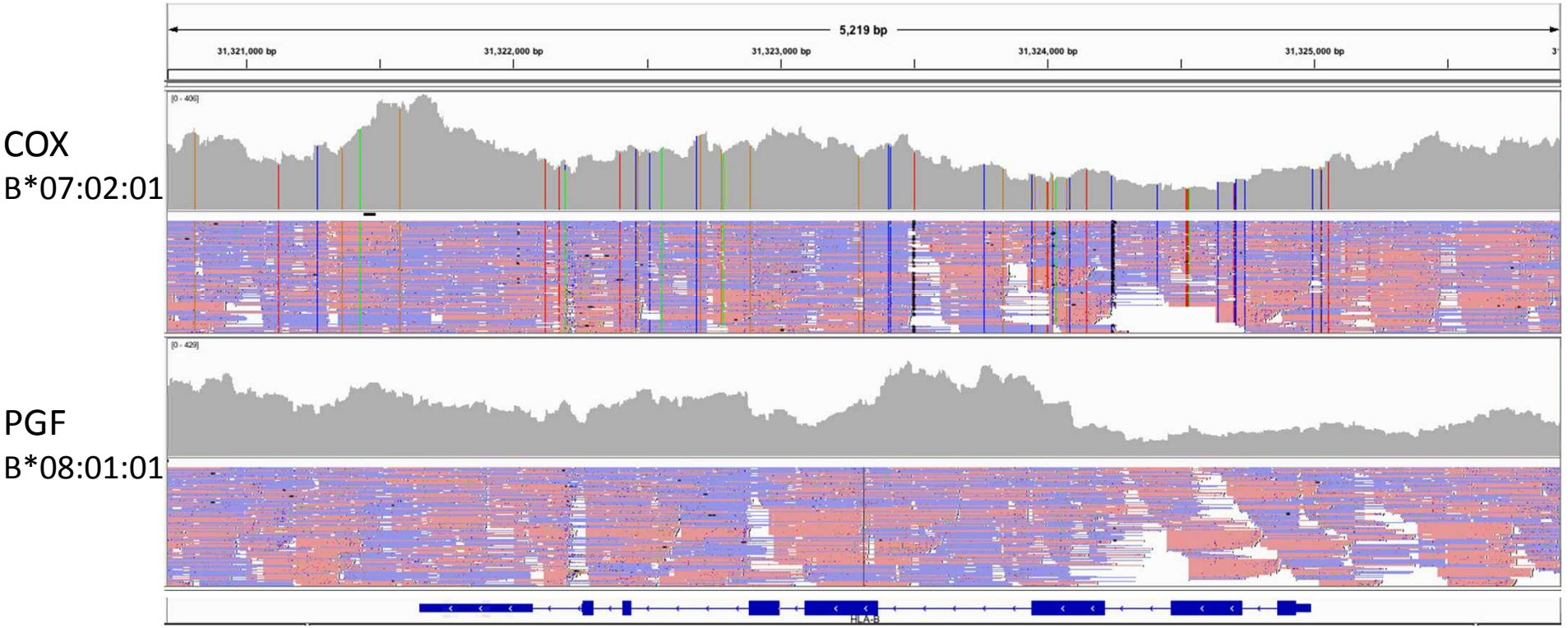
※リリース情報はこちら

分子データ

DRA000908

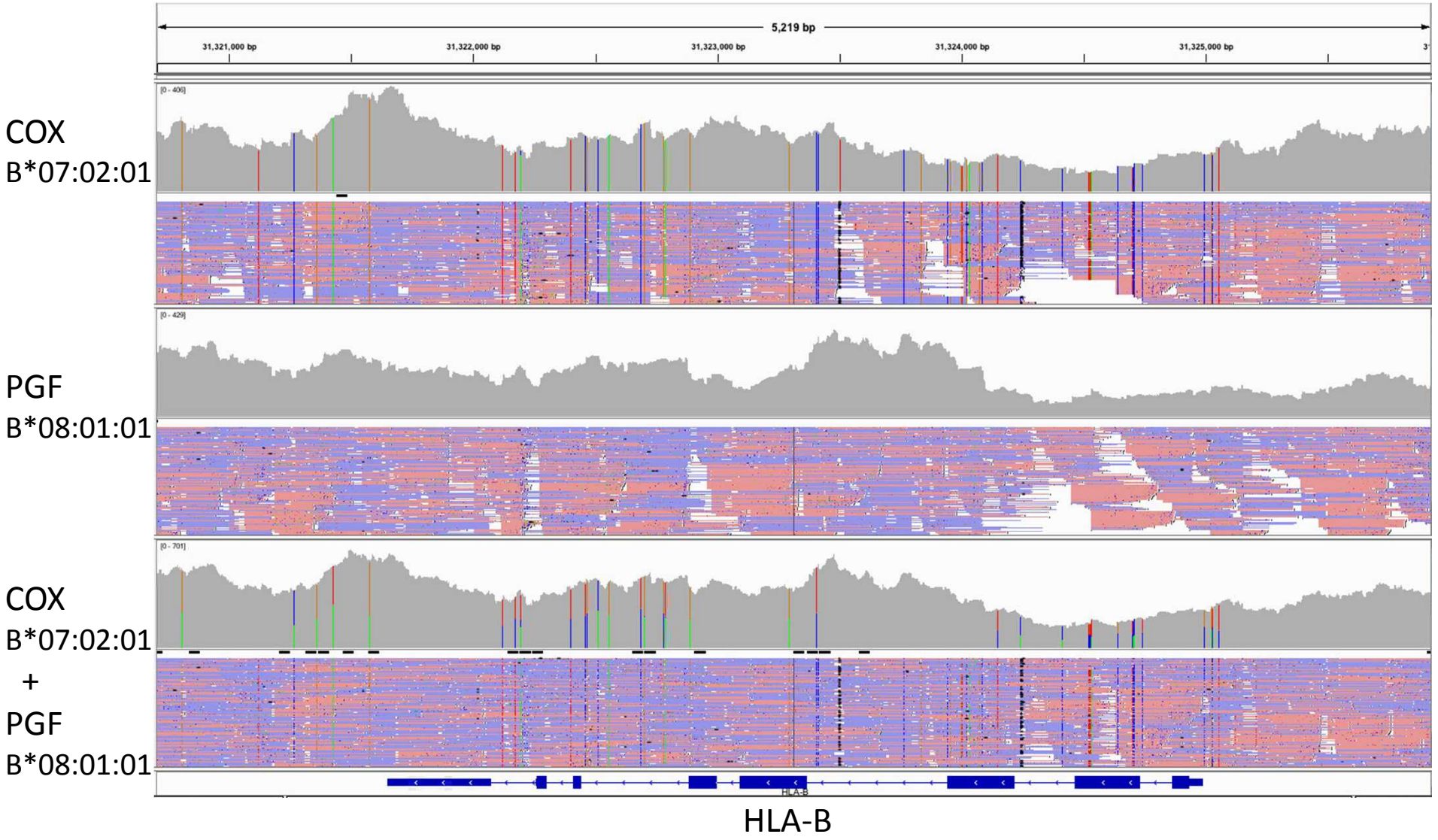
対象	セルライン33名分
規模	Target Capture
対象領域（Target Captureの場合）	HLA領域

セルラインにおけるHLA-BのNGSデータ

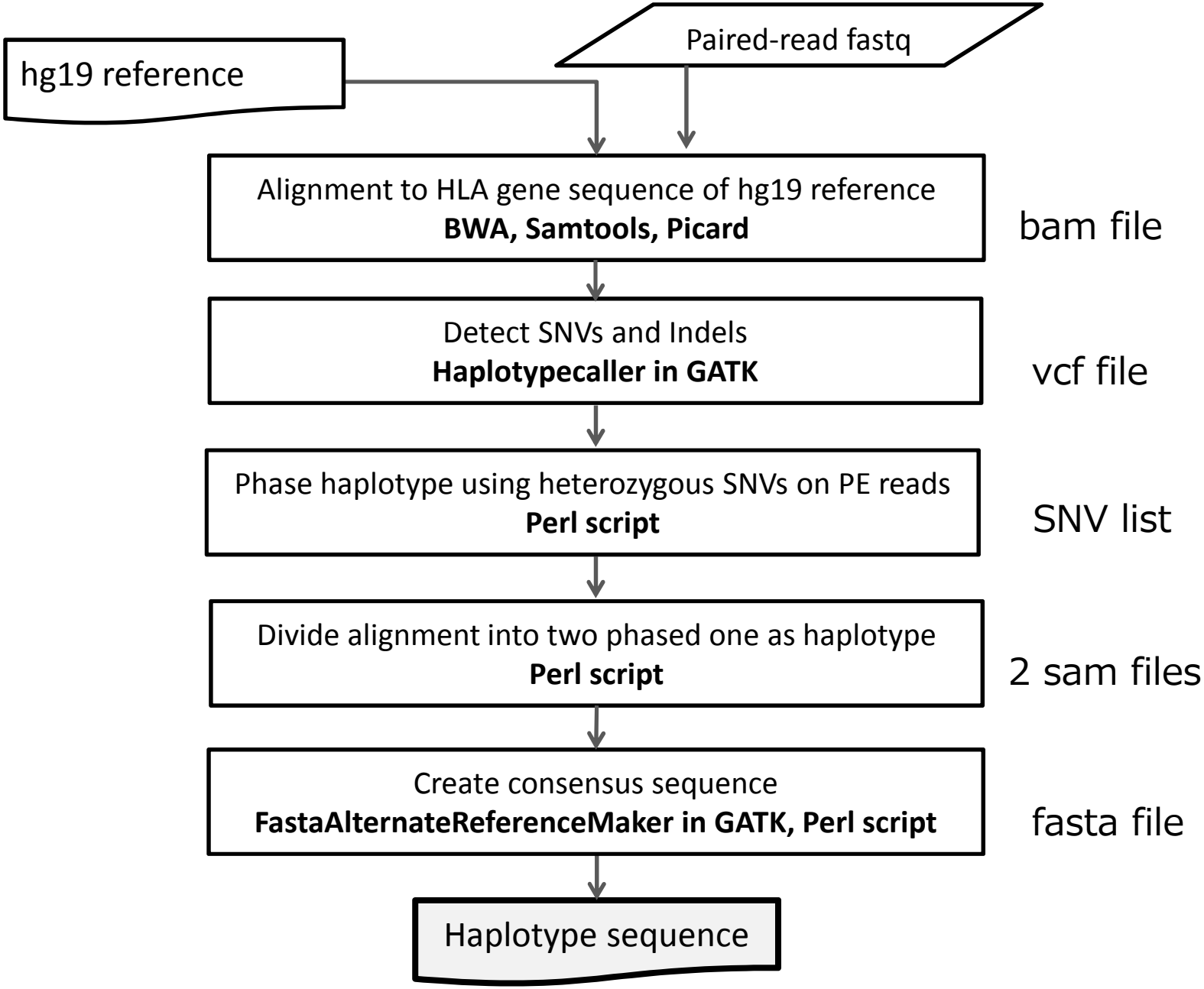


HLA-B

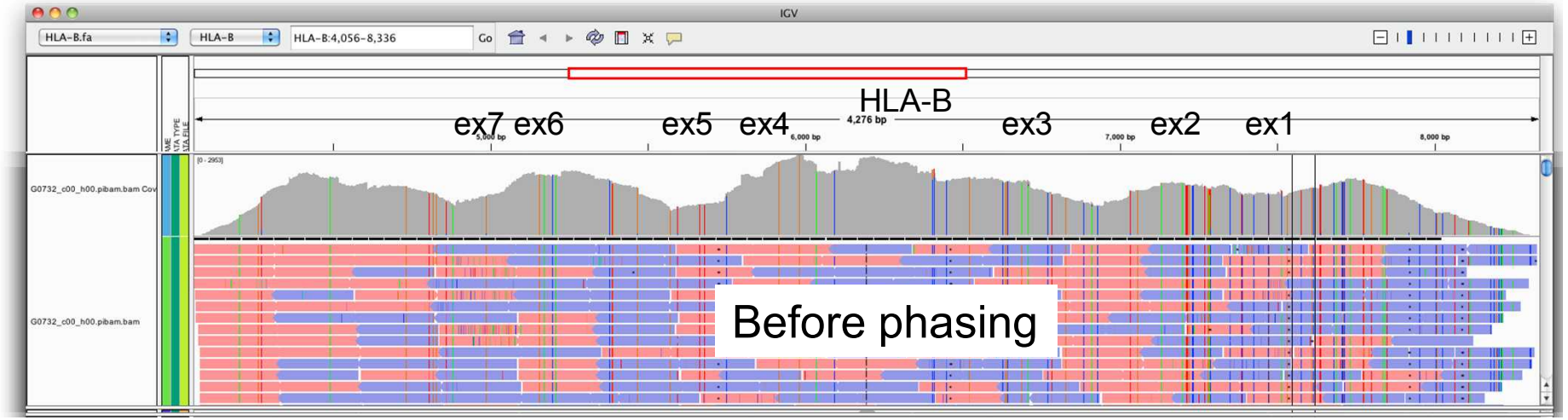
セルラインにおけるHLA-BのNGSデータ



Workflow of phase-defined sequencing



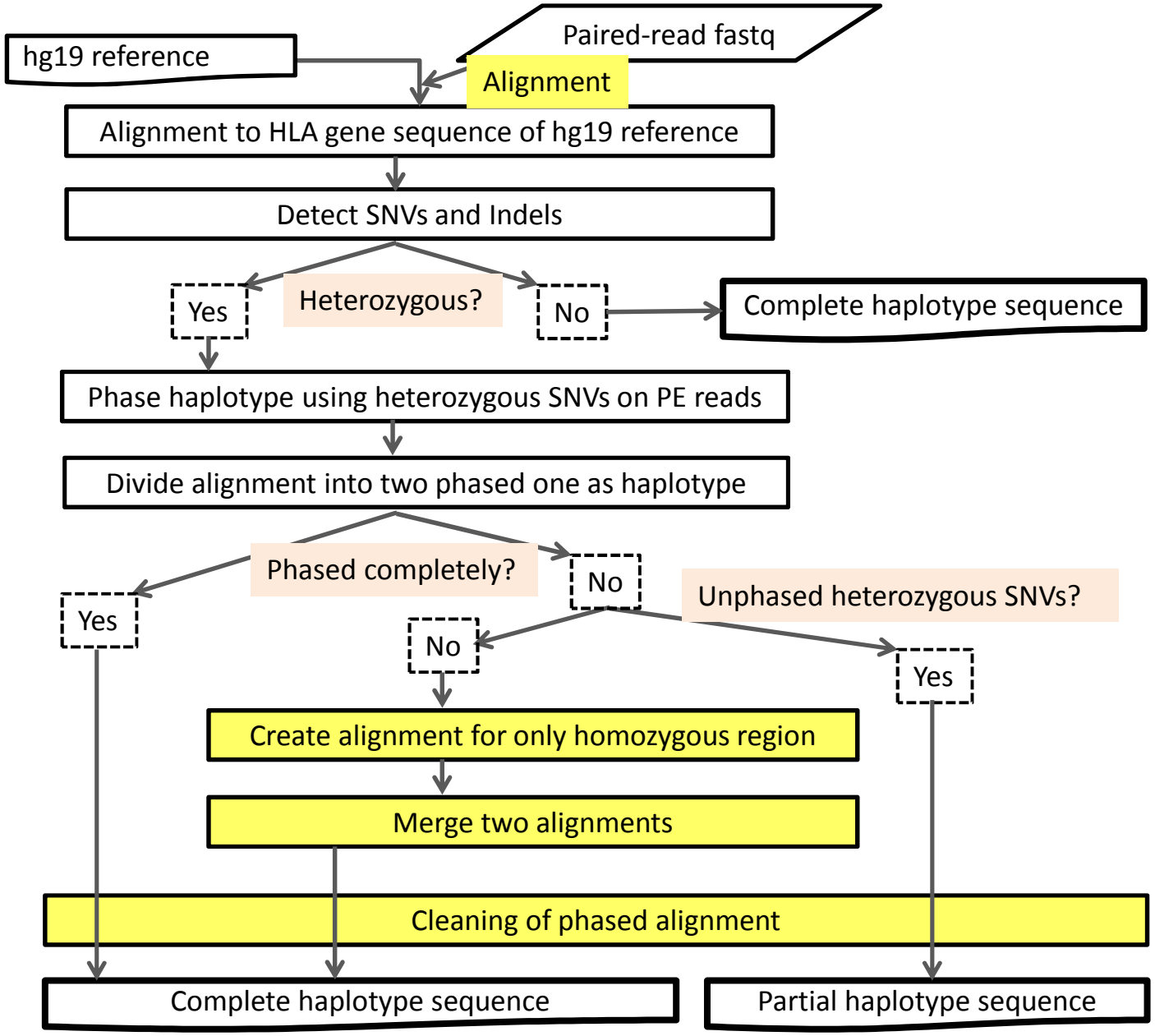
HLA遺伝子配列完全決定のワークフロー



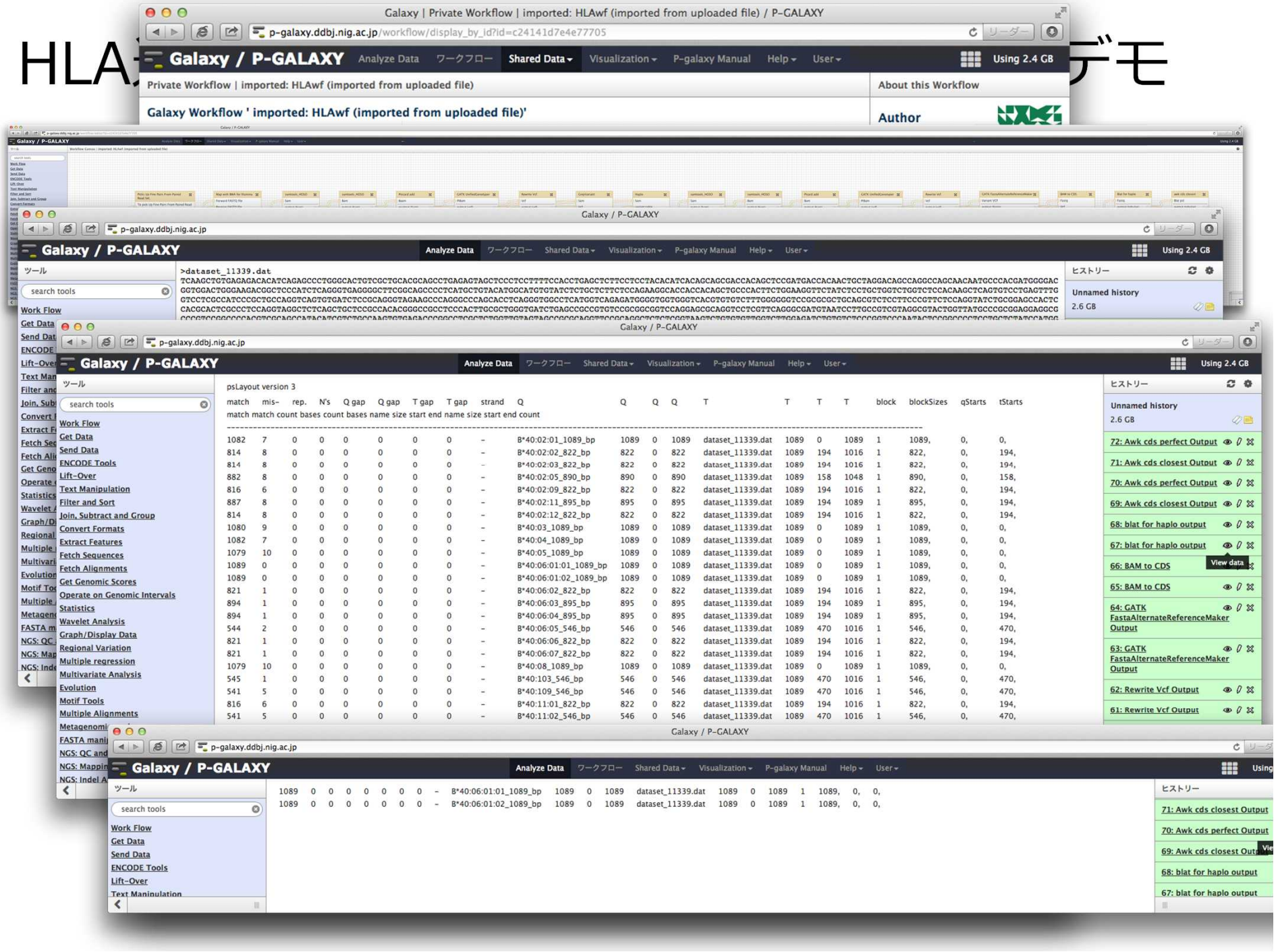
Phased haplotype 1 → *B*51:01:01*

Phased haplotype 2 → *B*15:02:01*

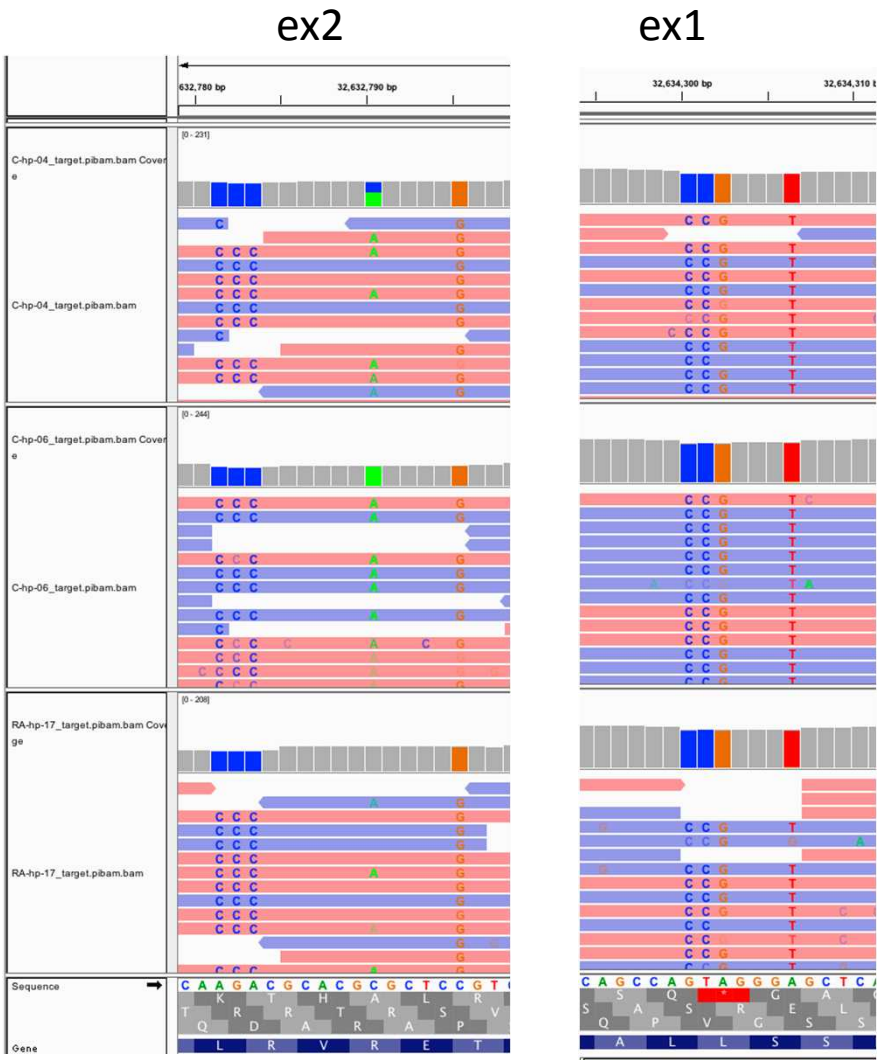
実データ解析へのパイプライン最適化



HLA


Galaxy | Private Workflow | imported: HLAwf (imported from uploaded file) / P-GALAXY
p-galaxy.ddbj.nig.ac.jp/workflow/display_by_id?id=c24141d7e4e77705
Galaxy / P-GALAXY
Analyze Data
ワークフロー
Shared Data
Visualization
P-galaxy Manual
Help
User
Using 2.4 GB
Private Workflow | imported: HLAwf (imported from uploaded file)
About this Workflow
Galaxy Workflow ' imported: HLAwf (imported from uploaded file)'
Author
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
Statistics
Wavelet Analysis
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Motif Tools
Multiple Alignments
>dataset_11339.dat
ヒストリー
Unnamed history
2.6 GB
psLayout version 3
72: Awk cds perfect Output
71: Awk cds closest Output
70: Awk cds perfect Output
69: Awk cds closest Output
68: blat for haplo output
67: blat for haplo output
66: BAM to CDS
65: BAM to CDS
64: GATK FastaAlternateReferenceMaker Output
63: GATK FastaAlternateReferenceMaker Output
62: Rewrite Vcf Output
61: Rewrite Vcf Output
View data

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



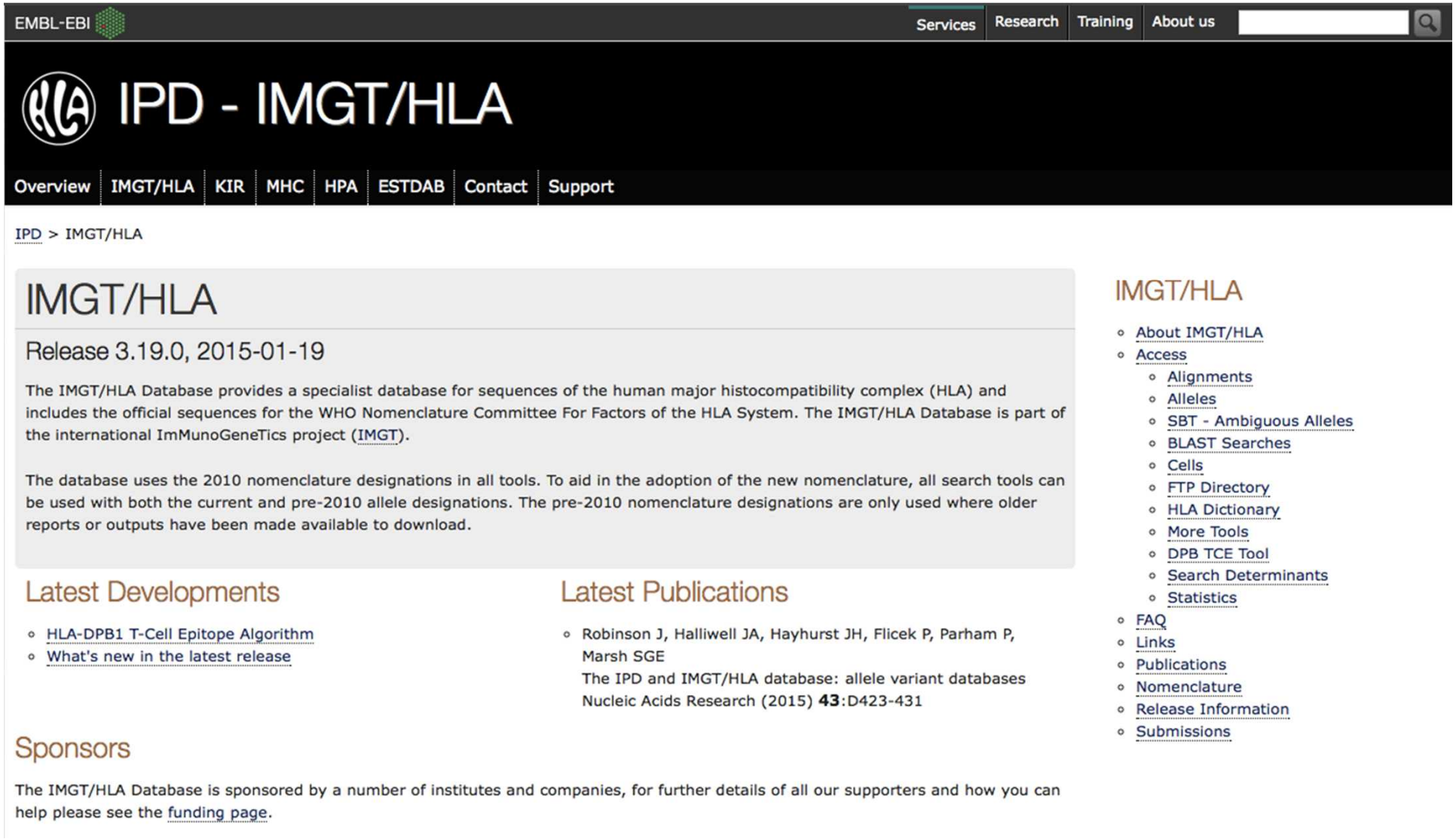
A-C -> DQB1*04:01
C-T -> DQB1*04:02

A-T -> New?
C-T -> DQB1*04:02

A-T -> New?

C-T -> DQB1*04:02

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



DQB1*04:01 の完全長塩基配列

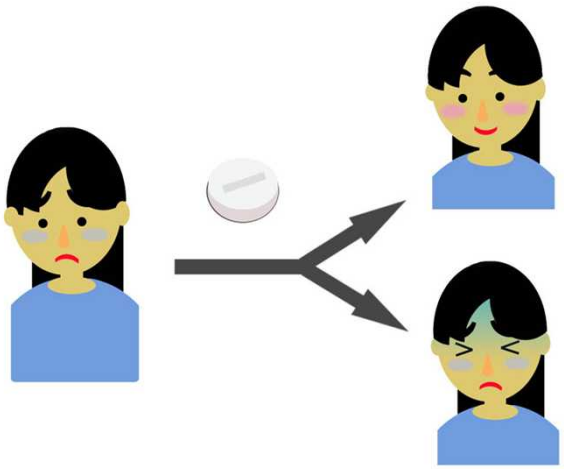
```
   1  GCAATTTTCT CTCCCCTGAA ATATGATCCC CACTTAATTT GCCCTACTGA AAGAATCCCA
  61  AGTATAAAAA CAACCAGTTT TTAATCAATA TTACAAAGAT GTTTACTGTT GAATCGCATT
 121 TTTCTTTGGC TTCTTAAAAT CCCTTAGGCA TTCAATCTTC AGCTCTTCTA TAATTGAGAG
 181 GAAGTTTTCA CCTCAAATGT TCATCCAGTG CAATTTGAAG ACGTCACAGT GCCAGGGACT
 241 GAATTGAGAA CCTTCACAAA AAAAAATGTC TGCCTAGAGA CAGATTAGGT CCTTCAGCTC
 301 CAGTGCTGAT TGGTTCCTTT CCAAAGGACC ATCCAATCCT GCCACGCAGG GAAACATCCA
 361 CAGGTTTTTA TTCTTTCTGC CAGGTACATC AGATCCATCA GGTCCAAGCT GTGTTGACTA
 421 CCACTACTTT TCCCTTCGTC TCAATTATGT CTTGGAAGAA GGCTTTGCGG ATCCCTGGAG
 481 GCCTTCGGGT AGCAACTGTG ACCTTGATGC TGGCGATGCT GAGCACCCCG GTGGCTGAGG
 541 GCAGAGACTC TCCCGGTAAG TGCAGGGCCA CTGCTCTCCA GAGCCGCCAC TCTGGGAACA
 601 GGCTCTCCTT GGGCTGGGGT AGGGGGATGG TGATCTCCAT GATCTCGGAC ACAATCTTTC

6301 CAATGGTCTC TGTTCATGGT ATATTTGCTG CTATGAGGAT CAAGACTTAG GGTCGAAGTT
6361 TGCCAGTTTC TAGGAATCTC CAGAGGTTGT TCCCCAGAAC CAAGCCTTAA CTTTGGTGGT
6421 ATCTTCTTGT GAAATGTGAA GCCAGAACCA CAGCTTAAAT GTTAGACAAG AGGATGATGC
6481 CCACTTTGTG CCACATGTTG GTGGCTACTG CCTGTAGGCA TTTTCCAGTG ACTGAAAGAG
6541 GCTGCTAGTG GTAGGGATGA GGTATCATCC AATTTTCTAA AAAGATTGAA CCCTTCATAT
6601 TCCCCAGAAG AGTAACAGCT GTTCCGCCAC TTCCCACATA TCTGCATCAA GCTGAAGTTC
6661 TGTGTCCTCA CGAGCTGATT TCACCTTTGC ACAGATCTTG CGGGAGGTGA CAATAATACA
6721 TTCTGGACCT CAGCTTTCTC TGTCTGAAGC TGCAGGGGGC CCCTGAGGGG TGGGGGAGAT
6781 TGCAGGCCCA CCAGCGTACC CTGTGCTGAT CATCCCTCTT CTCTCTTCTT CAGGGCTCCT
6841 GCACTGACTC CTGAGACTAT TTTAACTGGG ATTGGTTATC ACTTTTCTGT 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