Supplementary Algorithm, Tables and Figures

ATHLATES: Accurate Typing of Human Leukocyte Antigen Through Exome Sequencing

Chang Liu[1,6], Xiao Yang[2,6], Brian Duffy[3], Thalachallour Mohanakumar[1,4], Robi D. Mitra[4,5] Michael C. Zody[2] & John D. Pfeifer[1]

[1]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA. [2]Genome Sequencing and Analysis Program, Broad Institute of MIT & Harvard, Cambridge, Massachusetts, USA. [3]HLA Laboratory, Barnes-Jewish Hospital, St. Louis, Missouri, USA. [4]Department of Surgery, Washington University School of Medicine, St. Louis, Missouri, USA. [5]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to X.Y. (xiaoyang@broadinstitute.org) or J.D.P. (pfeifer@path.wustl.edu).

LIST OF SUPPLEMENTARY ALGORITHMS

Supplementary Algorithm 1. Contig generation via Union-Find algorithm.

LIST OF SUPPLEMENTARY TABLES

Supplementary Table 1. Sequences of typed HLA genes included in the reference.

Supplementary Table 2. Characteristics of exome-seq data used for *in silico* HLA typing.

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation.

Supplementary Table 4. A sample report generated by ATHLATES. Twenty one alleles were initially identified to be within distance 2. The first 4 alleles have distance 0 and in the meanwhile, they correspond to 2 types of protein coding sequences. According to candidate allele selection criterion, the candidate allele list consists of these 4 alleles, where in total 10 allelic pairs (4 homo- and 6 hetero-zygous) were resort to scoring function. The pairing between DQB1*02:01:01 and any one of DQB1*03:03:02:01, DQB1*03:03:02:02, and DQB1*03:03:02:03 gives the equivalent minimum score as compared to the pairing among DQB1*03:03:02:01, DQB1*03:03:02:02, and DQB1*03:03:02:03, where the latter cannot account for all the variants present in the contigs.

Supplementary Table 5. Statistical analysis of allelic bias within individual exons.

LEGENDS FOR SUPPLEMENTARY FIGURES

Supplementary Figure 1. Ambiguities resulted when merging paired-end reads (r_0, r_1). (A) r_0 and r_1 align to $allele_0$, where the aligned regions partly overlap. The bases highlighted in red indicate disagreement, which is denoted by a degenerate base S = (11) in the merged read *r'*. With r_2 aligning to the same region, base C is selected in the newly merged read *r''*. (B) r_0 and r_1 align to both $allele_1$ and $allele_2$. $Allele_1$ contains an additional base C highlighted in red compared to $allele_0$, within the gap between r_0 and r_1, resulting in two legitimate merging results *r'* and *r''*. *r''* is supported upon the identification of an additional read r_2 that is consistent with *r''*.

Supplementary Figure 2. Examples of preferring similarity to length when identifying the best match of an exon and the necessity of considering non-zero distance alleles. (A) The alignment of the second exons of the heterogeneous alleles A*02:01:01:01 and A*02:11:01 of sample HG01953 and the two contigs c_i and c_j identified by ATHLATES is shown. Contig c_i is identical to exon 2 of A*02:01:01:01 and contig c_j is a proper prefix of exon 2 of A*02:11:01. Because exon 2 of the two alleles differ only by the two bases highlighted, c_i naturally serves as a candidate hit for exon 2 of A*02:11:01 that has a length identical to exon 2 and a similarity over 99%. Nevertheless, c_j is shorter but has a 100% similarity. A*02:11:01 would have been missed without tolerating missing bases. (B) The alignment of the seventh exons of candidate alleles C*04:01:01:01 and C*04:30 and contig c_i identified by ATHLATES is shown. c_i is identical to exon 7 of C*04:01:01:01 and it is the only contig that can serve as a candidate hit for exon 7 of C*04:30. Without tolerating non-zero distance alleles, C*04:30 would have been missed.

Supplementary Figure 3. Variations in fold coverage among different target HLA genes and their exons. The median fold coverage is plotted for individual exons of each HLA gene. Error bars show 95% confidence intervals. The data are divided into panels by gene and groups by sample. Sample labels are shown below the bottom panel for the last gene.

Supplementary Figure 4. Schematic of principles for allelic pair inference. (A) Principle of parsimony. Assuming two contigs, $Contig_a$ and $Contig_b$, have been assembled and two candidate alleles, $allele_0$ and $allele_1$,have been identified. $Allele_0$ differs from $allele_1$ by one base, highlighted in red. ATHLATES would check three allelic pairs: ($allele_0$, $allele_1$), ($allele_0$, $allele_0$) and ($allele_1$, $allele_1$). We cannot rule out the possibility of ($allele_0$, $allele_1$) or ($allele_1$, $allele_1$) being the correct answer as $Contig_b$ could have supported $allele_1$ if the missing bases were TGG. However, the homozygous pair ($allele_0$, $allele_0$) is preferred as it is based on the fewest assumptions and sufficient to represent both contigs. (B) The allelic pair should capture as much information as presented in the data. Assuming four potential alleles (B*55:02:01, B*56:11, B*35:03:01, B*35:60) have been identified as candidates for HLA-B of sample

HG01873 and every exon of each allele are present in the contigs. It turns out that these four alleles share a subset of exons. As labeled in the table, for a given exon (1-4), whenever two alleles share the same exon sequence, they are assigned the same label A or B: *e.g.* the second exon between B*55:02:01 and B*56:11 are identical, hence, they are assigned the same label A. In allelic pair inference, a total of 10 possible allelic combinations (4 homozygous and 6 heterozygous) of these four alleles are considered. Only the pair B*55:02:01 and B*35:03:01 could account for all types of haplotypes of exons supported by data, resulting in heterozygosity (both A and B alleles) at all four exons. This example also illustrates a resolution of the phase of different exons.

Supplementary Figure 5. Diversity of the HLA genotypes of included study samples. The genetic diversity of the 15 samples included in this study is visualized by mapping all their alleles onto the phylogenetic trees of each gene typed. The alleles are mapped at the four-digit resolution (same protein sequence corresponding to Sanger sequenced exons) and are highlighted in colors corresponding to each gene. The phylogenetic trees are generated using RAxML (v.7.3.3). The genetic diversity can be appreciated based on the coverage and pattern of colored alleles. The detailed specificity of included alleles can be found in **Supplementary Table 3**.

Supplementary Figure 6. An example of exom-seq data of the individual NA18526 (1st row) that has inadequate coverage over target HLA genes. The five target HLA genes of this individual were typed previously. ATHLATES reported no typing result for any of the target gene as the assembled contigs did not unambiguously support any candidate allele due to insufficient exon coverage. The alignment is generated using BLAST between contigs of a target gene and one of the known cDNA sequences of this gene. For DQB1 gene, no contig was produced by ATHLATES. HLAminer, on the other hand, reported typing results for HLA-A, B, C and DRB1 genes, all of which are discordant with the known type except HLA-B, where the known type is among one of the seven equally supported candidates. The other 4 samples with inadequate coverage are NA18605 (2nd row), HG01515 (3rd row), HG01049 (4th row), and HG00731 (5th row). The bars at the bottom row show exons of each gene.

Supplementary Figure 7. Comparison of HLA typing results between ATHLATES and the conventional Sanger method grouped by genes and samples. The left panel shows the number and type of allelic pairs that are reported by both methods in concordance. In 31 out of 65 typing results one unique allelic pair is reported, while two or more allelic pairs that harbor intronic polymorphisms are reported in the rest by both methods. The right panel demonstrates the number and type allelic pairs ruled out by ATHLATES but are encountered by Sanger method after the first round of sequencing. The causes of these extra allelic pairs include polymorphisms in exons that are not sequenced, polymorphisms at positions not covered by

Sanger sequences, and cis/trans ambiguities (or same diploid sequence shared by two or more different allelic combinations). Many of these extra allelic pairs are subsequently ruled out by additional rounds of sequencing or PCR as detailed in Supplementary Table 3; this is associated with significantly increased cost and turnaround time.

Supplementary Figurer 8. Differential fold coverage at variant positions between validated allelic pairs for each individual. The fold coverage is plotted for each allele of a heterozygous allelic pair (colored in blue and red) at positions where they differ. In case of homozygosity, the coverage is plotted for all alignment positions with symbols colored in red only. Dotted lines are exon boundaries.

Supplementary Figure 9. Size of allelic bias in discrete exons. Differences in fold coverage between two heterozygous alleles at each variant position are plotted as the percentage of total coverage at the same variant position. Each point represents the difference in fold coverage (y axis) at each variant position. Data are grouped by different exons and then arranged by genes (columns) and samples (rows); involved allelic pairs are shown to the right of each panel. The extent of allelic bias observed suggests that the bait for exome capture in the HLA region may benefit from additional optimization.

Supplementary Algorithm 1 Contig Generation via Union-Find algorithm.

Require: merged paired-end reads

Initialize each read to be a contig, record base frequency for each position
$l \leftarrow$ max read length
Decompose contigs into l-mers, for each l-mer track its contig position
Sort l-mers by the order of decreasing frequency
Each contig is initialized to be a root node of a tree
repeat
  for each l-mer **do**
    for each pair of contigs (C_i, C_j) indexed by this l-mer **do**
      if the roots of C_i and C_j differ in the tree **then**
        Obtain the two contigs corresponding to the two roots
        Generate alignment for these two contigs by matching the l-mer
        if the alignment is consistent **then**
          Merge two contigs form a new contig
          Assign the merged contig to one root and assign this
                root as the parent to the root of the other contig
          Accumulate base count for each contig position
        end if
      end if
    end for
  end for
  Decrease l by a constant number (default 10)
until $l \leq$ minimum length (default 40)

Supplementary Table 1. Sequences of typed HLA genes included in the reference*

	gDNA sequences			cDNA sequences		
Genes†	Count	Median length (bp)	Range (bp)	Count	Median length (bp)	Range (bp)
HLA-A	112	3332.5	2903 - 3518	1884	546	540 - 1163
HLA-B	149	3312	1208 - 3340	2489	546	531 - 1208
HLA-C	101	3343	2700 - 3368	1382	546	544 - 1197
HLA-DRB1	26	13463.5	10299 - 16120	1092	270	222 - 801
HLA-DQB1	18	7107	6800 - 7480	165	270	148 - 810

*Based on IMGT/HLA Database Release 3.8.

†The other off-target genes in the reference include DMA, DMB, DOA, DOB, DPA1, DPB1, DQA1, DRA, DRB3, DRB4, DRB5, E, F, G, H, J, K, L, MICA, MICB, P, TAP1, TAP2, V. They are less polymorphic compared to the HLA genes being typed, and both gDNA and cDNA sequences available are included.

Supplementary Table 2. Characteristics of exome-seq data used for *in silico* HLA typing

Sample	Run ID	Study*	Sequencing facility†	Population	Country	Capture platforms	Instrument	Read count	Exome coverage (≥10X)
Samples with adequate coverage									
HG01756	SRR359102	1000 Genomes	WUGI	Iberian	Spain	V2 (SureSelect)	GAIIx	190644090	97%
HG01757	SRR359103	1000 Genomes	WUGI	Iberian	Spain	V2 (SureSelect)	GAIIx	192645934	97%
HG01872	SRR359298	1000 Genomes	WUGI	Kinh	Vietnam	V2 (SureSelect)	GAIIx	200139823	97%
HG01873	SRR359295	1000 Genomes	WUGI	Kinh	Vietnam	V2 (SureSelect)	GAIIx	203842382	97%
HG01886	SRR360655	1000 Genomes	WUGI	African	Barbados	V2 (SureSelect)	GAIIx	182195662	97%
HG01953	SRR360288	1000 Genomes	WUGI	Peruvian	Peru	V2 (SureSelect)	GAIIx	182525408	96%
HG01968	SRR360391	1000 Genomes	WUGI	Peruvian	Peru	V2 (SureSelect)	GAIIx	193999086	96%
HG02014	SRR360148	1000 Genomes	WUGI	African	Barbados	V2 (SureSelect)	GAIIx	174553282	97%
HG02057	SRR359301	1000 Genomes	WUGI	Kinh	Vietnam	V2 (SureSelect)	GAIIx	204351266	97%
NA20313	SRR359098	1000 Genomes	WUGI	African	USA	V2 (SureSelect)	GAIIx	194492515	98%
NA20313R‡	SRR359108	1000 Genomes	WUGI	African	USA	V2 (SureSelect)	GAIIx	195282918	98%
NA18507	0635_7_ATCGAGC	Internal Validation	GTAC	Yoruba	Nigeria	V3 (SureSelect)	HiSeq2000	192099768	96%
NA19129	0635_6_CACCTCC	Internal Validation	GTAC	Yoruba	Nigeria	V3 (SureSelect)	HiSeq2000	208046924	97%
NA19240	0635_7_TACTCTA	Internal Validation	GTAC	Yoruba	Nigeria	V3 (SureSelect)	HiSeq2000	189722418	97%
NA19240R‡	0636_4_CTCAATG	Internal Validation	GTAC	Yoruba	Nigeria	V3 (SureSelect)	HiSeq2000	177544782	96%
Samples with inadequate coverage									
HG01515	SRR231271	1000 Genomes	WUGI	Iberian	Spain	V2 (SureSelect)	HiSeq2000	80709178	92%
HG01049	SRR107086	1000 Genomes	BI	Puerto Rican	Puerto Rico	V2 (SureSelect)	HiSeq2000	151803340	95%
HG00731	SRR107083	1000 Genomes	BI	Puerto Rican	Puerto Rico	V2 (SureSelect)	HiSeq2000	151983152	96%
NA18526	ERR031854	1000 Genomes	BGI	Chinese	China	V1 (Nimblegen)	HiSeq2000	50730605	80%
NA18605	ERR031873	1000 Genomes	BGI	Chinese	China	V1 (Nimblegen)	HiSeq2000	50730605	80%

*1000 Genomes project (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data/); the data from the internal validation study are available upon request.

†WUGI, Washington University Genome Institute; GTAC, Genome Technology Access Center at Washington University; BI, Broad Institute. All samples were sequenced as paired-end reads (2 X 101 bps).

‡Duplicate sequencing of NA20313 and NA19240.

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
HG01756	HLA-A	30:02:01	66:01	30:01	136.1	30:02:01	66:01				30:02:01	66:01
				30:04	136.1	30:25	66:05	Zrr				
				30:02	127.2							
				66:01	137.2							
	HLA-B	18:01:01:01	41:02:01	18:26	157.8	18:01:01:01	41:02:01				18:RRG	41:02
		18:01:01:02	41:02:01	18:01:01:01	157.8	18:01:01:02	41:02:01					
				18:01:01:02	147.2	18:17N	41:02:01	Sr-				
				41:01	152	18:14	41:03:02	Z				
				45:04	57.7	18:39	41:02:01	Zr-				
				45:01	57.7	18:69	41:02:01	Zr-				
				50:01	57.7							
				15:83	57.7							
	HLA-C	05:01:01:01	17:01:01:01	05:01	162.7	05:01:01:01	17:01:01:01				05:01	17:MN
		05:01:01:01	17:01:01:02	17:01	197.2	05:01:01:01	17:01:01:02					
		05:01:01:02	17:01:01:01			05:01:01:01	17:02	-r				
		05:01:01:02	17:01:01:02			05:01:01:01	17:03					
						05:01:01:02	17:01:01:01					
						05:01:01:02	17:01:01:02					
						05:01:01:02	17:02	-r				
						05:01:01:02	17:03					
						05:01:12	17:01:04	Z				
						05:29:01	17:05	Zrr				
	HLA-DRB1	03:01:01:01	03:01:01:01	03:01	162.4	03:01:01:01	03:01:01:01				03:01	03:01
		03:01:01:01	03:01:01:02	07:01	87.4	03:01:01:01	03:01:01:02					
		03:01:01:02	03:01:01:02			03:01:01:01	03:01:08	S				
						03:01:01:02	03:01:01:02					
						03:01:01:02	03:01:08	S				
						03:01:08	03:01:08	S				
	HLA-DQB1	02:01:01	02:01:01	02:01	141.3	02:01:01	02:01:01				02:01:01	02:01:01
HG01757	HLA-A	01:01:01:01	02:01:01:01	01:03	249.1	01:01:01:01	02:01:01:01				01:BMMP	02:01
		01:01:01:01	02:01:01:02L	01:01	249.1	01:01:01:01	02:01:01:02L	-r				
		01:01:01:01	02:01:01:03	02:01	116.4	01:01:01:01	02:01:01:03					
		01:01:01:02N	02:01:01:01	02:03	116.4	01:01:01:02N	02:01:01:01	r-				
		01:01:01:02N	02:0:01:02L	68:08	69.2	01:01:01:02N	02:0:01:02L	rr				
		01:01:01:02N	02:01:01:03	11:02	74.1	01:01:01:02N	02:01:01:03	r-				
						01:01:13	02:01:02	Z				
						01:01:21	02:01:09	Z				
						01:14	02:101:01	Zrr				
						01:30	02:246	Zrr				
						01:66	02:338	Zrr				
						01:98	02:262	Zrr				
						01:100	02:346	Zrr				
						01:104	02:90	Zrr				
						02:36	36:04	Z-r				
	HLA-B	18:01:01:01	57:01:01	57:01	304	18:01:01:01	57:01:01		18:01		18:01	57:01
		18:01:01:02	57:01:01	18:01:01:02	216.8	18:01:01:02	57:01:01					
				18:03	218.1	18:17N	57:01:01	Sr-				
						18:36	57:26	Zrr				
						18:49	57:60	Zr-				
						18:68	57:10	Zrr				
	HLA-C	07:01:01:01	07:01:01:01	07:01	146.2	07:01:01:01	07:01:01:01				07:WTR	07:WTR
		07:01:01:01	07:01:01:02			07:01:01:01	07:01:01:02					
		07:01:01:02	07:01:01:02			07:01:01:01	07:01:02					
						07:01:01:01	07:06					
						07:01:01:01	07:18					
						07:01:01:02	07:01:01:02					
						07:01:01:02	07:01:02					
						07:01:01:02	07:06					
						07:01:01:02	07:18					
						07:01:02	07:01:02					
						07:01:02	07:06					
						07:01:02	07:18					
						07:06	07:18					
						07:18	07:18					
	HLA-DRB1	03:01:01:01	07:01:01:01	07:01	273	03:01:01:01	07:01:01:01				03:01	07:01
		03:01:01:01	07:01:01:02	03:01	231.2	03:01:01:01	07:01:01:02					

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
		03:01:01:02	07:01:01:01	08:03	56.4	03:01:01:02	07:01:01:01					
		03:01:01:02	07:01:01:02			03:01:01:02	07:01:01:02					
						03:01:08	07:01:01:01	S				
						03:01:08	07:01:01:02	S				
	HLA-DQB1	02:01:01	03:03:02:01	02:01	223	02:01:01	03:03:02:01				02:01	03:03
		02:01:01	03:03:02:02	03:03	164.6	02:01:01	03:03:02:02					
		02:01:01	03:03:02:03			02:01:01	03:03:02:03					
						02:03	03:02:01	Z				
HG01872	HLA-A	11:02:01	24:07	11:02	300.3	11:02:01	24:07				11:PVND	24:07
		11:77[a]	24:07	11:01	293.3	11:110	24:07	Sr-				
				11:50	271.5							
				24:08	235.2							
				24:07	235.2							
				24:02	235.2							
				24:20	235.2							
	HLA-B	27:04:01	39:05:01	27:04	174.1	27:04:01	39:05:01				27:04:01	39:05:01
				27:25	91.4							
				39:34	106.1							
				40:02	90.6							
				40:06	90.6							
	HLA-C	08:01:01	12:02:02	08:21	235	08:01:01	12:02:02		08:01		08:01	12:02
				08:01	187.5	08:01:02	12:02:03	Z				
				12:03	176.8	08:01:06	12:02:07	Z				
				12:02	176.8	08:21	12:49	Zrr				
						08:22	12:02:02	r-				
	HLA-DRB1	08:03:02	12:02:01	07:01	114.2	08:03:02	12:02:01				08:03:02	12:02:01
				08:03	222.5							
				12:01	137.4							
				14:05	34.8							
				14:54:01	34.8							
	HLA-DQB1	06:01:01	03:01:01:01	06:01	192.3	06:01:01	03:01:01:01				06:01	03:01
		06:01:01	03:01:01:02	03:01	181	06:01:01	03:01:01:02					
		06:01:01	03:01:01:03			06:01:01	03:01:01:03					
HG01873	HLA-A	02:03:01	03:01:01:01	03:01	229.9	02:03:01	03:01:01:01			03:01	02:03	03:XKS
		02:03:01	03:01:01:03	02:03	110.2	02:03:01	03:01:01:02N	-r				
						02:03:01	03:01:01:03					
						02:171:02	03:50	Z-r				
	HLA-B	35:03:01	55:02:01	35:02	140.6	35:03:01	55:02:01				35:03:01	55:02:01
				35:14	140.6	35:06	55:46	Z-r				
				35:41	140.6							
				35:01	140.6							
				35:03	104.8							
				56:01	156.6							
				55:02	147.1							
				55:01	147.1							
				55:48	147.1							
				55:24	147.1							
				37:01	48.3							
	HLA-C	04:01:01:01	12:03:01:01	04:01	188	04:01:01:01	12:03:01:01		04:01:01:01		04:01	12:03
		04:01:01:02	12:03:01:01	04:03	150.4	04:01:01:02	12:03:01:01		04:01:01:02			
		04:01:01:03	12:03:01:01	12:03	122.5	04:01:01:03	12:03:01:01		04:01:01:03			
		04:01:01:04	12:03:01:01	01:08	97.8	04:01:01:04	12:03:01:01		04:01:01:04			
		04:01:01:05	12:03:01:01	01:03	97.8	04:01:01:05	12:03:01:01		04:01:01:05			
		04:01:01:01	12:03:01:02	01:30	97.8	04:01:01:01	12:03:01:02		04:01:01:01			
		04:01:01:02	12:03:01:02	01:02	97.8	04:01:01:02	12:03:01:02		04:01:01:02			
		04:01:01:03	12:03:01:02	15:17	78.8	04:01:01:03	12:03:01:02		04:01:01:03			
		04:01:01:04	12:03:01:02	15:02	78.8	04:01:01:04	12:03:01:02		04:01:01:04			
		04:01:01:05	12:03:01:02	15:16	78.8	04:01:01:05	12:03:01:02		04:01:01:05			
				15:05	78.8	04:01:05	12:03:20	Z	04:01:05			
						04:09N	12:03:01:01	S				
						04:09N	12:03:01:02	S				
						04:30	12:03:01:01	Sr-				
						04:30	12:03:01:02	Sr-				
						04:82	12:03:01:01	r-				
						04:82	12:03:01:02	r-				
						04:94:01	12:24	Zrr				

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
	HLA-DRB1	08:02:01	14:05:01	14:05	253.9	08:02:01	14:05:01				08:02:01	14:05:01
				08:03	180.1							
				07:01	89.1							
				04:03	32.7							
	HLA-DQB1	04:02:01	05:03:01:01	05:03	134.6	04:02:01	05:03:01:01				04:02	05:03
		04:02:01	05:03:01:02	03:03	50.5	04:02:01	05:03:01:02					
				03:02	52.3							
				03:05	52.3							
HG01886	HLA-A	30:02:01	74:01	30:02	150.4	30:02:01	74:01				30:02	74:AB
				30:01	119.5	30:02:01	74:02:01:01					
				30:04	119.5	30:02:01	74:02:01:02					
				32:01	95.4							
				74:01	95.4							
				74:11	103.9							
	HLA-B	15:03:01	57:03:01	57:01	267.8	15:03:01	57:03:01				15:MJMN	57:03
				57:11	267.8	15:103	57:03:01	Sr				
				57:06	267.8	15:151	57:57	Zrr				
				15:03	129.8	15:220	57:03:01	r-				
	HLA-C	02:10	07:01:02	07:01	126	02:02:05	07:01:07		02:10		02:10	07:WTR
				02:11	127.4	02:10	07:01:01:01	Z				
				02:02	127.4	02:10	07:01:01:02	Z				
						02:10	07:01:02	Z				
						02:10	07:06	Z				
						02:10	07:18	Z				
						02:12	07:210	Zrr				
						02:16:01	07:22	Zr-				
						02:27:02	07:09	Z-r				
	HLA-DRB1	11:01:02	13:02:01	13:02	144.9	11:01:02	13:02:01				11:01:02	13:02:01
				07:01	112.3	11:09	13:23:01	Z-r				
				11:01	123.6	11:14:01	13:05:02	Z				
				08:03	62.8	11:14:02	13:05:01	Z				
						11:19:03	13:63	Z-r				
						11:20	13:14:03	Zr-				
	HLA-DQB1	05:02:01	06:09	06:09	160.1	05:02:01	06:09				05:02:01	06:09
				05:03	140.8							
				05:01	112.4							
HG01953	HLA-A	02:01:01:01	02:11:01	02:01	120	02:01:01:01	02:11:01				02:01	02:11
		02:01:01:02L	02:11:01			02:01:01:02L	02:11:01	r-				
		02:01:01:03	02:11:01			02:01:01:03	02:11:01					
	HLA-B	15:04	35:05:01	15:07	159.9	15:04	35:05:01				15:04	35:05:01
				15:32	159.9	15:07:01	35:37	Z-r				
				15:01	159.9							
				58:01	156.1							
				35:14	244							
	HLA-C	01:02:01	04:01:01:01	01:02	150.4	01:02:01	04:01:01:01				01:02	04:JERF
		01:02:01	04:01:01:02	04:03	142.8	01:02:01	04:01:01:02					
		01:02:01	04:01:01:03	04:06	142.8	01:02:01	04:01:01:03					
		01:02:01	04:01:01:04	04:01	174.4	01:02:01	04:01:01:04					
		01:02:01	04:01:01:05			01:02:01	04:01:01:05					
						01:02:01	04:09N	S				
						01:02:01	04:30	S-r				
						01:02:01	04:82	-r				
						01:02:04	04:01:11	Z				
						01:02:06	04:01:08	Z				
						01:02:11	04:01:01:01					
						01:02:11	04:01:01:02					
						01:02:11	04:01:01:03					
						01:02:11	04:01:01:04					
						01:02:11	04:01:01:05					
						01:02:11	04:09N	S				
						01:02:11	04:30	S-r				
						01:02:11	04:82	-r				
						01:14	04:29	Zrr				
						01:17	04:10	Z				
						01:58	04:33	Zrr				
	HLA-DRB1	04:11:01	09:01:02	07:01	154.6	04:11:01	09:01:02				04:11:01	09:01:02

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
				04:03	169.3							
	HLA-DQB1	03:02:01	03:03:02:01	03:02	113	03:02:01	03:03:02:01				03:02	03:03
		03:02:01	03:03:02:02			03:02:01	03:03:02:02					
		03:02:01	03:03:02:03			03:02:01	03:03:02:03					
HG01968	HLA-A	02:01:01:01	68:01:02	68:01	198.7	02:01:01:01	68:01:02				02:01	68:FKZ
		02:01:01:02L	68:01:02	02:48	114.5	02:01:01:02L	68:01:02	r-				
		02:01:01:03	68:01:02	02:01	114.5	02:01:01:03	68:01:02					
						02:01:01:01	68:11N	S-r				
						02:01:01:02L	68:11N	Srr				
						02:01:01:03	68:11N	S-r				
						02:04	68:23	Z				
						02:22:01	68:08:02	Z				
						02:34	68:03:01	Z				
						02:35:01	68:05	Z				
						02:217:02	68:75	Z-r				
						02:245	68:83	Zrr				
	HLA-B	07:02:01	40:02:01	40:02	256.3	07:02:01	40:02:01				07:TDVB	40:02
				40:06	256.3	07:02:03	40:02:04	Z				
				07:02	154.1	07:05:01	40:18	Z-r				
						07:06	40:18	Z-r				
						07:61	40:02:01	Sr-				
						07:91	40:50	Zr-				
						07:143	40:40	Zr-				
						07:161N	40:02:01	r-				
	HLA-C	03:04:01:01	07:02:01:01	03:03	100.9	03:04:01:01	07:02:01:01		03:04		03:04	07:WCP
		03:04:01:01	07:02:01:02	03:04	100.9	03:04:01:01	07:02:01:02					
		03:04:01:01	07:02:01:03	07:02	124.3	03:04:01:01	07:02:01:03					
		03:04:01:02	07:02:01:01			03:04:01:01	07:50	S-r				
		03:04:01:02	07:02:01:02			03:04:01:02	07:02:01:01					
		03:04:01:02	07:02:01:03			03:04:01:02	07:02:01:02					
						03:04:01:02	07:02:01:03					
						03:04:01:02	07:50	S-r				
						03:04:19	07:02:04	Z				
						03:04:26	07:02:10	Z				
						03:04:27	07:02:32	Z				
						03:32	07:51	Zrr				
						03:35	07:10	Zr-				
						03:38:02	07:39	Z-r				
						03:40:01	07:29	Z				
						03:41	07:133	Zrr				
						03:64:02	07:127	Z-r				
						03:90	07:27:01	Zr-				
						03:92	07:186	Zrr				
	HLA-DRB1	01:03	09:01:02	07:01	158.5	01:03	09:01:02				01:03	09:01:02
				01:01	168.5							
	HLA-DQB1	03:03:02:01	05:01:01:01	05:01	118	03:03:02:01	05:01:01:01				03:03	05:01
		03:03:02:02	05:01:01:01	03:03	118.1	03:03:02:02	05:01:01:01					
		03:03:02:03	05:01:01:01			03:03:02:03	05:01:01:01					
		03:03:02:01	05:01:01:02			03:03:02:01	05:01:01:02					
		03:03:02:02	05:01:01:02			03:03:02:02	05:01:01:02					
		03:03:02:03	05:01:01:02			03:03:02:03	05:01:01:02					
HG02014	HLA-A	02:01:01:01	36:01	01:01	246.7	02:01:01:01	36:01				02:01	36:01
		02:01:01:02L	36:01	36:01	195.1	02:01:01:02L	36:01	r-				
		02:01:01:03	36:01	02:01	269.3	02:01:01:03	36:01					
	HLA-B	35:01:01:01	40:01:02	40:79	286.4	35:01:01:01	40:01:01				35:TDS	40:01
		35:01:01:02	40:01:02	40:01	286.4	35:01:01:01	40:01:02					
				07:41	188	35:01:01:01	40:88	Z-r				
				07:02	188	35:01:01:02	40:01:01					
				07:33	188	35:01:01:02	40:01:02					
				07:50	188	35:01:01:02	40:88	Z-r				
				53:01	243.2	35:01:04	40:01:04	Z				
				35:02	243.2	35:01:17	40:01:17	Z				
				35:41	243.2	35:01:23	40:01:01					
				35:01	243.2	35:01:23	40:01:02					
						35:01:23	40:88	Z-r				
						35:04:01	40:52	Z-r				

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
						35:10	40:25	Z-r				
						35:15	40:63	Zrr				
						35:20:01	40:07	Z-r				
						35:28	40:106	Z-r				
						35:20:01	40:07	Z-r				
						35:28	40:106	Z-r				
						35:34	40:38	Z-r				
						35:42:01	40:01:01	r-				
						35:42:01	40:01:02	r-				
						35:42:01	40:88	Zrr				
						35:64:02	40:87:02	Z				
						35:68:01	40:160:01	Zrr				
						35:68:02	40:160:02	Z				
						35:119	40:01:01	Zr-				
						35:119	40:01:02	Zr-				
						35:119	40:88	Zrr				
	HLA-C	03:04:01:01	04:01:01:01	04:01	260.4	03:04:01:01	04:01:01:01				03:04	04:JERF
		03:04:01:01	04:01:01:02	04:03	221	03:04:01:01	04:01:01:02					
		03:04:01:01	04:01:01:03	04:06	221	03:04:01:01	04:01:01:03					
		03:04:01:01	04:01:01:04	03:02	213.3	03:04:01:01	04:01:01:04					
		03:04:01:01	04:01:01:05	03:04	213.3	03:04:01:01	04:01:01:05					
		03:04:01:02	04:01:01:01	15:17	36.2	03:04:01:01	04:09N	S				
		03:04:01:02	04:01:01:02	15:02	36.2	03:04:01:01	04:30	S-r				
		03:04:01:02	04:01:01:03	15:16	36.2	03:04:01:01	04:82	-r				
		03:04:01:02	04:01:01:04	15:05	36.2	03:04:01:02	04:01:01:01					
		03:04:01:02	04:01:01:05			03:04:01:02	04:01:01:02					
						03:04:01:02	04:01:01:03					
						03:04:01:02	04:01:01:04					
						03:04:01:02	04:01:01:05					
						03:04:01:02	04:09N	S				
						03:04:01:02	04:30	S-r				
						03:04:01:02	04:82	-r				
						03:04:25	04:01:23	Z				
						03:04:26	04:01:39	Z				
						03:07	04:29	Z-r				
						03:10	04:114	Z-r				
						03:28	04:04:01	Zr-				
						03:32	04:33	Zrr				
						03:35	04:110	Zrr				
						03:38:02	04:10	Z				
						03:92	04:08	Zr-				
						03:98	04:85	Zrr				
						03:163	04:36	Zrr				
	HLA-DRB1	01:01:01	15:01:01:01	07:01	104.4	01:01:01	15:01:01:01				01:01	15:01
		01:01:01	15:01:01:02	15:01	197.9	01:01:01	15:01:01:02					
		01:01:01	15:01:01:03	01:01	209.9	01:01:01	15:01:01:03					
		01:01:01	15:01:01:04	01:02	209.9	01:01:01	15:01:01:04					
						01:01:01	15:01:17					
	HLA-DQB1	05:01:01:01	06:02:01	06:02	218.5	05:01:01:01	06:02:01				05:01	06:02
		05:01:01:02	06:02:01	05:01	222.2	05:01:01:02	06:02:01					
HG02057	HLA-A	02:03:01	31:01:02	31:01	269.3	02:03:01	31:01:02				02:03:01	31:01:02
				02:03	148.3	02:171:02	31:30	Z-r				
				32:01	52.7							
				74:01	52.7							
				74:11	52.7							
				11:02	38.9							
	HLA-B	13:01:01	48:01:01	48:01	224.8	13:01:01	48:01:01				13:01	48:01
				13:01	152.2	13:01:01	48:03:01	Z				
						13:01:01	48:03:02	Z				
						13:01:01	48:21	Z-r				
						13:01:05	48:01:01					
						13:01:05	48:03:01	Z				
						13:01:05	48:21	Z-r				
						13:02:01	48:01:01					
						13:02:01	48:03:01	Z				
						13:02:01	48:03:02	Z				

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
						13:02:01	48:21	Z-r				
	HLA-C	03:03:01	03:04:04[b]	03:04	104.1	03:03:01	03:04:21				03:BPSK	03:04
				03:03	110.2	03:04:21	03:20N	S-r				
	HLA-DRB1	11:01:01	13:12:01	11:01	190.6	11:01:01	13:12:01			13:12	11:CTPB	13:12
				07:01	99.1	11:01:08	13:12:01					
				08:03	35.4	11:19:01	13:49	Z				
						11:37:01	13:30	rr				
						11:97	13:12:01	Sr-				
	HLA-DQB1	03:01:01:01	03:01:01:01	03:01	157.6	03:01:01:01	03:01:01:01				03:01	03:01
		03:01:01:01	03:01:01:02			03:01:01:01	03:01:01:02					
		03:01:01:01	03:01:01:03			03:01:01:01	03:01:01:03					
		03:01:01:02	03:01:01:02			03:01:01:02	03:01:01:02					
		03:01:01:02	03:01:01:03			03:01:01:02	03:01:01:03					
		03:01:01:03	03:01:01:03			03:01:01:03	03:01:01:03					
NA18507	HLA-A	23:01:01	30:01:01	30:01	313.4	23:01:01	30:01:01		23:01		23:01	30:01
				23:01	222.3	23:17	30:01:01	S				
				31:01	32.4	23:21	30:11:02	Zr-				
				31:04	32.4							
				74:01	31.8							
				74:11	31.8							
				34:01	47.3							
	HLA-B	15:03:01	42:01:01	42:01	250.8	15:03:01	42:01:01		15:03		15:03:01	42:01:01
				15:18	181.7	15:103	42:01:01	Sr-	15:103		15:103	42:01:01
				15:03	190.9	15:220	42:01:01	r-				
	HLA-C	02:10	17:01:01:01	17:01	200.4	02:10	17:01:01:01				02:10	17:MN
		02:10	17:01:01:02	02:11	120.2	02:10	17:01:01:02					
				02:02	120.2	02:10	17:02	-r				
				15:17	59.5	02:10	17:03					
				15:02	59.5							
				15:05	59.5							
	HLA-DRB1	08:04:01	13:02:01	08:03	171.6	08:04:01	13:02:01				08:04:01	13:02:01
				13:02	251.8							
				11:01	126.7							
				11:04	126.7							
				07:01	122.7							
	HLA-DQB1	03:01:04	06:09	06:09	168.3	03:01:04	06:09				03:01:04	06:09
				03:01	153.9	03:01:04	06:05:01	Z				
NA19129	HLA-A	02:02	23:01:01	23:01	330.2	02:02	23:01:01			23:01	02:02	23:01
				02:05	146.4	02:02	23:17	S				
				02:02	152.4	02:102	23:39	Zrr				
	HLA-B	07:02:01	58:01:01	07:02	230.6	07:02:01	58:01:01		07:02		07:02	58:01
				58:01	243.5	07:02:13	58:01:05	Z				
						07:18:01	58:18	Zrr				
						07:24	58:37	Zrr				
						07:26	58:22	Zrr				
						07:61	58:01:01	Sr-				
	HLA-C	07:02:01:01	07:18	07:01	249.6	07:01:01	07:02:01:01		07:02	07:18	07:02	07:18
		07:02:01:02	07:18	07:02	230.7	07:01:01	07:02:01:02					
		07:02:01:03	07:18	07:19	162.8	07:01:01	07:02:01:03					
						07:01:01	07:50	S-r				
						07:01:02	07:02:01:01					
						07:01:02	07:02:01:02					
						07:01:02	07:02:01:03					
						07:01:02	07:50	S-r				
						07:02:01:01	07:06					
						07:02:01:01	07:18					
						07:02:01:02	07:06					
						07:02:01:02	07:18					
						07:02:01:03	07:06					
						07:02:01:03	07:18					
						07:06	07:50	S-r				
						07:18	07:50	S-r				
						07:19	07:27:01	Z				
	HLA-DRB1	01:01:01	11:02:01	11:04	224.5	01:01:01	11:02:01				01:01:01	11:02:01
				01:01	203.2							
				01:02	189.1							

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
				07:01	91.6							
	HLA-DQB1	03:19	05:01:01:01	05:01	196.5	03:19	05:01:01:01				03:19	05:01
		03:19	05:01:01:02	03:01	153.2	03:19	05:01:01:02					
NA19240[d]	HLA-A	30:01:01	68:02:01:01	68:02	217.9	30:01:01	68:02:01:01				30:01	68:02
		30:01:01	68:02:01:02	30:01	209.3	30:01:01	68:02:01:02					
		30:01:01	68:02:01:03	11:02	46.6	30:01:01	68:02:01:03					
				34:01	46.8							
	HLA-B	35:01:01:01	57:03:01	57:01	177.3	35:01:01:01	57:03:01		35:01	57:03	35:01	57:03
		35:01:01:02	57:03:01	35:02	122.6	35:01:01:02	57:03:01					
				35:41	122.6	35:01:23	57:03:01					
				35:01	122.6	35:04:01	57:01:01	Z				
				35:14	122.6	35:04:01	57:21	Z-r				
				35:03	122.6	35:04:01	57:40	Z-r				
				35:08	122.6	35:42:01	57:03:01	r-				
				35:05	122.6	35:68:01	57:03:01	r-				
				53:01	113.1	35:119	57:03:01	r-				
	HLA-C	04:01:01:01	18:02	04:01	215.4	04:01:01:01	18:01		04:01	18:02	04:01	18:02
		04:01:01:02	18:02	18:01	155	04:01:01:01	18:02					
		04:01:01:03	18:02			04:01:01:02	18:01					
		04:01:01:04	18:02			04:01:01:02	18:02					
		04:01:01:05	18:02			04:01:01:03	18:01					
						04:01:01:03	18:02					
						04:01:01:04	18:01					
						04:01:01:04	18:02					
						04:01:01:05	18:01					
						04:01:01:05	18:02					
						04:09N	18:01	S				
						04:09N	18:02	S				
						04:30	18:01	Sr-				
						04:30	18:02	Sr-				
						04:82	18:01	r-				
						04:82	18:02	r-				
	HLA-DRB1	12:01:01	16:02:01	16:02	219.9	12:01:01	16:02:01				12:DUKV	16:02
				12:01	297.2	12:06	16:02:01	Sr-				
				15:01	106.6	12:10	16:02:01	Sr-				
				15:03	106.6	12:17	16:02:01	Sr-				
				07:01	90.5							
				14:05	34.8							
				14:54:01	34.8							
				08:03	39.4							
	HLA-DQB1	03:01:01:01	05:02:01	03:01	162.6	03:01:01:01	05:02:01				03:01	05:02
		03:01:01:02	05:02:01	05:03	70.2	03:01:01:02	05:02:01					
		03:01:01:03	05:02:01	05:01	41.6	03:01:01:03	05:02:01					
NA20313[d]	HLA-A	03:01:01:01	68:02:01:01	NR		03:01:01:01	68:02:01:01				03:XKS	68:02
		03:01:01:01	68:02:01:02			03:01:01:01	68:02:01:02					
		03:01:01:01	68:02:01:03			03:01:01:01	68:02:01:03					
		03:01:01:03	68:02:01:01			03:01:01:02N	68:02:01:01	r-				
		03:01:01:03	68:02:01:02			03:01:01:02N	68:02:01:02	r-				
		03:01:01:03	68:02:01:03			03:01:01:02N	68:02:01:03	r-				
						03:01:01:03	68:02:01:01					
						03:01:01:03	68:02:01:02					
						03:01:01:03	68:02:01:03					
						03:07	68:54	Z-r				
	HLA-B	35:01:01:01	53:01:01	NR		35:01:01:01	53:01:01		35:01	53:01	35:01	53:01
		35:01:01:02	53:01:01			35:01:01:02	53:01:01		35:42	53:01	35:42	53:01
						35:01:23	53:01:01					
						35:27	53:05					
						35:42:01	53:01:01	r				
	HLA-C	04:01:01:01	04:01:01:01	04:01	38.3	04:01:01:01	04:01:01:01				04:JERF	04:JERF
		04:01:01:01	04:01:01:02			04:01:01:01	04:01:01:02					
		04:01:01:01	04:01:01:03			04:01:01:01	04:01:01:03					
		04:01:01:01	04:01:01:04			04:01:01:01	04:01:01:04					
		04:01:01:01	04:01:01:05			04:01:01:01	04:01:01:05					
		04:01:01:02	04:01:01:02			04:01:01:01	04:09N					
		04:01:01:02	04:01:01:03			04:01:01:01	04:30					
		04:01:01:02	04:01:01:04			04:01:01:01	04:82					

Supplementary Table 3. HLA typing results using exome-seq data and the laboratory validation

		Exome reads				Laboratory validation						
		ATHLATES		HLAminer*		SBT#			Additional typing‡		Final type assigned	
Sample	Gene	Allele 1	Allele 2	Alleles	Confidence	Allele 1	Allele 2	Note†	Allele 1	Allele 2	Allele 1	Allele 2
		04:01:01:02	04:01:01:05			04:01:01:02	04:01:01:02					
		04:01:01:03	04:01:01:03			04:01:01:02	04:01:01:03					
		04:01:01:03	04:01:01:04			04:01:01:02	04:01:01:04					
		04:01:01:03	04:01:01:05			04:01:01:02	04:01:01:05					
		04:01:01:04	04:01:01:04			04:01:01:02	04:09N					
		04:01:01:04	04:01:01:05			04:01:01:02	04:30	r				
		04:01:01:05	04:01:01:05			04:01:01:02	04:82	r				
						04:01:01:03	04:01:01:03					
						04:01:01:03	04:01:01:04					
						04:01:01:03	04:01:01:05					
						04:01:01:03	04:09N					
						04:01:01:03	04:30	r				
						04:01:01:03	04:82	r				
						04:01:01:04	04:01:01:04					
						04:01:01:04	04:01:01:05					
						04:01:01:04	04:09N					
						04:01:01:04	04:30	r				
						04:01:01:05	04:01:01:05					
	HLA-DRB1	04:05:01	08:04:01[c]	NR		04:05:01	08:04:02				04:05:01	08:04:02
	HLA-DQB1	03:01:04	03:02:01	NR		03:01:04	03:02:01				03:01:04	03:02:01
						03:03:04	03:04	Z				

*HLAminer reports individual HLA alleles with confidence ≥ 30; no allelic pair is inferred by the program; NR, nothing reported.

#SBT, sequence based typing by Sanger's method. Results after the first round of sequencing are listed.

†r, rare allele in the population; Z, group specific primer (Z primer) available to sequence particular haplotypes; S denotes allelic pairs that do not need to be ruled out. The shaded allelic pairs were ruled out using Z primers during the second round of sequencing.

‡Performed by SSP and/or SSOP methods. SSP, sequence specific primers; SSOP, sequence specific oligonucleotide probes.

[a]A*11:77 is equally supported and cannot be ruled out.

[b]There is no read support for C*03:04:21.

[c]There is no read support for DRB1*08:04:02.

[d]The typing results for NA19240R and NA20313R are not shown. The typing by ATHLATES is completely concordant with NA19240 and NA20313, respectively; the typing by HLAminer does not completely agree with NA19240 and NA20313, respectively.

Supplementary Table 4. A sample report generated by ALTHLATES (for HG01757, HLA-DQB1)

Name	HD	Aln_len	cDNA_len	Similarity	Avg_cov	Missing Exons (ID, len) ; mismatches [ID, pos]
DQB1*03:03:02:03	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*03:03:02:02	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*03:03:02:01	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*02:01:01	0	772	786	1	144.622	(5, 0) (6, 14)
DQB1*02:02	1	772	786	0.998705	144.622	(5, 0) (6, 14) [3, 121]
DQB1*03:02:01	1	772	786	0.998705	134.253	(5, 0) (6, 14) [2, 157]
DQB1*03:31	1	618	618	0.998382	147.324	(1, 0) (5, 0) (6, 0) [3, 63]
DQB1*03:43	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [3, 141]
DQB1*03:41	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [2, 120]
DQB1*03:39	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [3, 98]
DQB1*03:38	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [2, 75]
DQB1*03:30	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [2, 13]
DQB1*03:03:04	1	552	552	0.998188	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [3, 224]
DQB1*02:01:04	1	552	552	0.998188	173.219	(1, 0) (4, 0) (5, 0) (6, 0) [2, 194]
DQB1*02:01:05	1	552	552	0.998188	173.219	(1, 0) (4, 0) (5, 0) (6, 0) [2, 206]
DQB1*02:07	1	552	552	0.998188	173.219	(1, 0) (4, 0) (5, 0) (6, 0) [2, 229]
DQB1*02:04	1	552	552	0.998188	173.219	(1, 0) (4, 0) (5, 0) (6, 0) [3, 123]
DQB1*03:33	1	522	522	0.998084	149.414	(1, 0) (4, 0) (5, 0) (6, 0) [3, 156]
DQB1*03:34	1	522	522	0.998084	149.414	(1, 0) (4, 0) (5, 0) (6, 0) [2, 45]
DQB1*03:32	2	552	552	0.996377	154.324	(1, 0) (4, 0) (5, 0) (6, 0) [2, 157] [3, 138]
DQB1*02:06	2	552	552	0.996377	173.219	(1, 0) (4, 0) (5, 0) (6, 0) [3, 121] [3, 225]

---------- Candidate Allelic Pairs ----------

Name	HD	Aln_len	cDNA_len	Similarity	Avg_cov	Missing Exons (ID, len) ; mismatches [ID, pos]
DQB1*03:03:02:03	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*03:03:02:02	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*03:03:02:01	0	772	786	1	134.253	(5, 0) (6, 14)
DQB1*02:01:01	0	772	786	1	144.622	(5, 0) (6, 14)

---------- Inferred Allelic Pairs ----------

DQB1*03:03:02:03	DQB1*02:01:01	0
DQB1*03:03:02:02	DQB1*02:01:01	0
DQB1*03:03:02:01	DQB1*02:01:01	0

HD: Hamming distance.

Aln_len: alignment length, number of cDNA bases supported by contigs.

cDNA_len: total length of cDNA of an allele.

Avg_cov: average coverage for an allele.

Missing Exons (ID, len): ID, the identity of a exon not considered for calculation of Hamming distance; len, length of the indicated exon as documented in the IMGT/HLA database.

mismatches[ID, pos]: the position (pos) of a mismatch in the exon of indicated identity (ID) when compared to its best hit in contigs.

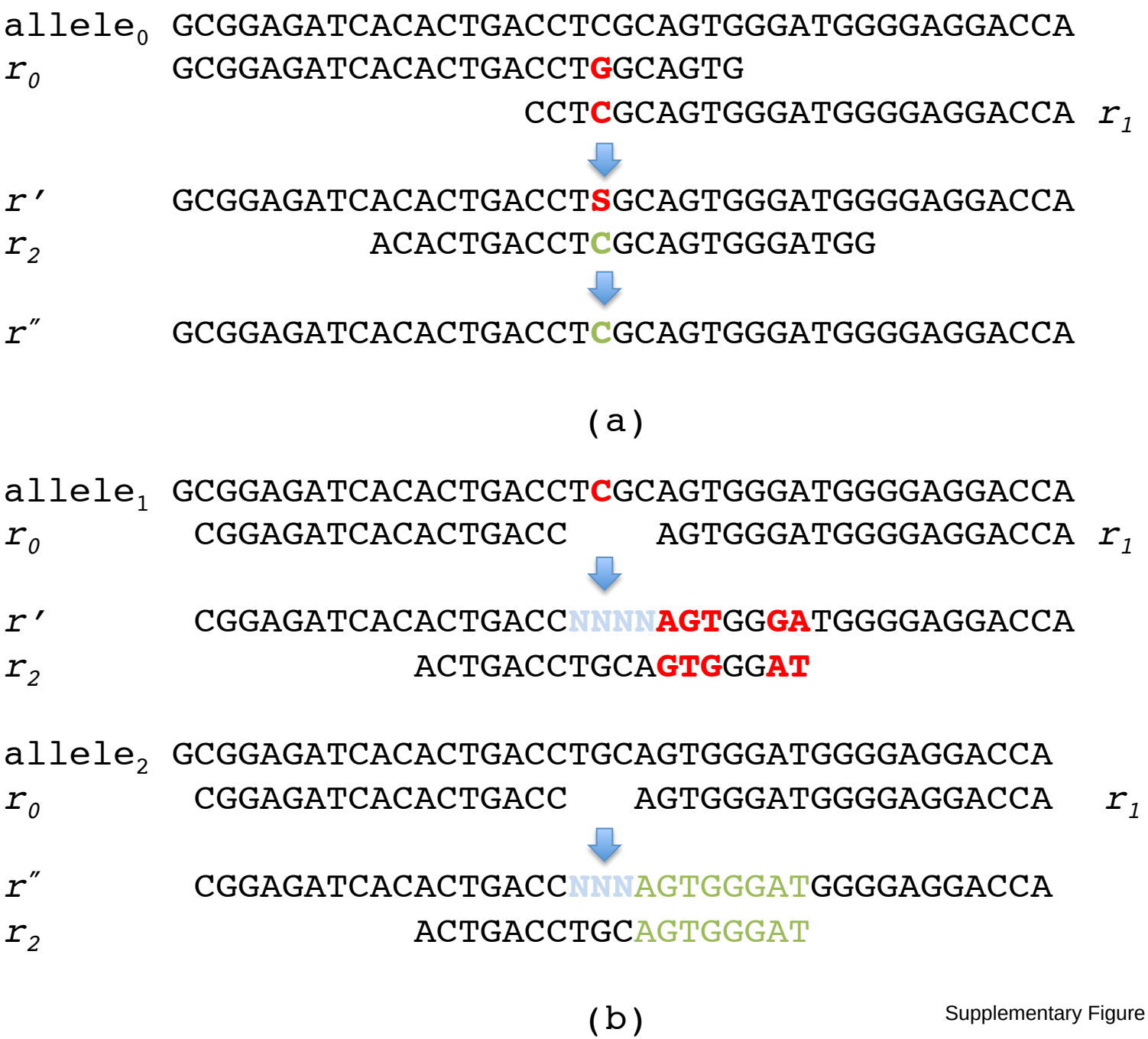
Supplementary Table 5. Statistical analysis of allelic bias within individual exons [a]

			Allele1 - Allele2											
			Exon 1		Exon 2		Exon 3		Exon 4		Exon 5		Exon 6	
Genes	Sample	Allele pair (1-2)	Z	*p*	Z	*p*	Z	*p*	Z	*p*	Z	*p*	Z	*p*
HLA-A	HG01756	A*30:02:01-A*66:01			-2.558[b]	.011	-3.328[b]	.001	-2.803[b]	.005	-1.826[b]	.068	-1.414[b]	.157
	HG01757	A*01:01:01:01-A*02:01:01:01	-1.342[b]	.180	-1.872[b]	.061	-3.717[c]	.000	-2.936[b]	.003	-1.604[b]	.109	-1.414[c]	.157
	HG01872	A*11:02:01-A*24:07	-.447[c]	.655	-3.012[b]	.003	-.094[b]	.925	-2.023[b]	.043	-2.032[c]	.042	-1.414[b]	.157
	HG01873	A*02:03:01-A*03:01:01:01	-1.342[c]	.180	-.530[c]	.596	-3.298[b]	.001	-2.934[c]	.003	-1.604[c]	.109	-1.342[b]	.180
	HG01886	A*30:02:01-A*74:01	-1.000[b]	.317	-2.809[b]	.005	-2.807[b]	.005	-2.803[b]	.005	-2.521[b]	.012	-1.414[b]	.157
	HG01953	A*02:01:01:01-A*02:11:01			-1.342[b]	.180								
	HG01968	A*02:01:01:01-A*68:01:02			-2.936[c]	.003	-1.275[b]	.202						
	HG02014	A*02:01:01:01-A*36:01	-1.342[c]	.180	-2.941[c]	.003	-3.523[b]	.000	-2.937[c]	.003	-1.604[c]	.109	-1.414[c]	.157
	HG02057	A*02:03:01-A*31:01:02	-1.633[c]	.102	-1.966[c]	.049	-1.351[c]	.177	-1.992[c]	.046	-.365[b]	.715		
	NA18507	A*23:01:01-A*30:01:01	-1.342[c]	.180	-3.749[c]	.000	-.255[c]	.798	-2.023[c]	.043	-1.572[c]	.116	-1.414[b]	.157
	NA19129	A*02:02-A*23:01:01	-1.342[c]	.180	-3.531[b]	.000	-3.063[c]	.002	-3.518[c]	.000	-2.032[b]	.042		
	NA19240	A*30:01:01-A*68:02:01:01			-2.936[b]	.003	-3.519[b]	.000	-2.981[b]	.003	-1.604[c]	.109	-1.342[c]	.180
	NA19240R	A*30:01:01-A*68:02:01:01			-1.808[b]	.071	-3.520[b]	.000	-2.904[b]	.004	-1.604[c]	.109	-1.342[c]	.180
	NA20313	A*03:01:01:01-A*68:02:01:01			-2.207[b]	.027	-3.182[c]	.001	-3.059[b]	.002	-1.604[c]	.109	-1.414[c]	.157
	NA20313R	A*03:01:01:01-A*68:02:01:01			-2.201[c]	.028	-3.189[c]	.001	-3.061[b]	.002	-1.604[c]	.109	-1.342[c]	.180
HLA-B	HG01756	B*18:01:01:01-B*41:02:01	-.850[b]	.395	-3.296[b]	.001	-2.606[b]	.009	-.447[c]	.655	.000[d]	1.000		
	HG01757	B*18:01:01:01-B*57:01:01	-1.604[b]	.109	-3.260[b]	.001	-1.844[b]	.065			-1.604[c]	.109		
	HG01872	B*27:04:01-B*39:05:01	-2.536[b]	.011	-3.301[c]	.001	-3.299[b]	.001	-.552[b]	.581	-.447[b]	.655		
	HG01873	B*35:03:01-B*55:02:01	-1.826[b]	.068	-3.062[b]	.002	-2.949[b]	.003						
	HG01886	B*15:03:01-B*57:03:01	-2.032[b]	.042	-4.376[b]	.000	-2.521[b]	.012			-1.604[c]	.109		
	HG01953	B*15:04-B*35:05:01	-1.890[b]	.059	-.966[c]	.334	-3.186[c]	.001	-1.069[c]	.285				
	HG01968	B*07:02:01-B*40:02:01	-2.524[b]	.012	-3.625[b]	.000	-.211[c]	.833						
	HG02014	B*35:01:01:01-B*40:01:02	-2.032[b]	.042	-2.936[b]	.003	-.119[b]	.906	-1.342[b]	.180	-1.826[c]	.068		
	HG02057	B*13:01:01-B*48:01:01	-2.264[b]	.024	-3.727[c]	.000	-2.416[b]	.016	-1.342[b]	.180	-1.342[b]	.180		
	NA18507	B*15:03:01-B*42:01:01	-1.841[c]	.066	-2.675[c]	.007	-2.515[c]	.012	-.447[c]	.655	-1.604[b]	.109		
	NA19129	B*07:02:01-B*58:01:01	-.322[c]	.748	-4.546[b]	.000	-2.305[b]	.021	-1.342[b]	.180	-1.826[b]	.068		
	NA19240	B*35:01:01:01-B*57:03:01			-3.064[b]	.002	-2.371[b]	.018	-.365[b]	.715	-1.604[c]	.109		
	NA19240R	B*35:01:01:01-B*57:03:01			-3.929[b]	.000	-2.205[b]	.027	-1.461[b]	.144	-.535[b]	.593		
	NA20313	B*35:01:01:01-B*53:01:01			-2.214[b]	.027								
	NA20313R	B*35:01:01:01-B*53:01:01			-2.214[b]	.027								
HLA-C	HG01756	C*05:01:01:01-C*17:01:01:01	-.677[b]	.498	-1.604[b]	.109	-.118[c]	.906	-1.362[c]	.173	-2.133[c]	.033		
	HG01872	C*08:01:01-C*12:02:02			.000[d]	1.000	-2.807[b]	.005						
	HG01873	C*04:01:01:01-C*12:03:01:01			-.238[b]	.812	-2.103[b]	.035	-2.023[c]	.043	-.535[b]	.593		
	HG01886	C*02:10-C*07:01:02	-1.342[b]	.180	-2.936[b]	.003	.000[d]	1.000	-2.936[c]	.003	-2.675[c]	.007	-.447[c]	.655
	HG01953	C*01:02:01-C*04:01:01:01			-3.061[b]	.002	-1.947[b]	.051	-1.604[b]	.109				
	HG01968	C*03:04:01:01-C*07:02:01:01	-1.841[b]	.066	-.674[c]	.500	-2.513[c]	.012	-3.059[c]	.002	-.178[c]	.859	-1.342[c]	.180
	HG02014	C*03:04:01:01-C*04:01:01:01			-3.101[c]	.002	-3.066[c]	.002	-1.069[b]	.285				
	HG02057	C*03:03:01-C*03:04:04			-1.342[c]	.180								
	NA18507	C*02:10-C*17:01:01:01	-1.841[c]	.066	-2.023[c]	.043	-1.512[c]	.130	-.140[c]	.889	-1.604[c]	.109		
	NA19240	C*04:01:01:01-C*18:02	-1.841[b]	.066	-2.823[b]	.005								
	NA19240R	C*04:01:01:01-C*18:02	-1.826[b]	.068	-2.552[b]	.011								
HLA-DRB1	HG01757	DRB1*03:01:01:01-DRB1*07:01:01:01	-1.826[c]	.068	-1.090[c]	.276	-.734[b]	.463						
	HG01872	DRB1*08:03:02-DRB1*12:02:01			-3.921[b]	.000					.000[d]	1.000		
	HG01873	DRB1*08:02:01-DRB1*14:05:01			-2.657[b]	.008								
	HG01886	DRB1*11:01:02-DRB1*13:02:01			-1.129[c]	.259								
	HG01953	DRB1*04:11:01-DRB1*09:01:02	-1.342[b]	.180	-3.232[b]	.001	-.178[c]	.859						
	HG01968	DRB1*01:03-DRB1*09:01:02	-1.604[b]	.109	-4.626[b]	.000	-1.683[c]	.092						
	HG02014	DRB1*01:01:01-DRB1*15:01:01:01			-3.826[b]	.000	-2.703[b]	.007						
	HG02057	DRB1*11:01:01-DRB1*13:12:01			-2.371[b]	.018								
	NA18507	DRB1*08:04:01-DRB1*13:02:01			-2.920[c]	.004								
	NA19129	DRB1*01:01:01-DRB1*11:02:01			-3.137[b]	.002	-1.073[b]	.283						
	NA19240	DRB1*12:01:01-DRB1*16:02:01	-1.761[b]	.078	-4.459[c]	.000	-2.134[c]	.033						
	NA19240R	DRB1*12:01:01-DRB1*16:02:01	-.674[c]	.500	-4.458[c]	.000	-2.240[c]	.025						
	NA20313	DRB1*04:05:01-DRB1*08:04:01			-.337[c]	.736	-.474[c]	.635						
	NA20313R	DRB1*04:05:01-DRB1*08:04:01			-.644[b]	.520	-.652[c]	.515						
HLA-	HG01757	DQB1*02:01:01-DQB1*03:03:02:01	-2.023[c]	.043	-3.138[b]	.002	-1.214[c]	.225	-1.342[c]	.180				

Supplementary Table 5. Statistical analysis of allelic bias within individual exons [a]

			Allele1 - Allele2									
			Exon 1		Exon 2		Exon 3		Exon 4		Exon 5		Exon 6	
Genes	Sample	Allele pair (1-2)	Z	*p*	Z	*p*	Z	*p*	Z	*p*	Z	*p*	Z	*p*
DQB1	HG01872	DQB1*06:01:01-DQB1*03:01:01:01	-.365[c]	.715	-4.108[b]	.000	-.135[c]	.893	-2.023[b]	.043				
	HG01873	DQB1*04:02:01-DQB1*05:03:01:01	-1.077[b]	.282	-4.458[c]	.000	-1.956[b]	.050	-2.366[c]	.018				
	HG01886	DQB1*05:02:01-DQB1*06:09	-1.841[c]	.066	-3.724[c]	.000	-2.366[c]	.018	-1.342[c]	.180				
	HG01968	DQB1*03:03:02:01-DQB1*05:01:01:01	-2.673[c]	.008	-4.763[c]	.000	-.445[b]	.656	-1.992[c]	.046				
	HG02014	DQB1*05:01:01:01-DQB1*06:02:01	-1.841[c]	.066	-3.221[b]	.001	-2.023[c]	.043	-1.342[b]	.180				
	NA18507	DQB1*03:01:04-DQB1*06:09			-3.982[c]	.000	-1.836[c]	.066						
	NA19129	DQB1*03:19-DQB1*05:01:01:01			-4.369[c]	.000	-1.957[c]	.050						
	NA19240	DQB1*03:01:01:01-DQB1*05:02:01	-1.367[c]	.172	-2.853[c]	.004	-2.401[b]	.016	-2.527[c]	.012				
	NA19240R	DQB1*03:01:01:01-DQB1*05:02:01	-2.677[c]	.007	-2.779[c]	.005	-2.223[b]	.026	-2.524[c]	.012				
	NA20313	DQB1*03:01:04-DQB1*03:02:01			-2.201[c]	.028	-1.826[c]	.068						
	NA20313R	DQB1*03:01:04-DQB1*03:02:01			-2.201[c]	.028	-1.826[c]	.068						

[a]Wilcoxon Signed Ranks Test was performed to compare fold coverages of heterozygous alleles at positions where they differ within each exons; p values lower than 0.05 are shaded in red, with the smallest value shaded with the darkest red; homozygous cases are omitted. [b]Allele1 has higher coverage (based on negative ranks). [c]Allele2 has higher coverage (based on positive ranks). [d]The sum of negative ranks equals the sum of positive ranks.



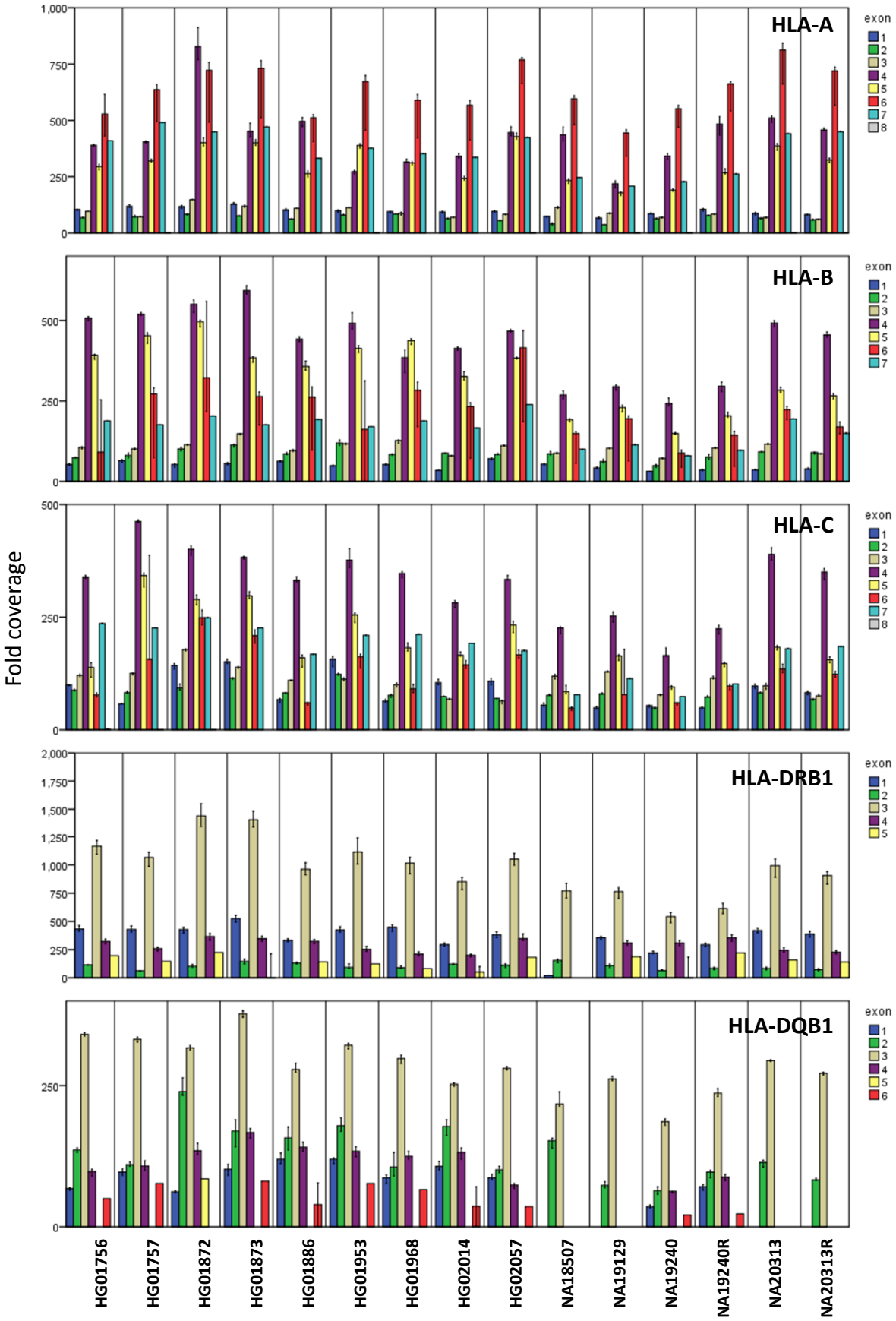
Supplementary Figure 1

```
A*02:01:01:01   GCTCTCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCT
A*02:11:01      GCTCTCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCT
Contig i        GCTCTCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCT
Contig j        GCTCTCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGAGCCCCGCTTCATCGCAGTGGGCT

A*02:01:01:01   ACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAG
A*02:11:01      ACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAG
Contig i        ACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAG
Contig j        ACGTGGACGACACGCAGTTCGTGCGGTTCGACAGCGACGCCGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAG

A*02:01:01:01   AGCAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTGG
A*02:11:01      AGCAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGATTGACCGAGTGGACCTGG
Contig i        AGCAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTGG
Contig j        AGCAGGAGGGTCCGGAGTATTGGGACGGGGAGACACGGAAAGTGAAGGCCCACTCACAGATTGACCGAGTGGACCTGG

A*02:01:01:01   GGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCG
A*02:11:01      GGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCG
Contig i        GGACCCTGCGCGGCTACTACAACCAGAGCGAGGCCG
Contig j        GGACCCTGCGCGGCT--------------------
```

(a)

```
C*04:01:01:01   CCAGCAACAGTGCCCAGGGCTCTGATGAGTCTCTCATCGCTTGTAAAG
A*04:30         CCACCAACAGTGCCCAGGGCTCTGATGAGTCTCTCATCGCTTGTAAAG
Contig i        CCAGCAACAGTGCCCAGGGCTCTGATGAGTCTCTCATCGCTTGTAAAG
```

(b)



Supplementary Figure 3

$allele_0$ GCGGAGATCACACTGACCTCGCAGTGGGATGGGGAGGACCA
$allele_1$ GCGGAGATCACACTGACCTGGCAGTGGGATGGGGAGGACCA

$Contig_a$ GCGGAGATCACACTGACCTCGCAGTGGGATGGGGAGGACCA
$Contig_b$ GCGGAGATCACACTGACCNNNCAGTGGGATGGGGAGGACCA

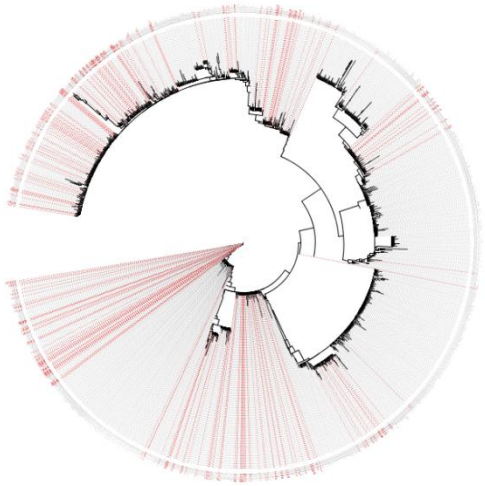
(a)

Candidate Alleles	Exon1	Exon2	Exon3	Exon4
B*55:02:01	A	A	A	A
B*56:11	A	A	B	A
B*35:03:01	B	B	B	B
B*35:60	B	B	A	A

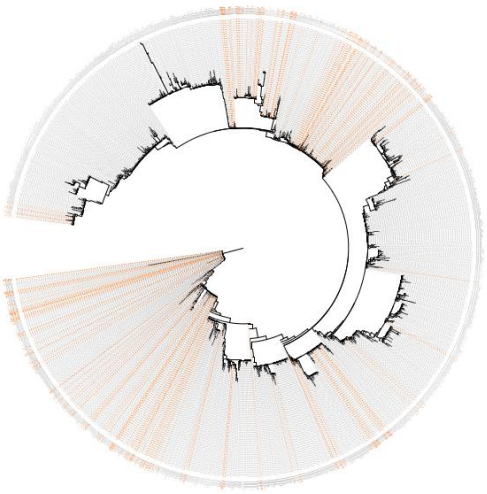
(b)

Supplementary Figure 4

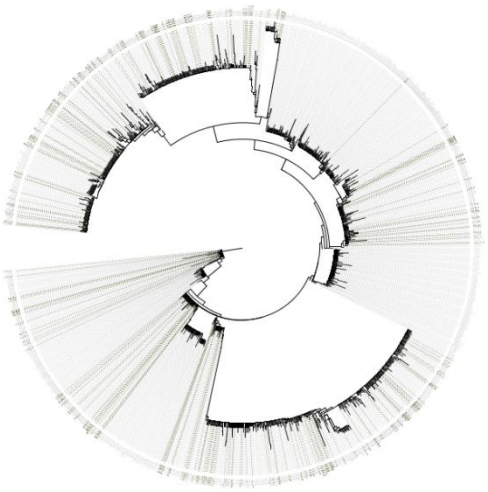
HLA-A



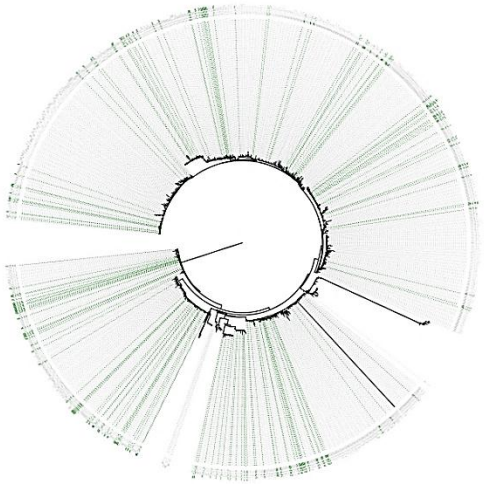
HLA-B



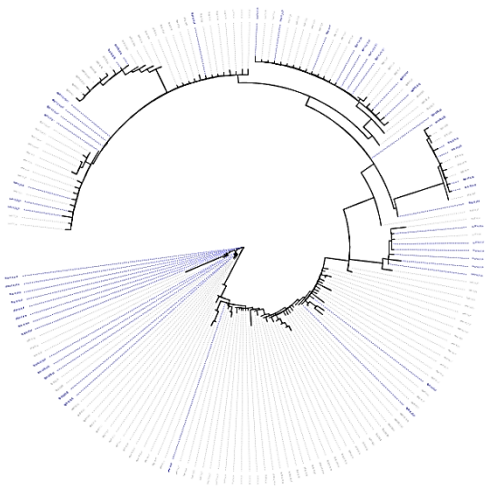
HLA-C



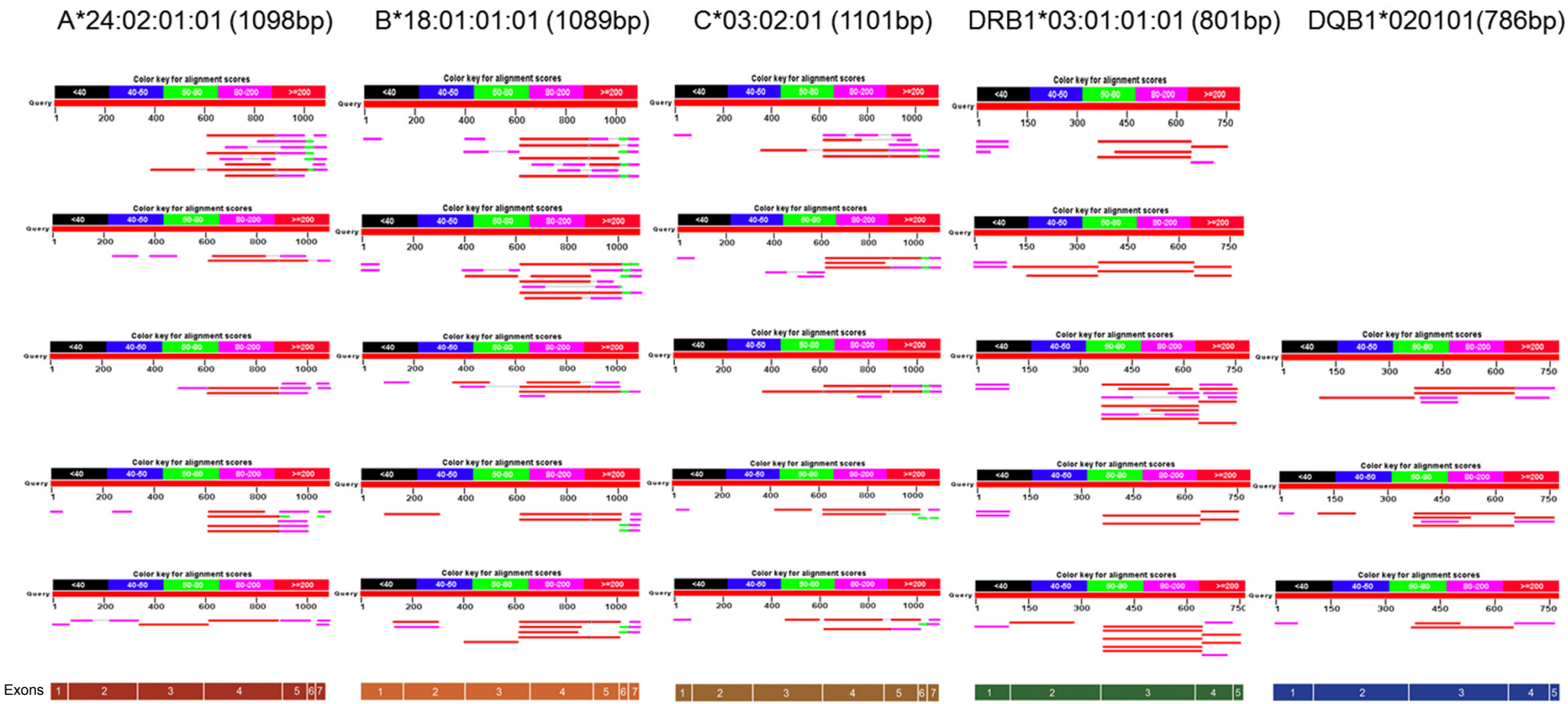
HLA-DRB1



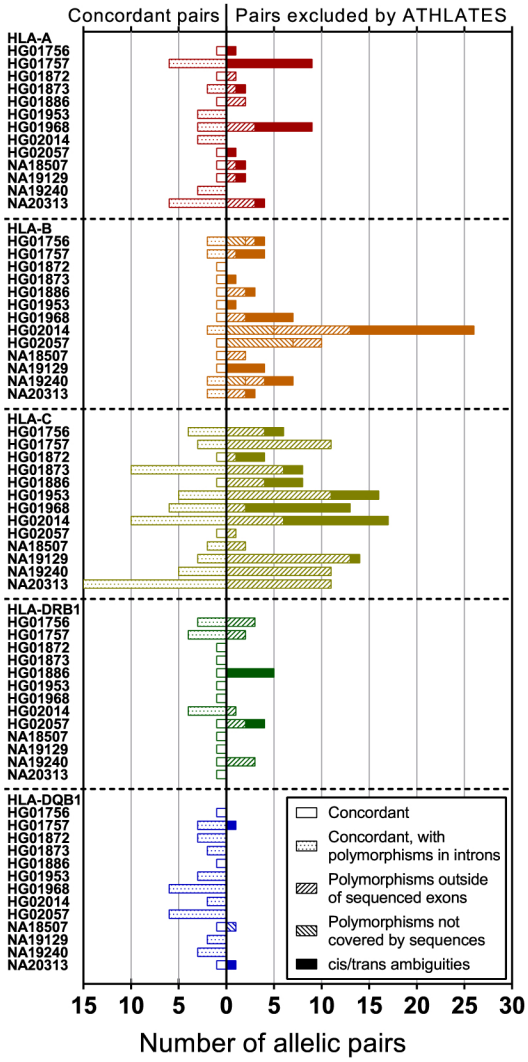
HLA-DQB1



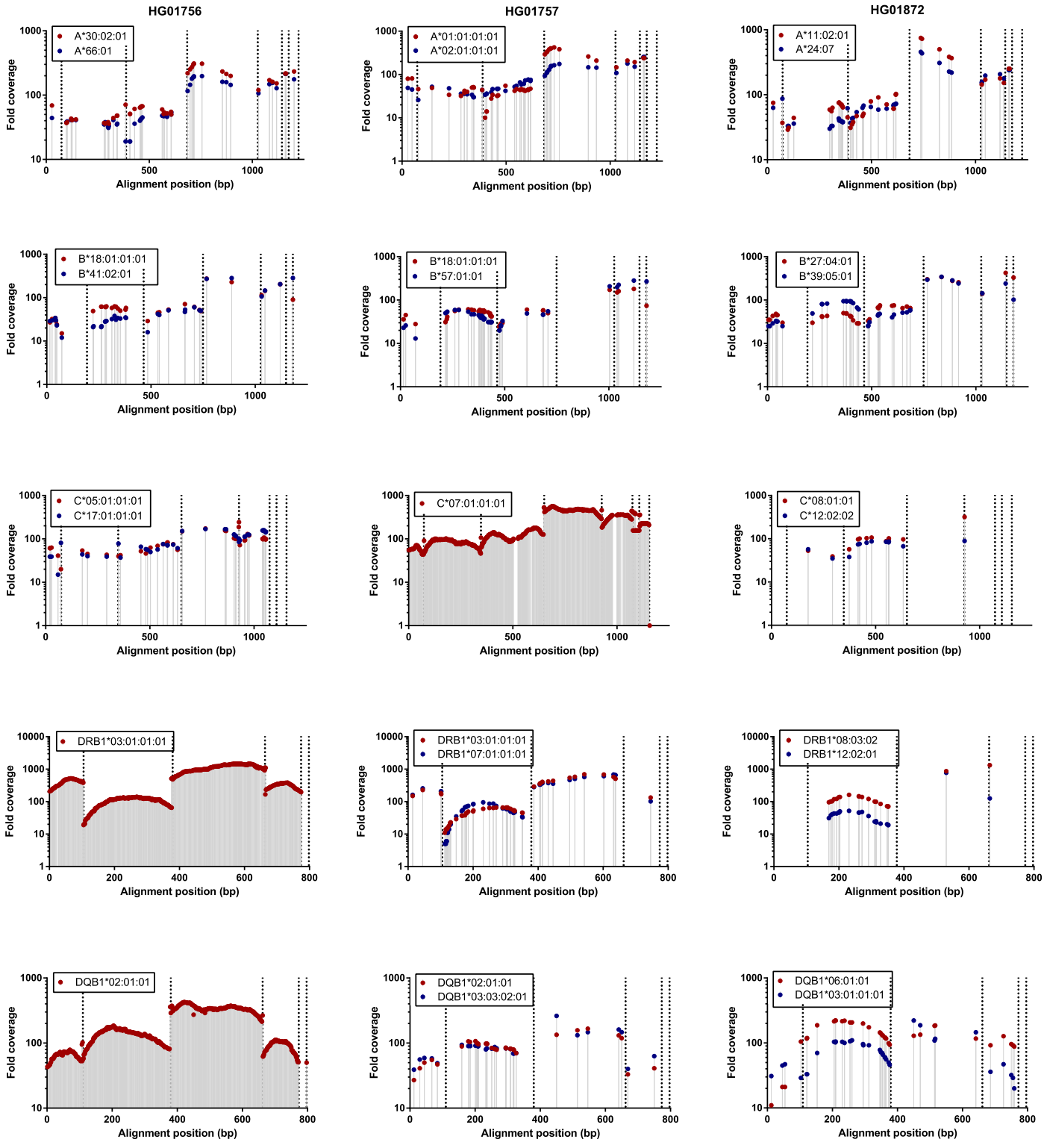
Supplementary Figure 5



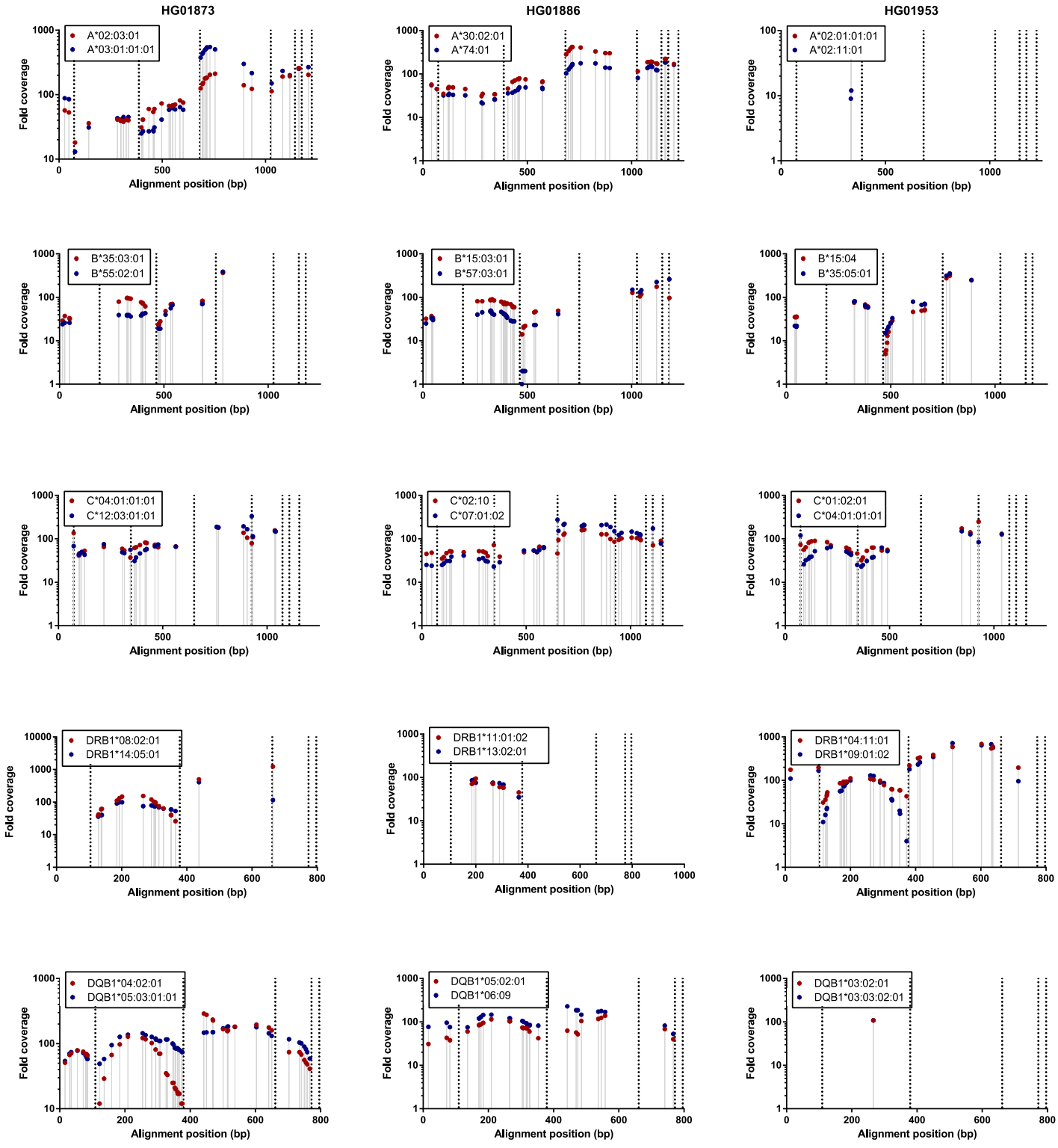
Supplementary Figure 6



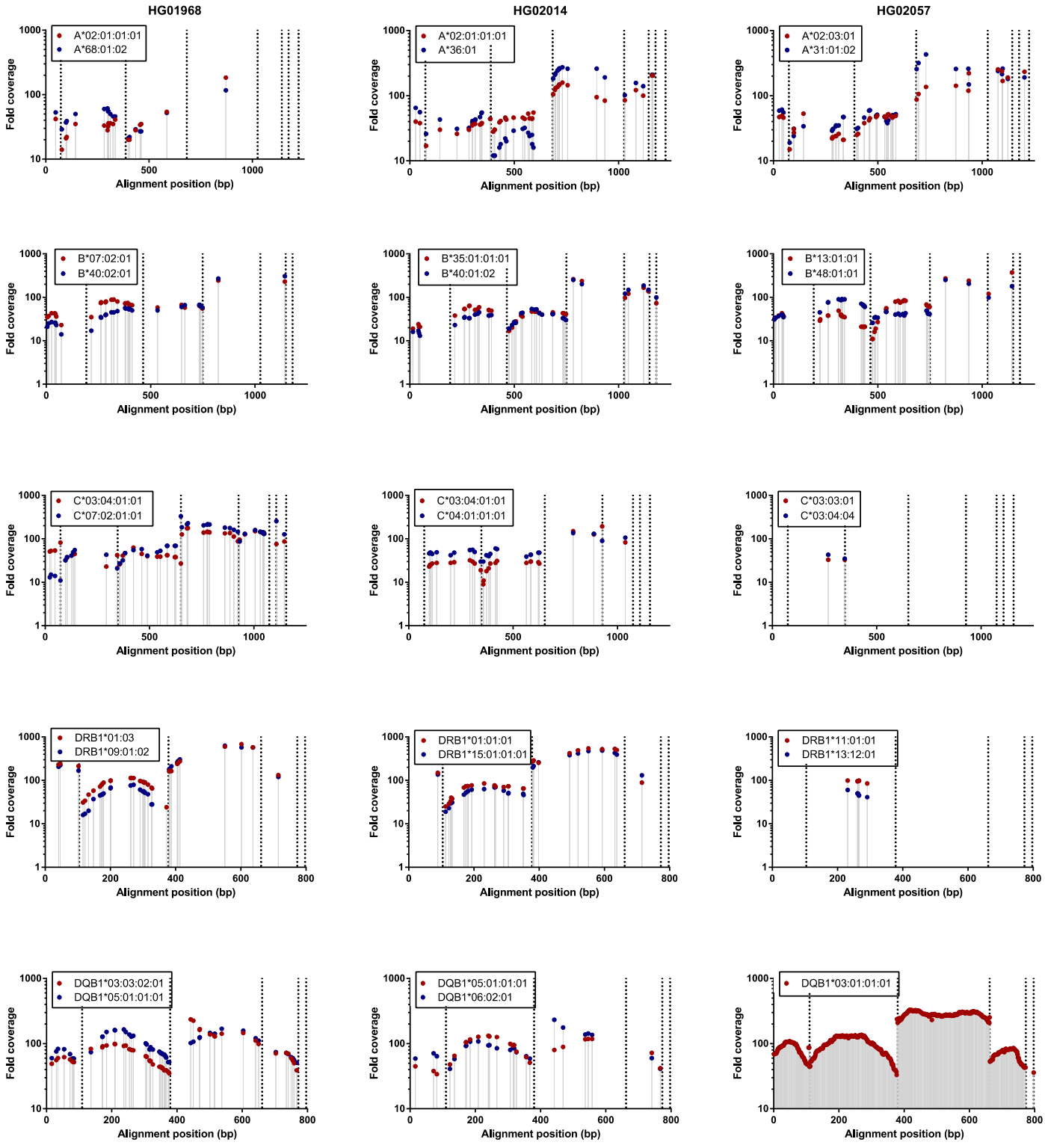
Supplementary Figure 7



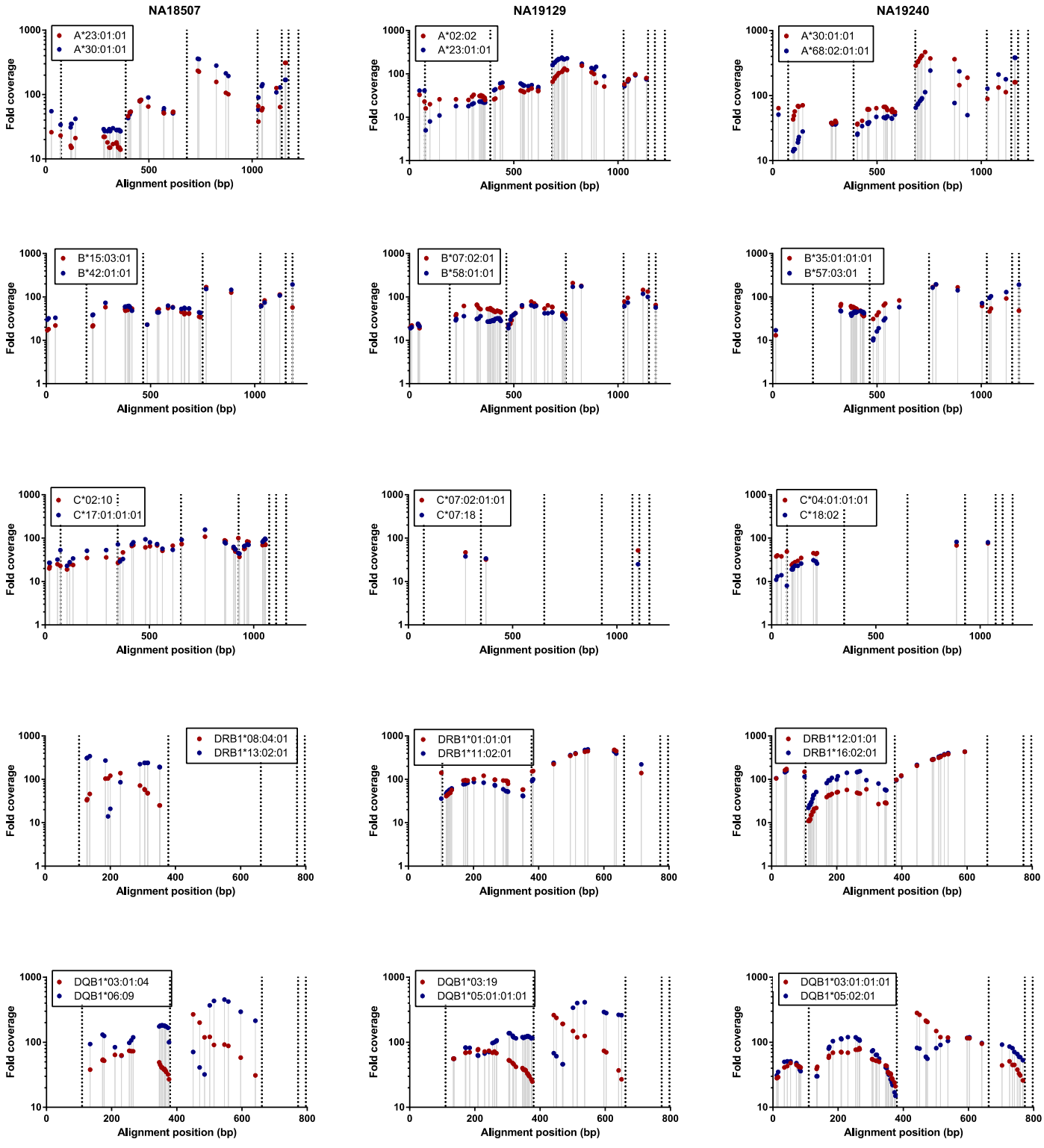
Supplementary Figure 8



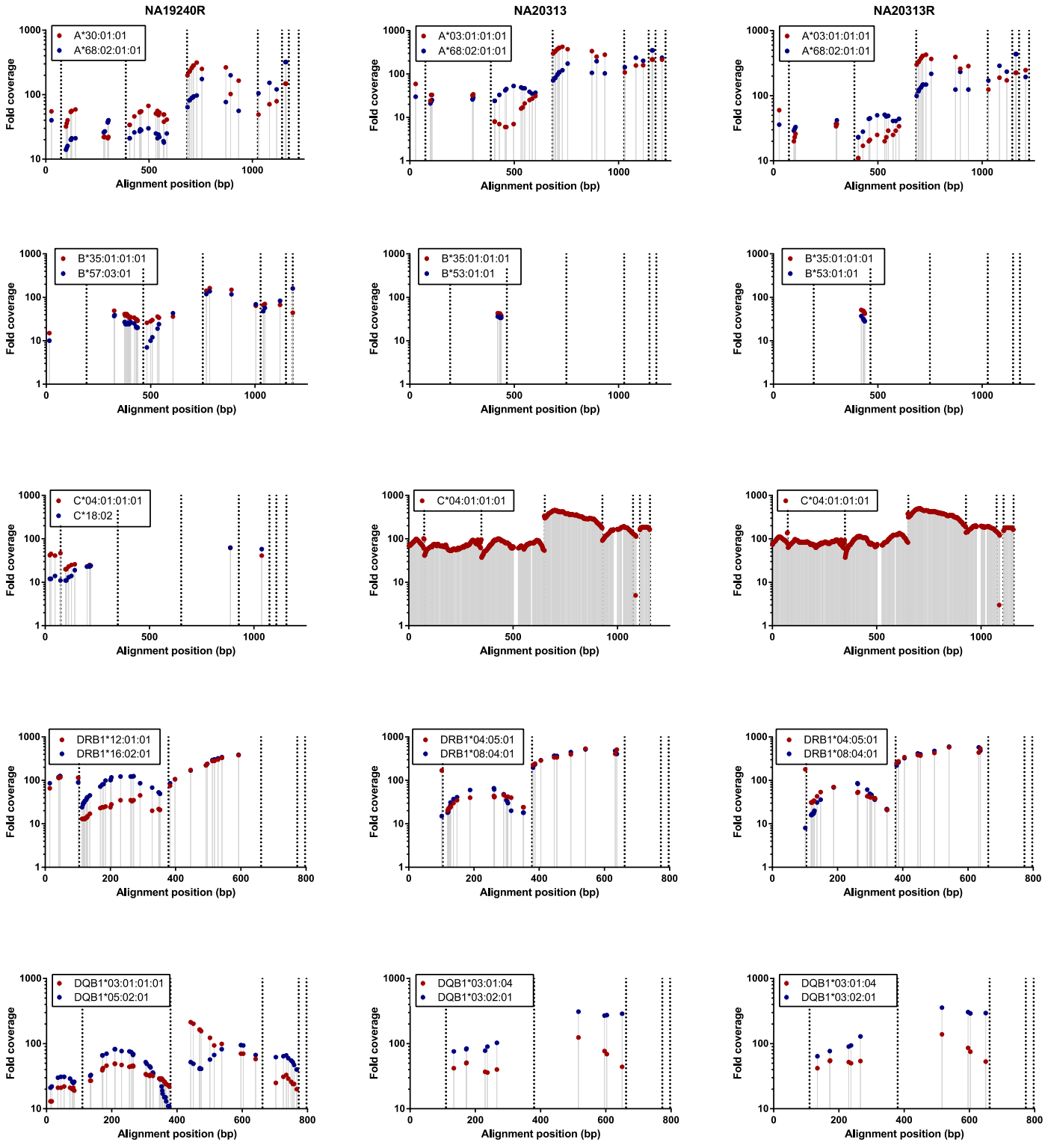
Supplementary Figure 8 (continued)



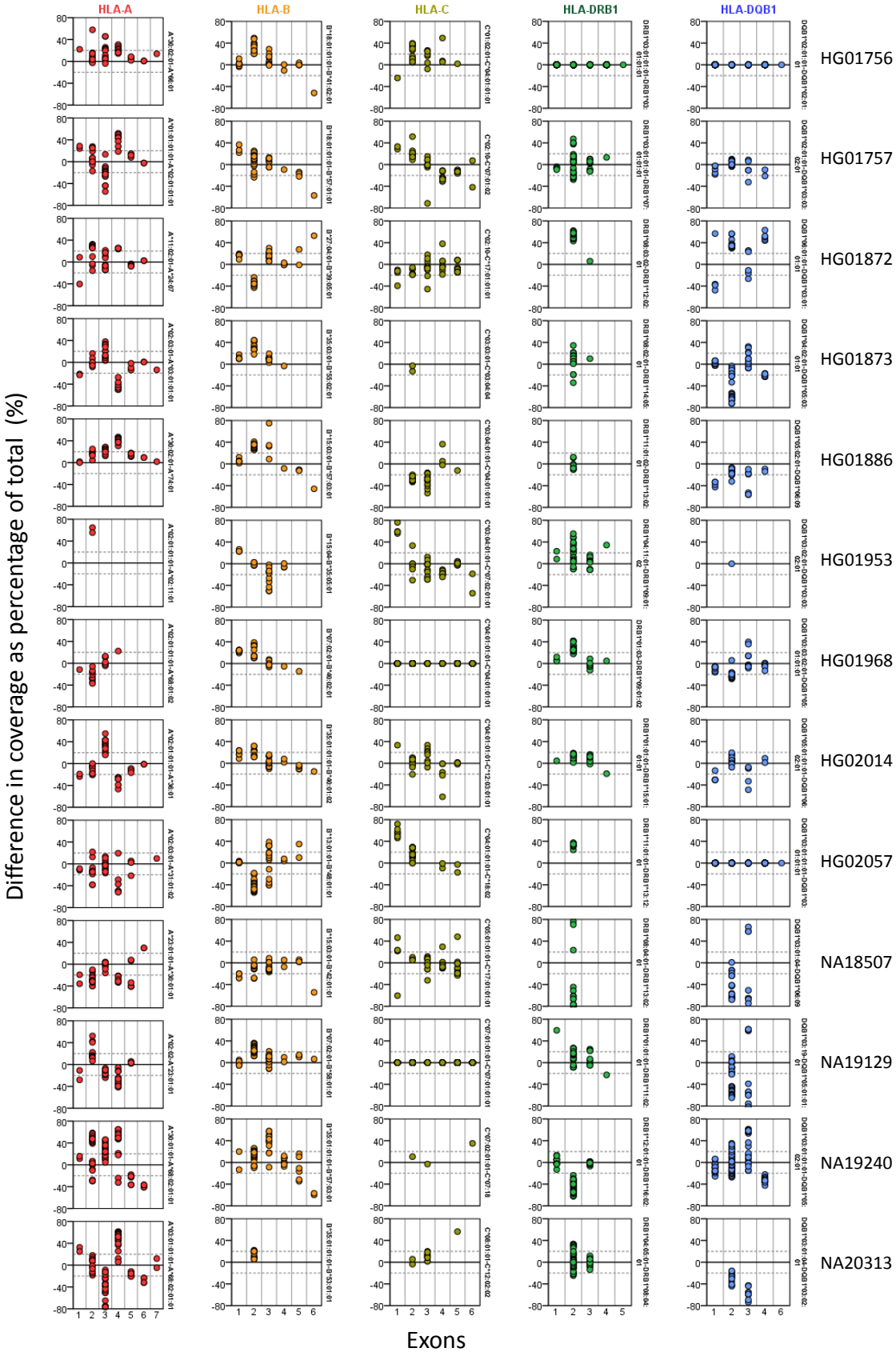
Supplementary Figure 8 (continued)



Supplementary Figure 8 (continued)



Supplementary Figure 8 (continued)



Supplementary Figure 9