

Customer feedback on products

Application : Development of a method for typing the major histocompatibility gene complex in humans (HLA) using a next-generation sequencer (NGS)

Product Name : Kapa Biosystems Library Amplification Kit (KK 2611, KK2612)

This application note was kindly provided by Dr. Kazuyoshi Hosomichi of Division of Human Genetics of National Institute of Genetics, Japan.

Purpose and background of development

The next-generation sequencing (NGS) technology realizes not only large-scale genomic sequencing but also rapid resequencing of targeted regions of interest on the genome from multiple samples. Targeted resequencing is an efficient approach for detecting, verifying and screening mutations, and we are applying this approach to develop an HLA typing method based on resequencing of the major histocompatibility gene complex (HLA) genes in humans.

The HLA region located on chromosome 6 region 6p21.31 is a 3.8-Mb segment comprising HLA genes involved in immune response. The genomic region exhibits an exceptionally high polymorphism in the human genome, with a vast number (over 8,700 in total) of HLA alleles being identified from the 6 HLA loci (HLA-A, -B, -C, -DR, -DQ and -DP). The HLA alleles are of great medical interest as they are associated with protection and severity of nearly 100 diseases including lifestyle-related diseases (e.g. diabetes), autoimmune diseases (e.g. rheumatism), cancer, GVH disease associated with hematopoietic stem cell transplantation and viral infections. In addition, antigen presentation by HLA molecules has potential applications in development of vaccines to prevent infections as well as new forms of cancer immunotherapy. Moreover, several reports have been made on strong association (odds ratio of over 1,000) between specific HLA alleles and drug-induced adverse effects such as Stevens-Johnson syndrome, toxic epidermal necrosis and drug-induced hypersensitivity syndrome, indicating the large contribution of HLA alleles to drug-induced adverse effects.

Two methods of HLA typing are currently used, i.e. the PCR-sequence specific oligonucleotide (PCR-SSO) method using fluorescent beads and the PCR-sequencing based typing (PCR-SBT) based on direct sequencing, but these technologies have not been replaced for more than 10 years. PCR-SSO can identify high-frequency alleles but not low-frequency alleles. Meanwhile, PCR-SBT, which directly determines the genome sequence, yields ambiguous typing results due to phase ambiguity for more than 8,000, 18,000 and 5,000 HLA allele combinations in HLA-A, -B and -C loci, respectively.

This application note provides an example case, focusing on the problems and solutions associated with library preparation in our project of resequencing HLA genes. For detailed description of the experiment and analysis methods, please refer to the reference.

Methods of experiment

Workflow

Genomic DNA derived from human peripheral blood



Library preparation using Nextera DNA Sample Prep kit

KAPA Library Amplification Kit

Amplification reagents from other manufacturers



MiSeq next-generation sequencer from Illumina



Data analysis using bwa, SAMtools, GATK and original programs



Data visualization by Integrative Genomics Viewer (IGV)

KAPA Library Amplification Kit

Reaction mixture composition

2xKapa HiFi HS RM	25uL
50xNextera Primer Cocktail	5uL
Index1 primer	5uL
Index2 primer	5uL
Tagmented Library	10uL
	50uL RXN

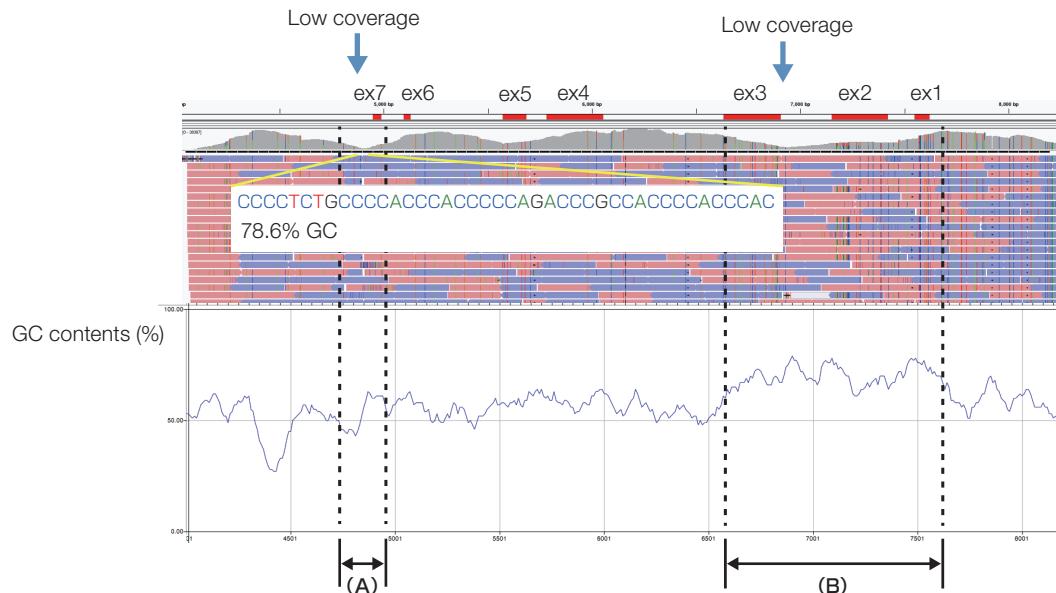
PCR cycle

Initial Extension	72°C	3min
Denaturation	98°C	30sec
Denaturation	98°C	10sec
Annealing	63°C	30sec
Extension	72°C	3min
Hold	10°C	

7cycles

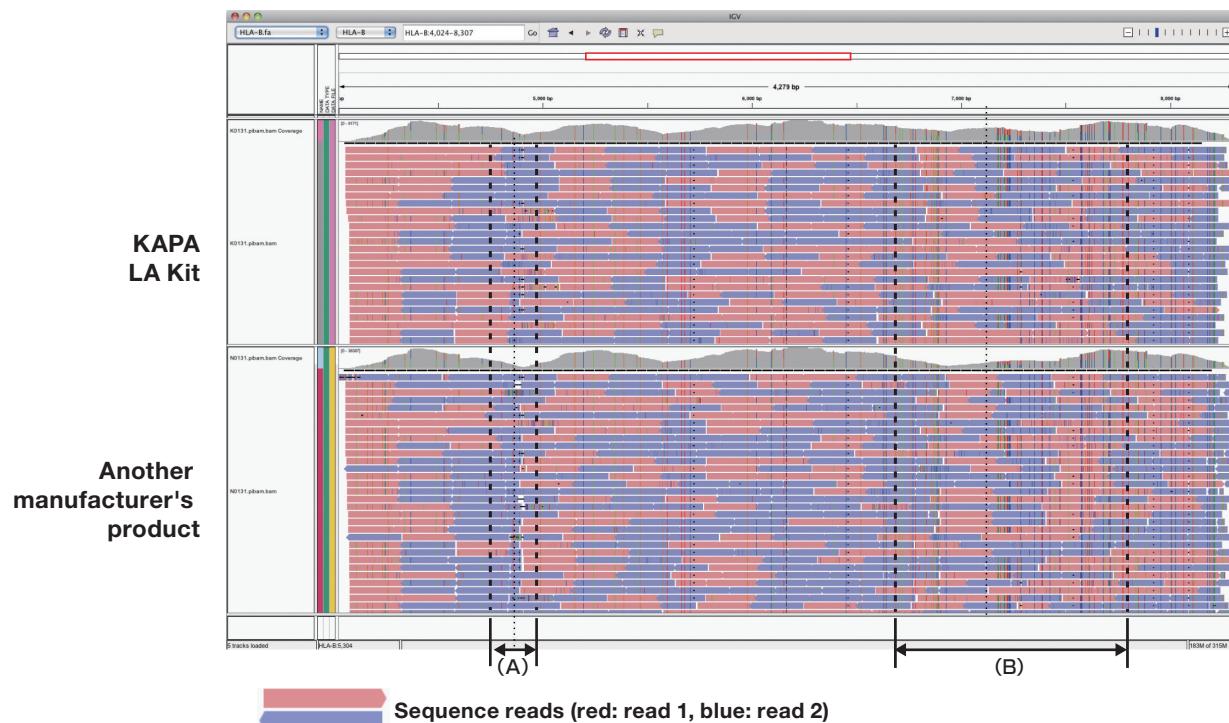
Results

Relation between GC content and coverage



The weakness of next-generation sequencing is that the read coverage tends to be lower in regions with high GC contents. In HLA genes, the most important exons 2 and 3 have high GC contents. In our analysis, we integrated two or more single-nucleotide variation (SNV) data as a physical data from one paired-end sequence read and defined the phase to construct two HLA gene sequences. Low coverage was one of the factors hindering defining of the phase, which needed to be solved.

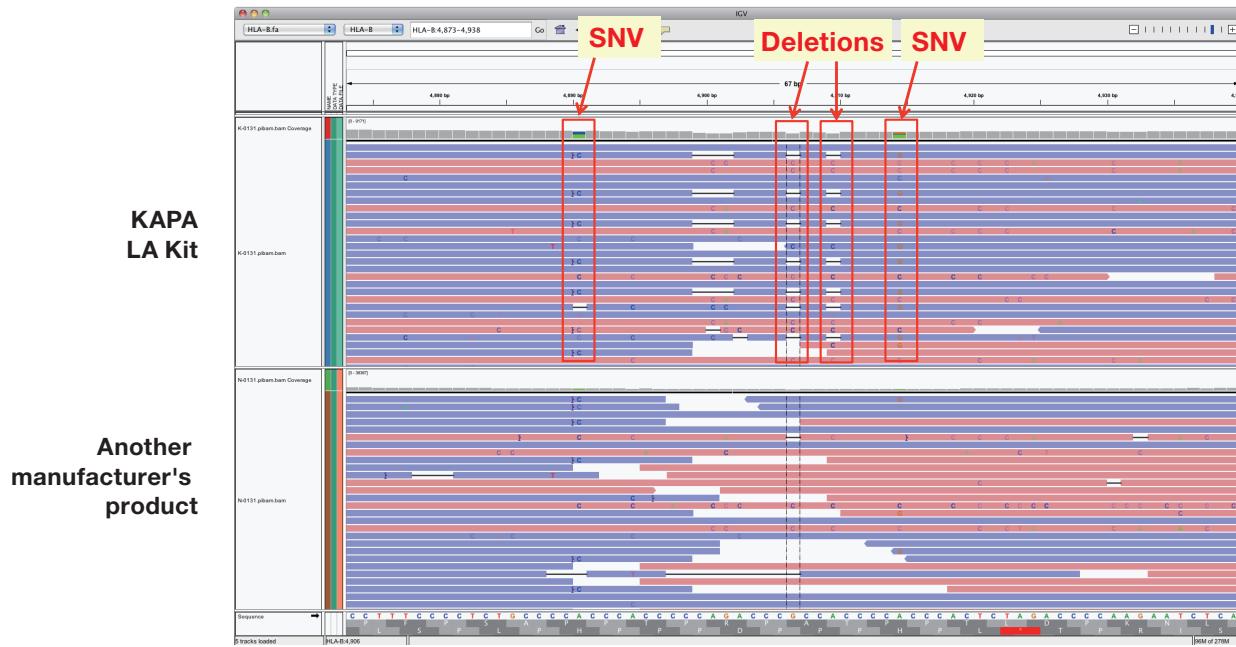
Visualization of analysis results Integrative Genomics Viewer (IGV)



Library preparation using the KAPA LA Kit resulted in several times larger number of reads successfully aligned compared to library preparation using another manufacturer's product in two regions (A) and (B) with extremely low coverage. As a result, the bias produced by GC content was eliminated, resulting in an alignment with a higher uniformity of coverage. The elimination of low-coverage regions leads to improved data accuracy with smaller number of data and is thus expected to realize sequencing cost reduction and higher throughput.

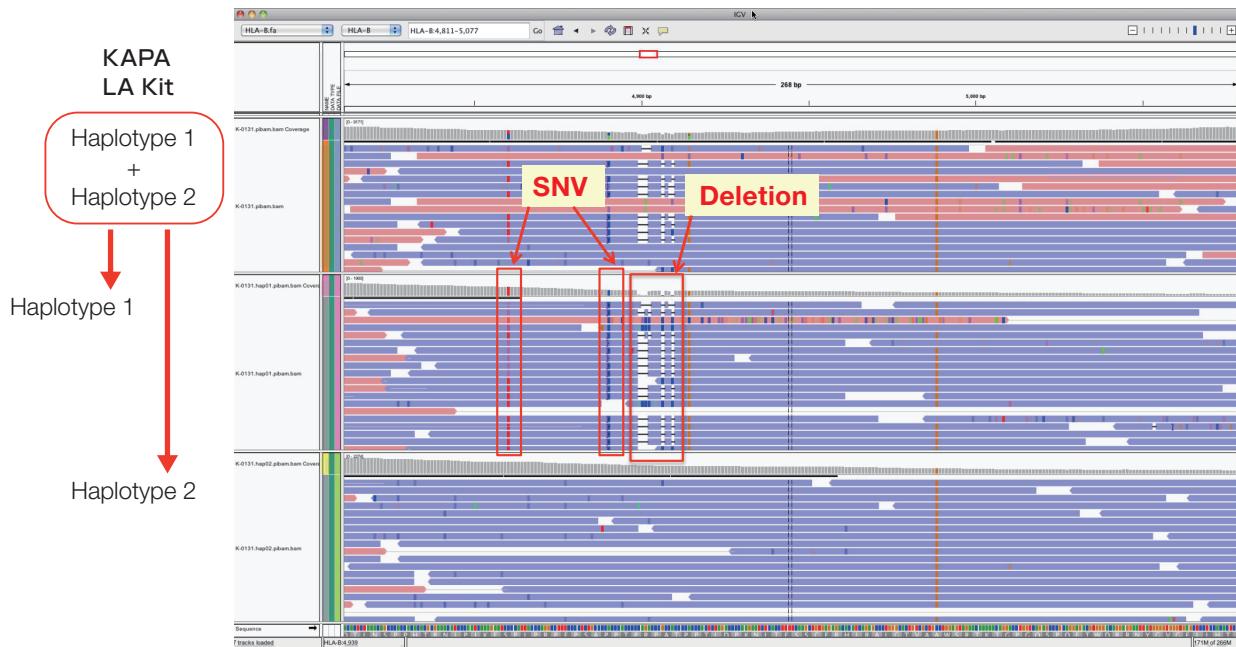
Results

Detection and comparison of deletions and SNVs



The library prepared using another manufacturer's product had extremely small number of reads aligned in the high GC content regions, and the small number of reads covering the region made it difficult to detect deletions. This was due to misalignment induced by a bias in the read distribution. Moreover, the detection sensitivity of SNV near deletions was reduced. On the other hand, the library prepared using the KAPA LA Kit had uniform distribution of reads aligned, even in the high GC content regions, resulting in successful detection of deletions and SNVs near the deletions.

Haplotype determination of deletions and SNVs



Moreover, since deletions and SNVs could be detected with a high reliability, we were able to classify these variants detected as heterozygotes into two haplotypes. Assignment of variants to either of the two chromosomes is essential to determining the HLA alleles and is thus the most important point in our analysis.

Results

Comparison of HLA typing results (HMA-B, cDNA database)

	PCR-SSO method (conventional method)		KAPA LA Kit		Another manufacturer's product		Depth	
Sample name	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	KAPA LA	Another manufacturer's product
Sample 1	B*54:01:01	B*57:01:01	B*54:01:01	B*57:01:01	No call	No call	5140.46	5329.84
Sample 2	B*38:01:01	B*58:01:01	B*38:01:01	B*58:01:01	No call	B*58:01:01	4852.13	5612.92
Sample 3	B*15:02:01	B*51:01:01	B*15:02:01	B*51:01:01	No call	No call	5692.53	5903.77
Sample 4	B*48:01:01	B*58:01:01	B*48:01:01	B*58:01:01	B*48:01:01	B*58:01:01	5948.66	5443.36
Sample 5	B*40:49	B*58:01:01	B*40:49	B*58:01:01	B*40:49	B*58:01:01	5878.22	3029.58

No call: haplotypes could not be fully distinguished in low-coverage regions, and full-length HLA gene sequence could not be obtained.

The evaluation of an HLA typing method entirely depends on the accuracy of determination of HLA alleles. When the analysis pipeline we developed was processed using the same parameters, the data obtained using the library prepared by another manufacturer's product required correction for determining the HLA alleles, whereas the data obtained using the KAPA LA Kit enabled accurate HLA allele typing using only the analysis pipeline. Automation of stable data processing is expected for processing multiple samples. Low-biased library preparation has the largest contribution to producing high-quality sequencing data, so the KAPA LA Kit was effective in reducing analysis load.

Acknowledgement

This study was a cooperative study with Professor Ituro Inoue of Division of Human Genetics of National Institute of Genetics and Dr. Marie Shimizu and Dr. Fumiaki Nakajima of Japan Red Cross Society Central Blood Research Institute. We greatly appreciate their cooperation in developing the method.

Conclusion

To date, low coverage in high GC content regions was a problem. Nextera DNA Sample Prep Kit is an excellent product enabling easy library preparation within a short time, but the use of the KAPA Library Amplification Kit for PCR amplification included in the protocol yielded sufficient number of sequence reads, even in the high GC content regions, and drastically reduced the undesirable GC bias. The reduced bias in coverage enables highly reliable analysis with a smaller number of sequence reads, realizing simultaneous sequencing of multiple samples. Although this method may not be versatile, in our case of analyzing the target gene with a high GC content, the use of the KAPA Library Amplification Kit was effective in determining the base sequence with a high throughput and at a low cost.

Reference (Japanese book)

Next-generation Sequencer (Jisedai Sequencer): Purpose-by-purpose Advanced Method (Mokuteki betsu advanced method) (Extra Issue of Cell Technology (Saibou Kougaku Bessatsu)), Shujunsha Co., Ltd.



Nippon Genetics Co.,Ltd <http://www.n-genetics.com>