REVIEW

The impact of next-generation sequencing technologies on HLA research

Kazuyoshi Hosomichi¹, Takashi Shiina², Atsushi Tajima¹ and Ituro Inoue³

In the past decade, the development of next-generation sequencing (NGS) has paved the way for whole-genome analysis in individuals. Research on the human leukocyte antigen (HLA), an extensively studied molecule involved in immunity, has benefitted from NGS technologies. The HLA region, a 3.6-Mb segment of the human genome at 6p21, has been associated with more than 100 different diseases, primarily autoimmune diseases. Recently, the HLA region has received much attention because severe adverse effects of various drugs are associated with particular HLA alleles. Owing to the complex nature of the HLA genes, classical direct sequencing methods cannot comprehensively elucidate the genomic makeup of HLA genes. Thus far, several high-throughput HLA-typing methods using NGS have been developed. In HLA research, NGS facilitates complete HLA sequencing and is expected to improve our understanding of the mechanisms through which HLA genes are modulated, including transcription, regulation of gene expression and epigenetics. Most importantly, NGS may also permit the analysis of HLA-omics. In this review, we summarize the impact of NGS on HLA research, with a focus on the potential for clinical applications.

Received 18 May 2015; revised 10 June 2015; accepted 21 July 2015; published online 27 August 2015

INTRODUCTION

The sequencing of the entire human genome in 2007, after a 4-year process, provided important insights into our complete genomic makeup.¹ Subsequently, the genomic sequences of J. Watson, African, Chinese, Korean and Japanese genomes were reported.²⁻⁶ Analyses of the personal genomes of individuals have provided information on human genetic variation and complexity. Additionally, rapid progress in next-generation sequencing (NGS) technology has led to revolutionary changes in medical genomics, supplying massive sequencing data for human samples. Indeed, the 1000 Genome Project has already reported novel variants, both rare and common, from population-scale sequencing.⁷

Various study designs have been applied to NGS, including DNA target resequencing, RNA sequencing for transcriptome analysis, chromatin immunoprecipitation sequencing, bisulfite sequencing for methylene analysis and others. The Encyclopedia of DNA Elements (ENCODE) project has examined the role of 99% of non-protein-coding DNA,⁸ revealing substantial interactions between proteins and DNA and the transcription of functional elements other than mRNA encoding proteins. Moreover, various types of NGS technologies have been developed, including smaller-scale benchtop and long-read NGS systems. Benchtop NGS systems, such as GS Jr, ionPGM and MiSeq, allow researchers to make fine adjustments for various smaller-scale studies. For example, some panels of focused target genes, such as genes related to cancer and inherited diseases, are now available for sequencing.

Human leukocyte antigen (HLA) genes have a long research history as important targets in biomedical science and treatment. The HLA region on chromosome 6p21 comprises six classical HLA genes and at least 132 protein-coding genes. This region has important roles in regulation of the immune system as well as fundamental molecular and cellular processes.⁹ The sequencing of a continuous 3.6-Mb HLA genomic region with annotation of 224 genes was reported by the MHC Sequencing Consortium in 1999.¹⁰ In addition, the MHC Haplotype Project was carried out between 2000 and 2006 by the Sanger Institute, providing genomic sequences and gene annotations of eight different HLA-homozygous haplotypes to build a framework and resource for association studies of all HLA-linked diseases; these haplotypes were registered as UCSC hg19 or NCBI GRCh37 reference assemblies.¹¹⁻¹³ This small segment of 3.6 Mb occupies only 0.13% of the human genome but is associated with more than 100 diseases, mostly autoimmune diseases such as diabetes, rheumatoid arthritis, psoriasis and asthma. Furthermore, specific alleles of the HLA genes are strongly associated with hypersensitivities to specific drugs. For example, strong associations between carbamazepine-induced Stevens-Johnson syndrome or toxic epidermal necrolysis and HLA-B*15:02,¹⁴,¹⁵ abacavir-induced liver injury and HLA-B*57:01,¹⁶⁻¹⁹ and allopurinol-induced Stevens-Johnson syndrome or toxic epidermal necrolysis.
necrolysis and HLA-B*5801 have been reported in various populations. For a better understanding of the disease causality and adverse effects of drugs, the haplotype structure of the HLA region should be extensively and unambiguously determined. Therefore, a specific analytical procedure should be developed for completion of HLA sequencing and haplotype determination. NGS technologies have potential advantages over the Sanger method in the sequencing of HLA genes, that is, sequences of haplotype structure can be obtained at high throughput.

To date, several high-throughput HLA-typing methods using NGS have been developed. Importantly, HLA typing using NGS provides both high-throughput and high-resolution capabilities (Figure 1). Additionally, as reported by the ENCODE Project, HLA gene sequencing alone is not sufficient for developing a complete understanding of the genetic makeup of the HLA locus. The expression levels of HLA genes can have crucial roles in the pathogenesis of diseases; thus, detection of regulatory single-nucleotide variants (SNVs) and insertions and deletions (Indels) located outside of exons is necessary. If phase-defined complete sequencing of HLA genes, including functional regulatory regions, is performed, novel alleles associated with disease risks and adverse effects of drugs could be obtained, and the expression levels of genes that affect biological processes could be clarified.

PCR-BASED HLA SEQUENCING USING NGS

PCR-based methods, involving an amplicon-sequence step and a sequence capture step, are commonly used for library preparation. Most of the NGS-based HLA-typing methods have been developed using such techniques. In 2009, two HLA-typing methods using a Roche GS FLX system were reported (Table 1). The most popular HLA typing method based on the sequence capture technology. This method uses targeted capturing of the classical class I (HLA-A, HLA-C, and HLA-B) and class II (HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1) HLA genes. The DNA fragments from these eight HLA genes can be simultaneously enriched by a hybridization reaction in a single tube, without allele dropout, which is frequently observed in PCR-based methods. The results showed high accuracy for allele calls (99%) and identified errors in the IMGT/HLA reference database. It is also notable that the sequence capture method is generally applicable for NGS-based target resequencing of larger genomic regions and a larger number of genes than the PCR-based methods. On the basis of these features and the findings from the automated NGS-typing method, the sequence capture method has major advantages over PCR-based methods and is a promising method for HLA sequencing.

Impact of NGS on the study of HLA genes
K. Hosomichi et al

**Figure 1** HLA typing to provide sequencing data for the HLA gene(s) and regions. The HLA sequencing data of NGS could be analyzed from various points of view. The minimum scope of polymorphisms is the genotype of an SNV, and the maximum scope is the HLA haplotype sequence as a set of alleles from each HLA gene. The phase-determined sequence of the HLA allele can be applied for HLA typing as a reference. The resolution of HLA typing is classified into the following four categories: two-digit for alleles, four-digit for specific HLA proteins, six-digit for specific HLA coding sequence (CDS) and eight-digit as specific HLA genome sequence including untranslated regions and introns.
| Year | Author | Journal | Locus | PCR method | Target region | Sequencer | Data analysis |
|------|--------|---------|-------|------------|--------------|-----------|--------------|
| 2009 | Gabriel C et al. | Human Immunology | HLA-A, -B | PCR for each exon | Exons 1, 2, 3 and 4 | GS FLX (Roche) | AYA (Roche) Assign SBT (Conexio Genomics) |
| 2009 | Bentley G et al. | Human Immunology | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Exons 2, 3 and 4 of A, B and C; exon 2 of DRB1, DQA1, DQB1, DQB1 | GS FLX (Roche) | HLA typing software (Conexio Genomics) |
| 2009 | Lank SM et al. | Human Immunology | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS FLX (Roche) | Assign MPS software (Conexio Genomics) |
| 2010 | Lind C et al. | Human Immunology | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | Long PCR | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS FLX (Roche) | AYA (Roche) Assign SBT (Conexio Genomics) |
| 2010 | Holcomb CL et al. | Tissue Antigens | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS FLX (Roche) | Assign MPS software (Conexio Genomics) |
| 2011 | Erlich et al. | Human Immunology | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS FLX (Roche) | AYA (Roche) Assign SBT (Conexio Genomics) |
| 2011 | Wang C et al. | Proc Natl Acad Sci U S A | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | Long PCR | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GAIIx (Illumina) | HiSeq2000 (Illumina) MiSeq (Illumina) |
| 2012 | Shiina T et al. | Tissue Antigens | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPA1, -DPB1 | Long PCR | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | IonPGM (Thermo) |
| 2012 | Lank SM et al. | BMC Genomics | HLA class I, all HLA class II loci | RT-PCR | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | BLAT (GeneCodes) |
| 2013 | Moonsamy PV et al. | Tissue Antigens | HLA-A, -B, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS FLX (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Ringquist S et al. | PlOj One | HLA-DRB1 | PCR | Exon 2 | GS FLX (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Danzer M et al. | BMC Genomics | HLA-A, -B, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Hosomichi K et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Trachtenberg EA et al. | Methods Mol Biol | HLA-A, -B, -DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Hajeer AH et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Hosomichi K et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Smith AG et al. | Human Immunology | HLA-DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Ehenberg PK et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2013 | Zhou M et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Lan JC et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Ozaki Y et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | smith AG et al. | Human Immunology | HLA-DRB1, -DRB3/4/5, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Ehenberg PK et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Zhou M et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Lan JC et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
| 2015 | Ozaki Y et al. | BMC Genomics | HLA-A, -B, -DRB1, -DQA1, -DQB1, -DPB1 | PCR for each exon | Entire gene of A, B and C exons 2-3 of DRB1 and DQB1 | GS Junior (Roche) | Assign MPS software (Conexio Genomics) |
SEQUENCERS FOR HLA GENE SEQUENCING

Various sequencing machines have been developed for NGS HLA typing. The majority of published methods have been established using Roche sequencers. However, for the last few years, the Illumina MiSeq instrument has also been used for HLA typing (Table 1). The types of NGSs used for HLA typing may often change along with improvements in NGS technologies. The Pacific Biosciences PacBio RS II sequencer, which is capable of generating enormously long reads in a single molecule using real-time sequencing, was recently developed for HLA typing. Single molecule using real-time sequencing is highly effective in generating accurate, phased sequences of full-length alleles of HLA genes. Complete phasing of the HLA genes from single molecule using real-time sequencing may resolve phase ambiguity, which is a fundamental problem of conventional HLA-typing methods.

HLA TYPING FROM WGS AND WES DATA

During the past few years, whole exome sequencing (WES) has identified the causalities of a large number of Mendelian diseases by analyzing familial samples and/or sporadic patients. In addition, WES has facilitated the acquisition of massive amounts of data in various genome sequencing projects, such as the 1000 Genomes Project, NHLBI GO Exome Sequencing Project (https://esp.gs.washington.edu/drupal/) and UK10K project (http://www.uk10k.org), which are expected to improve our understanding of variations in the human genome. The sequence capture method has also been applied for WES using various kits, such as the Agilent Human All Exon kit, Roche SeqCap EZ Human Exome kit and Illumina TruSeq Exome Enrichment kit. The respective libraries of capture oligoprobes, which cover all human exons, are designed to target all exons of all HLA genes. For example, 820 exons of 182 genes in the HLA region are found in the Agilent Human All Exon kit design.
DNA sequence reads for HLA genes are included in the whole genome sequencing (WGS) and WES datasets, HLA typing could be carried out using both the datasets. Within WGS and WES datasets, HLA gene sequences represent only a small portion of the data, but these sequences are phased HLA gene sequences as HLA alleles. Therefore, HLA typing from WGS or WES datasets should be the key analysis method used to promote higher accuracy rates compared with those of the existing PCR-SSO or PCR-sequencing-based typing results.

**NGS HLA-TYPING SOFTWARE**

As described in Table 2, various HLA-typing software programs, including the aforementioned Omixon Target HLA and HLAmixer, as well as academic software and commercial software packages, have been developed for HLA typing from various types of data, including WGS, WES, RNA sequencing and amplicon datasets. As an example of HLA-typing software for WGS or WES, including a brief overview of the Omixon Target HLA typing system, is described in Figure 3.

For statistical methods, sequence reads are first aligned to the whole IMGT/HLA database (all known HLA alleles). Then, the best matching alleles are selected based on various alignment statistics, such as the number of reads covering exons and the extent of exons covered. During statistical analysis, only reads that are mappable as homologous to any allele in the IMGT/HLA database with a low number of mismatches should be stored. In the Omixon publication, which used data from the 1000 Genomes Project, the concordance rate between the NGS-based method and PCR-SSO was around 90%, which was not considered high. For the analysis, sequence reads from all exons of HLA genes were applied. At least 10 reads are required to counterbalance random noise. However, the sequence reads were not evenly distributed for each gene region, and the average depth implied that there may be holes in coverage. Another publication in which the authors utilized the HLAmixer software also mentioned 92.8% concordance rate between these two methods for allele group prediction. NGS HLA typing can call for all the

| HLA-typing software | URL | Read type | Reference |
|---------------------|-----|-----------|-----------|
| HLAmixer            | http://www.bcgsc.ca/platform/bioinfo/software/hlaminer | WGS/WES/RNA-seq/amplicon | Warren RL et al. |
| seq2HLA             | http://tron-mainz.de/tron-facilities/computational-medicine/seq2hla/ | RNA-seq | Boegel S et al. |
| ATHLATES            | https://www.broadinstitute.org/scientific-community/science/projects/viral-genomics/athlates | WGS/WES/amplicon | Liu C et al. |
| OptiType            | http://omictools.com/optitype-s6206.html | WGS/WES/RNA-seq | Szolek A et al. |
| HLAforest           | http://code.google.com/p/hlaforest | RNA-seq | Kim HJ et al. |
| PHLAT               | https://sites.google.com/site/projectphilat/ | WGS/WES/RNA-seq | Bai Y et al. |
| Phase-defined HLA sequencing | https://p-galaxy.ddbj.nig.ac.jp | Amplicon | Hosomichi K et al. |
| HLAreporter         | http://paed.hku.hk/genome/software.html | WGS/WES | Huang Y et al. |
| HLA-VBSeq           | http://nagasakilab.csml.org/hla/ | WGS | Nariai N et al. |
| HLAssign            | http://www.ikmb.uni-kiel.de/resources/download-tools/software/hlassign | Sequence capture | Wittig M et al. |
| Assign ATF (Conexio Genomics) | http://www.conexio-genomics.com | Amplicon | — |
| Omixon Target HLA (Omixon) | http://www.omixon.com/hlal | WGS/WES/amplicon | Major E et al. |
| NGScengine (Gen Dx) | http://www.gendx.com/products/ngscengine | Amplicon | — |
| GeMS (Scisco Genetics) | http://sciscogenetics.com/services/integrated-genotyping-system/ | Amplicon | — |

**Figure 3** Overview of data analysis for HLA typing using WGS/WES. Massive sequence reads from WGS/WES are aligned to the whole IMGT/HLA database (all known HLA alleles) to search for best matching alleles based on alignment statistics, number of reads covering exons and the extent of exon coverage. The HLA allele can be identified by only storing reads that are mappable as homologous to any allele in the IMGT/HLA database with a low number of mismatches by statistical analysis.
HLA genes recorded in the IMGT/HLA database and for novel HLA alleles. On the other hand, it is not currently possible to detect rare HLA alleles by PCR-SSO. Therefore, if sequence reads of rare HLA alleles are in WGS or WES, the HLA-typing results from NGS would be expected to be discordant with those from PCR-SSO. In the near future, the reliability of these HLA-typing methods from WGS and WES data may be improved. These programs are the next step in developing methods with greater specificity and sensitivity of HLA-typing results. In particular, the specificity is dependent on the HLA-typing resolution, for example, two-, four-, six- and eight-digit, each of which is based on the composition of the allele group, the specific allele protein, the specific DNA sequence with synonymous substitutions in the coding region and the specific DNA sequence of the entire gene, respectively. The high-resolution HLA typing of NGS is advantageous compared with the existing PCR-SSO and PCR-sequencing-based typing methods. In practice, it is not possible to execute complete eight-digit HLA typing because of limitations in the number of known HLA allelic sequences deposited in the IMGT/HLA database, where most HLA allelic sequences have been recorded as coding sequences or partial exons. Only HLA alleles recorded as full-length HLA gene sequences can be used for allele-call with eight-digit resolution. To put eight-digit resolution typing into practice, the NGS-based phase-defined complete sequencing methods for the HLA genes will be applicable as a high-resolution tool for the detection of novel alleles, and will facilitate the development of expanded databases with full-length HLA allelic sequences for eight-digit HLA typing. The success of complete HLA gene sequencing with high accuracy should be determined based on

Figure 4 Example of target resequencing to detect variants and functional prediction of the regulatory region. Target resequencing of the HLA region clarifies all variants in the target region. For example, several variants and approximately 2-kb deletions have been detected in the upstream region of HLA-DRA. (a) Alignment view of mapped reads (pink: forward strand read, purple: reverse strand read) in the alignment track for detection of SNVs (A: green, C: blue, G: yellow, T: red) and the deletion as displayed in the coverage track. (b) The region was located in cis-regulatory elements as active (H3K27ac-marked) enhancers and a DNase I-hypersensitive site defined by ENCODE chromatin immunoprecipitation sequencing and DNase-seq peaks. The deletion and SNVs may affect the expression level of HLA-DRA by influencing the binding of TFs.
high sequence read depth. In the case of HLA-B sequencing, the minimum depth for complete phasing was approximately 800 folds the average depth.\textsuperscript{37}

**HLA IN THE ENCODE PROJECT DATA**

In HLA research, NGS technologies have influenced HLA typing as well as our understanding of the functional regulatory regions within the HLA region, which could affect the expression of HLA genes. Thus far, HLA-associated diseases have been understood on the hypothesis that antigen presentation of HLA molecule affects the immune system. Four-digit HLA typing is sufficient to explain the importance of antigen presentation in disease causality. On the other hand, the HLA-associated phenotypes could also be affected by the expression levels of HLA genes or by allelic imbalance. In 2012, the ENCODE project succeeded in the systematic arrangement of transcript regions and transcription factor (TF)-binding sites in the genome, and showed the genomic patterns of chromatin structure and histone modifications.\textsuperscript{8} The achievements of the project also include the discovery of putative functional elements and domains within the HLA region. The knowledge obtained in the ENCODE project could be extended to understand HLA-associated diseases and phenotypes. Certainly diseases are associated with specific HLA alleles, and many variants within the HLA genes are also associated with HLA alleles in linkage disequilibrium; therefore, it is quite difficult to genetically determine which variant is associated with the disease because the disease is associated with a haplotype carrying the HLA alleles and many variants. Furthermore, there are limitations to genetic analyses with limited numbers of samples and minor genetic effects; however, the ENCODE project highlights the functional regions of the entire human genome including the HLA region.

Two examples of HLA genes and the associated diseases are described in Figure 4, Table 3 and Table 4. HLA-DRA is less polymorphic than HLA-A, B, -C or -DRB1. However, many variants have been observed in the upstream region among HLA-DRA alleles, particularly between the same six-digit HLA-DRA alleles, HLA-DRA*01:01:01 (Figure 4, unpublished data). A deletion of about 2 kb was also detected in the upstream region of HLA-DRA*01:02:02 by HLA target resequencing data. Before the completion of the ENCODE project, it was difficult to understand the effects of deletions. Now, we can see the possibility of a functional regulatory region around the deletion. Interestingly, two haploid genome sequences of HLA-DRA*01:01:01 had different sequences within the intron and upstream regions. Some of the variants also may affect the expression levels of HLA-DRA by mediating TF binding to the variants. The haplotype of the variants in the upstream region could be significantly different, even though the HLA allele was found to be the same as the HLA-DRA*01:01:01 sequence with six-digit resolution. The ENCODE project stressed the importance of complete HLA gene sequencing, including the upstream regulatory region, to determine the haplotype. In total, 3619 SNVs in the HLA region were selected as expression Quantitative Trait Loci (eQTL) SNVs for HLA gene expression (Table 3).\textsuperscript{35} These eQTL SNVs were identified in the RegulomeDB database (http://www.regulomedb.org), which have provided annotations of SNPs with known and predicted regulatory elements in the intergenic regions of the human genome. The database includes public datasets from the ENCODE project, in addition to GEO and publications. Known and predicted regulatory DNA elements from DNAse hypersensitivity, TF-binding sites and promoter regions that have been biochemically characterized to regulate transcription are also included. Recorded variants have been classified into various categories according to TF binding and target gene expression. The 3619 HLA eQTL SNVs are likely to affect the binding of TFs to mediate expression of the HLA gene. For variants and deletions near HLA-DRA, new hypotheses concerning the biological functions of the gene could be generated to improve our understanding of HLA-DRA-associated phenotypes.

In another example, 12 SNVs with regulatory functions have been shown to be associated with rheumatoid arthritis (Table 4). Of the lead SNVs, rs660895 is located 32.6 kb upstream of HLA-DRB1 (from the nearest transcription start site) and has been described as

| HLA gene | Number of eQTL SNVs |
|----------|---------------------|
| HLA-A    | 821                 |
| HLA-B    | 12                  |
| HLA-C    | 773                 |
| HLA-DRA  | 288                 |
| HLA-DRB1 | 580                 |
| HLA-DQA1 | 544                 |
| HLA-DQB1 | 473                 |
| HLA-DPA1 | 2                   |
| HLA-DPB1 | 126                 |
| Total    | 3619                |

Abbreviations: eQTL, expression Quantitative Trait Loci; HLA, human leukocyte antigen; SNV, single-nucleotide variant.

| Chr | Position | Lead SNP       | Distance to nearest TSS | GENCODE v7 location | ENCODE Score | RegulomeDB Score       |
|-----|----------|----------------|-------------------------|---------------------|--------------|------------------------|
| 6   | 29789171 | rs1610677      | 23 582 bp               | Intergenic region   | No regulatory annotation |
| 6   | 3137931  | rs1063635      | 9903 bp                 | Coding region       | Motif        |
| 6   | 32218989 | rs9296015      | 27 283 bp               | Intergenic region   | Motif        |
| 6   | 32282854 | rs6910071      | 49 238 bp               | Introgen region     | Motif        |
| 6   | 32429643 | rs9268853      | 68 359 bp               | Intergenic region   | Motif        |
| 6   | 32574171 | rs615672       | 35 847 bp               | Intergenic region   | Motif        |
| 6   | 32577380 | rs660895       | 32 638 bp               | Intergenic region   | ChIP-seq peak + DNasel-seq peak |
| 6   | 32602269 | rs9272219      | 44 749 bp               | Intergenic region   | Motif        |
| 6   | 32663851 | rs6457617      | 27 409 bp               | Intergenic region   | Motif        |
| 6   | 32663999 | rs6457620      | 27 557 bp               | Intergenic region   | Motif        |
| 6   | 32671103 | rs13192471     | 34 661 bp               | Intergenic region   | No regulatory annotation |
| 6   | 32680928 | rs7765379      | 44 486 bp               | Intergenic region   | No regulatory annotation |

Abbreviations: SNP, single nucleotide polymorphism; SNV, single-nucleotide variant; TSS, transcription start site.
CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P. et al. The diploid genome sequence of an individual human. PLoS Biol. 5, e254 (2007).
2 Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A. et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872–876 (2008).
3 Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G. et al. Accurate whole genome human sequencing using reversible terminator chemistry. Nature 456, 53–59 (2008).
4 Wang, J., Wang, W., Li, R., Liu, Y., Tian, G., Goodman, L. et al. The diploid genome sequence of an Asian individual. Nature 456, 60–65 (2008).
5 Kim, J. I., Ju, Y. S., Park, H., Kim, S., Lee, S., Yi, J. H. et al. A highly annotated whole-genome sequence of a Korean individual. Nature 460, 1011–1019 (2009).
6 Fujimoto, A., Nakagawa, H., Hosono, N., Nakano, K., Abo, T., Boroiovich, K. A. et al. Whole-genome sequencing and comprehensive variant analysis of an Indian individual using massively parallel sequencing. Nat. Genet. 42, 931–936 (2010).
7 1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M. et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).
8 Pennisi, E. ENCODE project writes eulogy for junk DNA. Science 337, 1159–1161 (2012).
9 Shinya, T., Hosomichi, K., Inoko, H. & Kukuchi, J. K. The HLA genomic loci map: expression, interaction, diversity and disease. J. Hum. Genet. 54, 15–39 (2009).
10 The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. Nature 401, 921–923 (1999).
11 Stewart, C. A., Horton, R., Alcocer, R. J., Ashurst, J. L., Atzeha, A. M., Coggill, P. et al. Complete MHC haplotype sequencing for common disease gene mapping. Genome Res. 14, 1176–1187 (2004).
12 Horton, R., Gibson, R., Coggill, P., Miretti, M., Alcocer, R. J., Almeida, J. et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 60, 1–18 (2008).
13 Traherne, J. A., Horton, R., Robert, A. N., Miretti, M. M., Hurles, M. E., Stewart, C. A. et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies novel clumping of segments in recent human history. PLoS Genet. 2, 189 (2006).
14 Alsfieic, A., Jorgensen, A. L., Williamson, P. R., Chadwick, D. W., Park, B. K. & Pirmohamed, M. HLA-B locus in Caucasian patients with carbamazepine hypersensitivity. Pharmacogenomics 7, 813–818 (2006).
15 Hung, S. I., Chung, W. H., Jee, S. H., Chen, W. C., Chang, Y. T., Lee, W. R. et al. Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. Pharmacogenet. Genomics 16, 297–306 (2006).
16 Hetherington, S., Hughes, A. R., Mosteller, M., Shortino, D., Baker, K. L., Spreen, W. et al. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet 359, 1121–1122 (2002).
17 Mailal, S., Nolan, D., Witt, C., Masel, G., Martin, A. M., Moore, C. et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet 359, 727–732 (2002).
18 Markert, A. M., Nolan, D., Almeida, J., Lechleitner, A., Jorgensen, A. L. et al. Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. Proc. Natl Acad. Sci. USA 101, 4180–4185 (2004).
19 Saag, M., Balu, R., Phillips, E., Brachman, P., Martorell, C., Burman, W. et al. High sensitivities of human leukocyte antigen-b*5701 as a marker for immunologically confirmed abacavir hypersensitivity in white and black patients. Clin. Infect. Dis. 46, 1111–1118 (2008).
20 Dainichi, T., Uchi, H., Maroi, Y. & Furue, M. Stevens-Johnson syndrome, drug-induced hypersensitivity syndrome, and antinuclear epiternal necrolysis caused by allopurinol in patients with a common HLA allele: what causes the diversity? Dermatology 215, 86–88 (2007).
21 Gabriel, C., Danzer, M., Hackl, C., Kopel, G., Hufnagl, P., Hofer, K. et al. Rapid high-throughput human leukocyte antigen typing by massively parallel sequencing for high-resolution allele identification. Hum. Immunol. 70, 960–964 (2009).
22 Bentley, G., Higuchi, R., Holgund, B., Goodridge, D., Sayer, D., Trachtberg, E. A. et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens 77, 206–217 (2011).
23 Lind, C., Fermiö, D., Mackiewicz, K., Heron, S., Rogers, M., Slavich, L. et al. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum. Immunol. 71, 1033–1042 (2010).
24 Lank, S. M., Wissman, R. W., Dudley, D. M. & O’Connor, D. H. A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping. Hum. Immunol. 71, 1011–1017 (2010).
25 Erlich, R. L., Jia, X., Anderson, S., Banks, E., Gao, X., Carrington, M. et al. Next-generation sequencing for HLA typing of class I loci. BMC Genomics 12, 62 (2011).
26 Holcomb, C. L., Holgund, B., Anderson, M. W., Blake, L. A., Bohrme, I., Egholm, M. et al. A multi-site study using high-resolution HLA genotyping by next-generation sequencing. Tissue Antigens 77, 206–217 (2011).
27 Washington, K., Krishnakumar, S., Wibthong, J., Babrabaren, F., Slepyanu, L., Su L. F. et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc. Natl Acad. Sci. USA 109, 8676–8681 (2012).
28 Shinya, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari A. et al. Super high resolution for single molecule-sequence-based typing of classical HLA loci at the D-digit level using next generation sequencers. Tissue Antigens 80, 305–316 (2012).
29 Lank, S. M., Gollbach, B. A., Creager H. M., Wisteman, R. W., Keskin, D. B., Reinherz, E. L. et al. Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed (DNA) amplicon pyrosequencing. BMC Genomics 13, 378 (2012).
30 Moonsamy, P. V., Williams, T., Bonella, P., Holcomb, C. L., Holgund, B. N., Milligan, G. M. et al. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplification library preparation. Tissue Antigens 81, 141–149 (2013).
31 Ringqvist, S., Bellone, G., Lu, Y., Roeder, K. & Trucco, M. Clustering and alignment of polymorphic sequences for HLA-DRB1 genotyping. PLoS ONE 8, e59835 (2013).
32 Danzer, M., Niklas, N., Stabentheiner, S., Hofer, K., Pröll, J., Stückler, C. et al. Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. BMC Genomics 14, 221 (2013).
33 Hosomichi, K., Jinam, T. A., Mitsuhasha, S., Nakao, K. H. & Inoue, I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics 13, 418 (2012).
34 Trachtengber, E. A. & Holcomb, C. L. Next-generation HLA typing using the 454 GS FLX system. Methods Mol. Biol. 1034, 197–219 (2013).
35 Ozaki, Y., Suzuki, S., Shigenari, A., Okudaira, Y., Kikkawa, E., Oka A. et al. HLA-DRB1, -DQB1, -DRB3, -DRB5 genotyping at a super-high resolution level using next generation sequencers. Tissue Antigens 83, 10–16 (2013).
36 Hajeer, A. H., Al Balwi, M. A., Aytül Uyar, F., Alheidan, Y., Alabdalrahman, A. I. Abdulkareem, I. et al. HLA-A, -B, -C, -DRB1 and -DQB1 allele and haplotype
Impact of NGS on the study of HLA genes

K Hosomichi et al.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/