Seminars in Immunopathology (2022) 44:15–28
https://doi.org/10.1007/s00281-021-00901-9

REVIEW

HLA imputation and its application to genetic and molecular fine-mapping of the MHC region in autoimmune diseases

Tatsuhiko Naito[1,2] · **Yukinori Okada**[1,3,4]

Received: 11 June 2021 / Accepted: 22 October 2021 / Published online: 16 November 2021

Abstract

Variations of human leukocyte antigen (HLA) genes in the major histocompatibility complex region (MHC) significantly affect the risk of various diseases, especially autoimmune diseases. Fine-mapping of causal variants in this region was challenging due to the difficulty in sequencing and its inapplicability to large cohorts. Thus, HLA imputation, a method to infer HLA types from regional single nucleotide polymorphisms, has been developed and has successfully contributed to MHC fine-mapping of various diseases. Different HLA imputation methods have been developed, each with its own advantages, and recent methods have been improved in terms of accuracy and computational performance. Additionally, advances in HLA reference panels by next-generation sequencing technologies have enabled higher resolution and a more reliable imputation, allowing a finer-grained evaluation of the association between sequence variations and disease risk. Risk-associated variants in the MHC region would affect disease susceptibility through complicated mechanisms including alterations in peripheral responses and central thymic selection of T cells. The cooperation of reliable HLA imputation methods, informative fine-mapping, and experimental validation of the functional significance of MHC variations would be essential for further understanding of the role of the MHC in the immunopathology of autoimmune diseases.

Keywords HLA · HLA imputation · MHC · Fine-mapping · Autoimmune diseases

Introduction

The major histocompatibility complex (MHC) region is located at 6p21.3 with spanning approximately 5 Mb in length [1]. The genes encoded by this region are clearly enriched for immune responses and inflammatory pathways [1, 2]. Consistently with its function, genetic variants in the MHC region contribute to the genetics of various human complex traits, especially autoimmune diseases and infectious diseases [3, 4]. The MHC is the region with the highest number of disease associations reported in genome-wide association studies (GWAS) [5]. These associations included those "non-autoimmune diseases," such as cardiovascular, metabolic, and neurological diseases, implying immune-related mechanisms behind the progression of these diseases and the broader significance of the MHC region [6, 7]. Among the genes densely present in the MHC region, human leukocyte antigen (HLA) genes are considered to explain most of the genetic heritability of MHC. HLA molecules mediate antigen presentation, which is a critical component in triggering the subsequent immune responses; thus, variations in HLA genes have been considered to associate with the risk of immune-related diseases directly. For a representative instance, in type 1 diabetes (T1D), the MHC region explains 42.8% of phenotypic variance, of which *HLA-DRB1*, *-DQA1*, and *-DQB1* haplotypes account for the most significant proportion at 29.6% [8].

This article is a contribution to the special issue on: Genetics and functional genetics of Autoimmune diseases - Guest Editors: Yukinori Okada & Kazuhiko Yamamoto

✉ Tatsuhiko Naito
tnaito@sg.med.osaka-u.ac.jp

1 Department of Statistical Genetics, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Osaka, Suita 565-0871, Japan

2 Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

3 Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan

4 Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita, Japan

Associations of single nucleotide polymorphisms (SNPs) with phenotypes of interest in GWAS typically do not indicate their direct causal roles but linkage with truly causal variants. To identify such causal variants (i.e., fine-mapping), comprehensive genotyping of regional variations including HLA allelic types for the target individuals is needed. However, the MHC region is one of the most challenging regions of the human genome to genotype because of its high degree of polymorphism and structural variations [9]. Thus, HLA typing is conducted with specific approaches, including traditional polymerase chain reaction (PCR)-based methods and next-generation sequencing (NGS). They are so labor-intensive, time-consuming, and expensive that they could not be applied to fine-mapping for large cohorts of GWAS [6, 10]. Subsequently, the genotypes of HLA alleles are indirectly imputed from SNP-level data using a pre-constructed HLA reference panel. HLA imputation has successfully contributed to the fine-mapping of causal HLA variants to delineate of the immunopathology of various diseases.

Beginning with a simple inference using tag SNPs [11, 12], various statistical HLA allelic imputation methods have been developed, each with its advantages and disadvantages for practical use. In this review, we discuss the recent advances and challenges in HLA imputation methods and available HLA reference panels. We also discuss the relationship between the MHC region and autoimmune diseases revealed by the fine-mapping and the current understanding of how HLA variations contribute to disease etiology.

Structure and definition of HLA

The MHC region is categorized into three sub-regions, namely, class I, II, and III (Fig. 1a) [1]. In the MHC class I region, three categories of genes are located: classical HLA class I genes (*HLA-A*, *-B*, and *-C*), non-classical HLA class I genes (*HLA-E*, *-F*, *-G*, *HFE*, and 12 pseudogenes), and the class I-like genes (*MICA*, *MICB*, and 5 pseudogenes). In the MHC class II region, there are two categories of HLA genes: classical HLA class II genes (*HLA-DR*, *-DP*, and *-DQ*) and non-classical HLA class II genes (*HLA-DM* and *-DO*). The remaining part is the class III region, where many of the genes are related to the immune system, such as complement (e.g., *C2*, *C4A*, and *C4B*) and inflammation system (e.g., *TNF*).

HLA class I molecules are expressed on the surface of nucleated cells and can present endogenous antigens to CD8+ T cells. While classical HLA class I genes are highly polymorphic and have distinct antigen-presenting ability, non-classical HLA class I genes are less polymorphic and have various functions. The structure of HLA class I molecules consists of a heavy chain consisting of three domains,

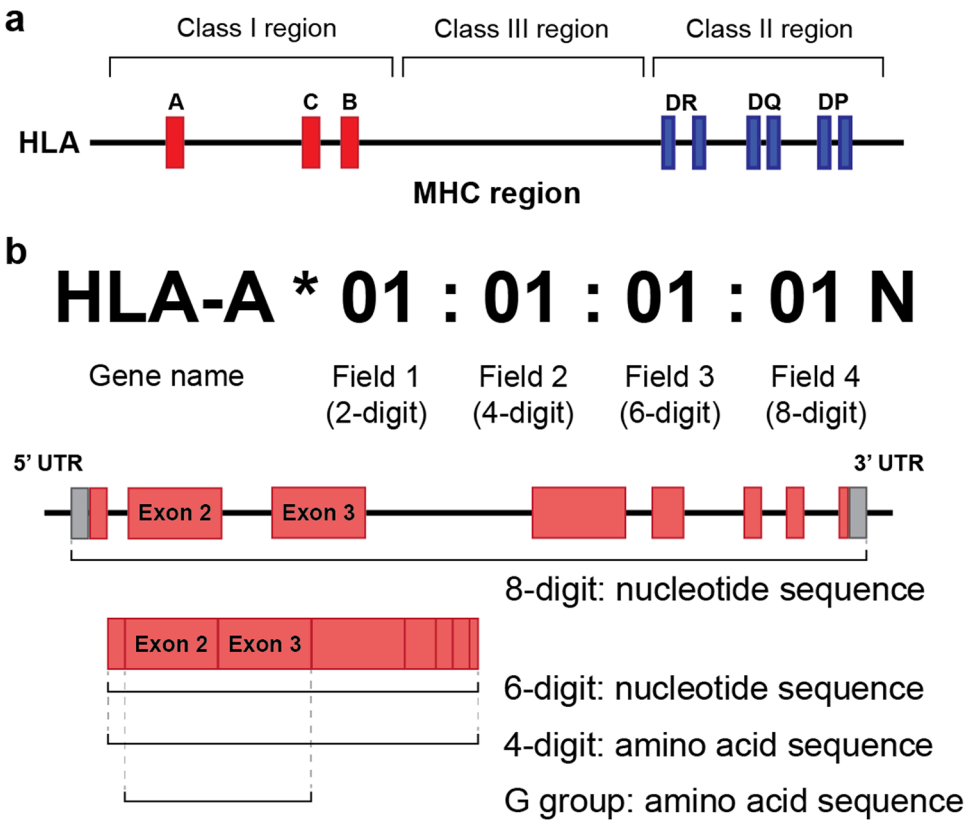

Fig. 1 Structure of the MHC region and nomenclature of HLA alleles. **a** The MHC region is categorized into class I, II, and III. Only classical HLA genes are illustrated along with their positions for simplicity. **b** The nomenclature of HLA alleles. HLA alleles are named hierarchically as four fields based on the resolution of sequences. The last letter denotes expression status, e.g., "N" indicates "not to be expressed."

α1, α2, and α3, and β2 microglobulin that constitutes one immunoglobulin-like domain.

HLA class II molecules are expressed on the surface of antigen-presenting cells, such as macrophages and dendritic cells, and function to present exogenous antigens to CD4+ T cells. The structure of HLA class II molecules consists of an alpha chain composed of two domains, α-domain consisting of α1 and α2 and β chain consisting of β1 and β2. Each HLA class I molecule (e.g., A, B, and C) is encoded by a single gene (e.g., *HLA-A*, *-B*, and *-C*, respectively). In contrast, for HLA class II, the heterodimer is formed from the products of two genes, e.g., *HLA-DQA1* and *HLA-DQB1* encode the α and β chains of DQ molecules, respectively. Although β chains of DR molecules are encoded by *HLA-DRB1*, there are additional loci encoding alternative DRβ chains in some haplotypes (e.g., *HLA-DRB3*, *-DRB4*, and *-DRB5*). The presence of the additional loci depends on the serogroup of *HLA-DRB1* gene on the same haplotype and named accordingly (e.g., *HLA-DRB4* corresponds to HLA-DRB1*04).

The rapid increase in the number of identified HLA alleles has led to the development of the nomenclature used to describe them. It is managed by the WHO Nomenclature Committee for Factors of the HLA System, and all identified HLA alleles are registered in the IMGT/HLA database [13]. In the HLA allele nomenclature, the HLA gene name is followed by numeric fields separated by colons that describe four levels of typing resolution (Fig. 1b). The first field or 2-digit resolution describes a serologically defined allele group, and the second field or 4-digit resolution indicates a unique protein sequence encoded by the allele within that group. Fields 3 and 4 resolutions show silent and non-coding polymorphisms, respectively. Traditionally, HLA typing was mainly based on the antigen-binding region (i.e., exons 2 and 3 for an HLA class I gene and by exon 2 for an HLA class II gene). However, with the increase in the number of HLA types registered in the database, many polymorphisms have been found outside the antigen-binding region. As a result, the G group was defined as a type of group in which the sequence of the antigen-binding region (i.e., exons 2 and 3 for class I and exon 2 for class II genes) to differentiate it from the 4-digit resolution allele [14].

HLA imputation methods for individual genotype data

HLA imputation was developed to fine-map the MHC region, characterized by complicated linkage disequilibrium (LD) structures and long-range haplotypes of regional variants and their corresponding HLA allelic types. HLA imputation uses a reference panel typed with both HLA and SNP genotypes to infer HLA genotypes from SNP information (Fig. 2). Starting with a simple inference using tag SNPs [11, 12], various HLA imputation methods have been developed to capture the complicated LD structure of the MHC region (Table 1).

Leslie et al. first reported a probabilistic approach for classical HLA allelic imputation based on the Li and Stephens haplotype model [21]. Its improved version was implemented as HLA*IMP targeted for the European population [22]. The Li and Stephens haplotype model is a theory of statistical genetics, stating that the genome sequence of an individual can be represented by recombination and a small number of mutations of those of other individuals [23]. They modeled the SNP haplotype background of individual HLA alleles and performed Bayesian inference to determine genotypes of HLA alleles [23]. Dilthey et al. developed a subsequent software program, HLA*IMP:02, which uses a haplotype graph approach with SNP data from multiple populations to address haplotypic heterogeneity [15]. HLA*IMP:02 is currently available in Thermo Fisher Scientific software for samples typed with its SNP-genotyping array. HLA*IMP:03 is web-based software, which uses random forest models [16]. SNP2HLA adopts an innovative approach in which multi-alleles of HLA genes are viewed as individual binary alleles and are imputed using Beagle, standard SNP genotype imputation software based on a haplotype graph approach [17]. One of its advantages is that SNP2HLA imputes HLA types and amino acid allele genotypes simultaneously. HIBAG (HLA Genotype Imputation with Attribute Bagging) estimates the likelihood of HLA alleles by the ensemble of multiple classifiers that model

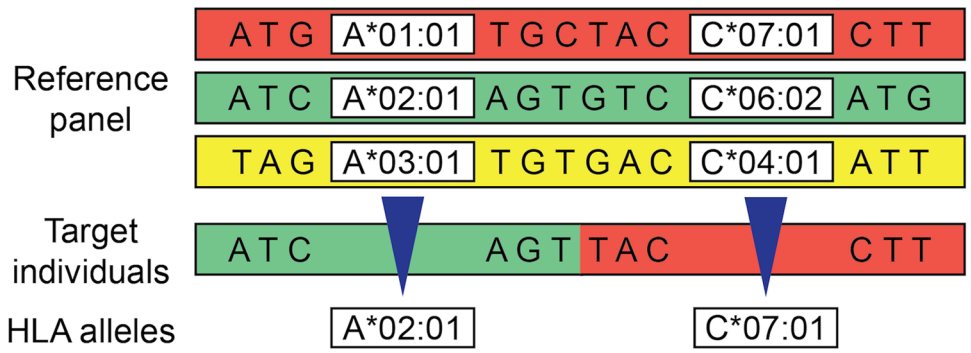

Fig. 2 An illustration of HLA imputation using a reference panel. An HLA reference panel contains individual data for which both SNP genotypes and HLA typing information are available. Based on LD information from a reference panel, it is possible to infer HLA allelic information of target individuals for whom only SNP genotype information is available.

Table 1 Comparison of HLA imputation software

Name	Type	Methods	URL	Reference
HLA*IMP:02	Stand-alone software	Haplotype-graph model	NA	[15]
HLA*IMP:03	Web application	Random forest model	http://imp.science.unimelb.edu.au/hla/	[16]
SNP2HLA	Shell script	Beagle with considering markers as binary alleles	http://software.broadinstitute.org/mpg/snp2hla/	[17]
HIBAG	R package	Bagging of multiple classifiers of EM algorithm	https://github.com/zhengxwen/HIBAG	[18]
CookHLA	Python script	Beagle with considering markers as binary alleles and embedding of markers on exons	https://github.com/WansonChoi/CookHLA	[19]
DEEP*HLA	Python script	Multi-task convolutional deep neural networks	https://github.com/tatsuhikonaito/DEEP-HLA	[20]

haplotypes and their frequencies based on expectation-maximization algorithm [18].

For widely used software, overall accuracy for high-quality reference panels is greater than 90% [24]. However, their accuracy tends to significantly decline as alleles were less frequent [19, 20]. Additionally, imputation accuracy in hyper-multi-allelic genes, such as *HLA-B* and *HLA-DRB1*, drops. In contrast, recently developed techniques presented their improvement of accuracy in such respects. CookHLA is similar to SNP2HLA in that it treats the multi-allelic HLA information as a set of binary markers but has several updates [19]. While SNP2HLA places each marker set in the center position of the gene, CookHLA embeds each marker set in the middle position of each polymorphic exon (i.e., exons 2, 3, and 4 for class I genes; and exons 2 and 3 for class II genes). It addresses the issue of LD decay with distance by effectively capturing the information of polymorphic exons. CookHLA repeats imputation for each exon and combines the posterior probabilities to make final consensus calls. Furthermore, CookHA uses Beagle v4 instead of v3, which was built in SNP2HLA. CookHLA achieved higher accuracy than SNP2HLA and HIBAG with significant superiority for less frequent alleles. For instance, CookHLA achieved 80% accuracy in alleles in frequency 0.1–0.5% for a European reference panel, whereas conventional methods presented 40−60% accuracy.

DEEP*HLA is also a recently published software, which uses a deep learning model to capture the complex LD structure of the MHC region. It utilizes the advantage of multi-task convolutional neural networks [20], which takes SNP input and impute alleles of multiple HLA genes belonging to the same preset group simultaneously (Fig. 3). Conventional imputation algorithms based on the Markov model of sequential information would show limited performance for imputing alleles without distant-dependent LD decay features. In contrast, DEEP*HLA was less dependent on distant-dependent LD decay, thanks to the nature of neural networks. DEEP*HLA was advantageous, especially for its low-frequency and rare alleles. It achieved around 80% accuracy for alleles with a frequency < 1% in most settings, while conventional methods presented 60−70% accuracy. Furthermore, DEEP*HLA was computationally efficient enough to be applied to biobank-scale data.

One aspect that determines which imputation software should be used is whether you have an HLA reference panel for a target population. HLA*IMP:02 and HLA*IMP:03 are pretrained with their reference data; thus, there is no need for your own. In contrast, the current version of HLA*IMP:02 and HLA*IMP:03 does not support a function for users to generate an imputation model using their own data locally. While SNP2HLA and CookHLA explicitly use reference haplotype data always, HIBAG and DEEP*HLA do not require these data once the trained models are generated. Since it is difficult to restore genotype information of individuals from the model parameters, their trained models could be publicly distributed or moved without ethical permission.

The formats of genotype data and HLA reference and nomenclature of HLA alleles are often unorganized, so that HLA association analysis has been laborious. HLA-TAPAS (HLA-Typing At Protein for Association Studies) is a sophisticated integrated pipeline, including data formatting, HLA reference panel construction, HLA imputation, and HLA association analysis. The imputation method of HLA-TAPAS adopts that of SNP2HLA in which Beagle v4 is used, unlike the original SNP2HLA software [25, 26]. As is the case with CookHLA, Beagle v4 supports multithreading, which would make it applicable for biobank-scale data.

HLA imputation for GWAS summary statistics

While privacy and ethical constraints often restrict access to individual GWAS genotype data, sharing GWAS summary statistics has become more prevalent. Imputation of summary statistics has been developed in this context [27]. In summary, statistics-based imputation and associations

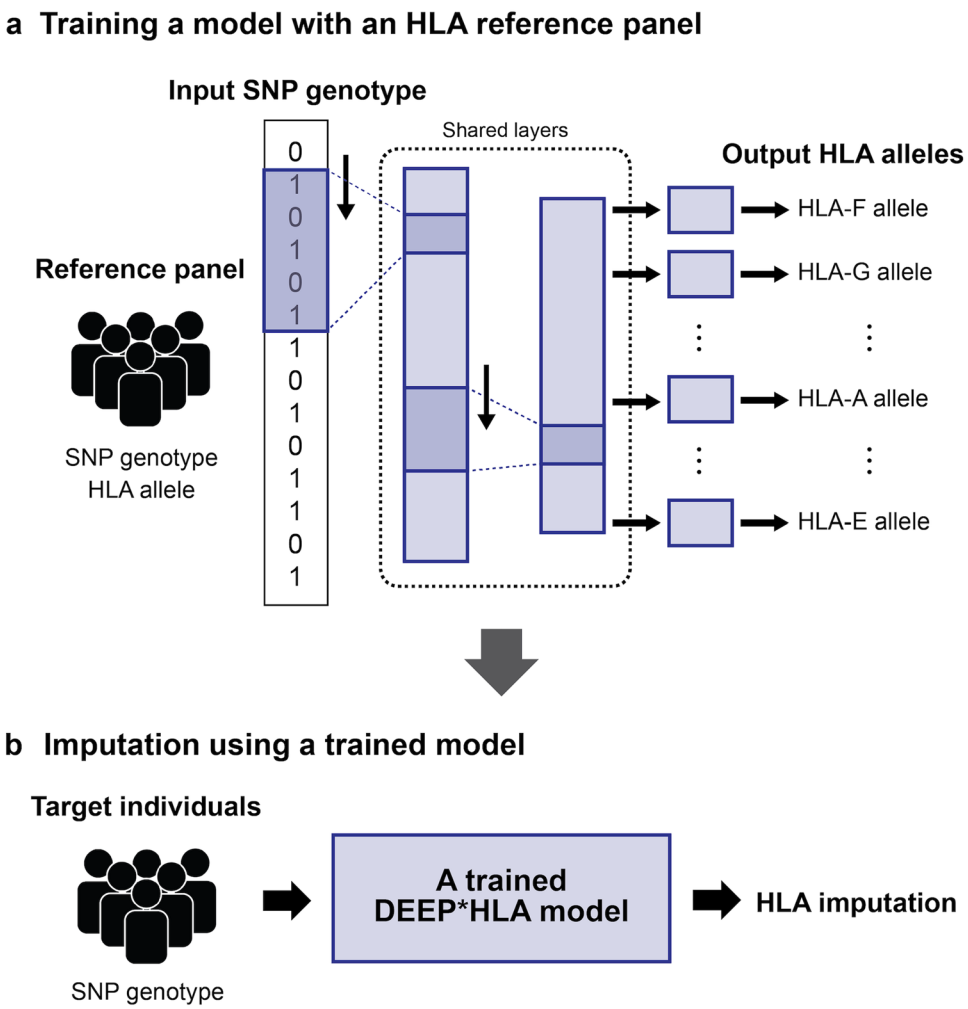

Fig. 3 The architecture and imputation strategy of DEEP*HLA. DEEP*HLA is a multi-task deep convolutional neural network model that takes SNP information and outputs genotype probabilities of HLA genes. In the training phase, DEEP*HLA learns the relationship between SNP genotype and HLA alleles from an HLA reference panel consisting of many individuals (**a**). In the imputation phase, a trained DEEP*HLA can perform HLA imputation from SNP genotype data without a reference panel (**b**).

of untyped alleles are inferred from the approximation of Z scores to a multivariate normal distribution. Li et al. extended this algorithm to the imputation of HLA association tests as DISH (Direct Imputing Summary association statistics of HLA variants) software [28]. Although it might be challenging to perform a detailed fine-mapping, such as haplotype analysis, reliable conditional analysis, and additional adjustment of covariates, DISH would play a sufficient role in obtaining meaningful inferences. We recently conducted trans-ethnic fine-mapping for Parkinson's disease (PD) by integrating GWAS summary statistics from different studies to detect functionally plausible risk-associated HLA variants [7].

Existing HLA imputation reference panels and challenges

A high-quality HLA reference panel is a prerequisite to achieve high-performance of HLA imputation. We summarized available HLA imputation reference panels in Table 2. In traditional HLA reference panels, HLA typing was conducted with PCR-based genotyping methods [17, 32, 34]. PCR-based HLA typing methods are limited to G group resolution or frequent alleles.

In contrast, high-throughput of NGS technologies and sophisticated HLA type inference algorithms has enabled higher resolution typing using either exons only or larger gene segments, including whole HLA genes [37, 38]. Reliance on a single reference sequence would be problematic in the assembly since MHC haplotypes have significant variations in genomic contents and length [1]. Thus, as exemplified by the genome graph-based methods developed by Dilthey et al. [39, 40], HLA type inference methods focusing on sequence diversity have been developed [41, 42]. Explanation of HLA type inference methods would be beyond the scope of this review and has been well discussed elsewhere [43]. Recently published HLA reference panels have mainly used NGS technologies. Kim et al. constructed a Korean reference panel with a hybrid method of the SSO method for *HLA-DRB1* and NGS for other HLA genes [30]. Zhou et al. performed deep sequencing of the MHC region

Table 2 A list of existing HLA reference panels

Population	HLA typing method	Sample size	URL	Year	Reference
Europeans	SSOP	5225	http://software.broadinstitute.org/mpg/snp2hla/	2013	[29]
Korean	NGS	413	https://sites.google.com/site/scbaehanyang/hla_panel	2014	[30]
East and South Asians	SSOP	530	http://software.broadinstitute.org/mpg/snp2hla/	2014	[31]
Japanese	SSOP	908	https://humandbs.biosciencedbc.jp/hum0028-v2	2015	[32]
Han-Chinese	NGS	10,689	http://gigadb.org/dataset/100156	2016	[33]
Japanese	NGS	1120	https://humandbs.biosciencedbc.jp/en/hum0114-v2	2019	[6]
Finnish	SSOP, SSP, SBT	1150	NA	2020	[34]
Europeans	NGS	401	NA	2020	[35]
Taiwanese	NGS	1012	NA	2020	[36]
Multi-ethnicity	NGS	21,546	https://github.com/immunogenomics/HLA-TAPAS/	2021	[26]

SSOP sequence-specific oligonucleotide probe, *NGS* next-generation sequencing, *SSP* sequence-specific primer, *SBT* sequencing-based typing

and constructed a Han-Chinese reference panel containing 10,689 normal healthy controls [33]. Hirata et al. constructed a Japanese reference panel based on NGS, which covered high-resolution allelic information of the extended MHC region, including non-classical HLA genes [6]. Squire et al. also constructed a reference panel mainly of Europeans, which included non-classical HLA and HLA-like genes [35].

Conventional HLA reference panels have been constructed for a single ancestry, considering that the LD and haplotype structure of the MHC region is highly ancestry specific. Thus, the ancestral background of a reference panel must be as close as possible to a target population. In contrast, a multiethnic reference panel is constructed with the expectation that its diversity could cover ethnically heterogeneous populations. Degenhardt et al. constructed a multiethnic reference panel by integrating multiple existing single-ancestry reference panels and demonstrated that it could maintain high accuracy across different ethnicities [29]. Luo et al. constructed a large-scale multiethnic reference panel (n = 21,456), including Admixed African, East Asian, European, and Latino, with short-read whole sequencing data. It achieved an accuracy of >90% at the G group resolution in all the ancestries. This reference panel is also publicly available in the Michigan Imputation Server, allowing direct imputation using this panel [44].

The currently available reference panels based on short-read NGS technologies could not capture the entire MHC region in a haplotype-preserving manner. Assembly of complete MHC haplotypes has been challenging due to several reasons, such as sequence homologies between the HLA genes and large structural variants, especially in the MHC class II region and class III region (e.g., C4 genes). This is problematic since such regions with high levels of structural variations may have a greater effect on the risk of some diseases. Indeed, it has been shown that structural variants in the C4 genes can largely explain the MHC genetic risk of schizophrenia [45]. Furthermore, Kamitaki et al. revealed that the C4 allele associated with the risk of schizophrenia could have a protective effect on SLE and SjS by imputation of C4 structural haplotypes using WGS data [46]. They demonstrated that the C4 gene variant explains the risk of SLE rather than its tagged HLA-DRB1*03:01, which had been presumed as the risk itself. In these diseases, the MHC genetic predisposition might be attributed not to precise interactions to specific autoantigens but to the continuous interaction of the immune system with a large number of potential autoantigens, which are modulated by C4 protein. These observations imply that fine-mapping using the current reference panels might distort our interpretation of the role of HLA variants in the etiology of some diseases and thus need to be improved.

In contrast, long-read sequencing technologies have been attracting attention as a potential method to solve these problems. Koren et al. reported the trio binning-based assembly of a diploid MHC with perfect HLA typing results [47]. Equivalent results were obtained with nanopore ultra-long reads [48]. Based on a combination of state-of-the-art long and short reads, Chin et al. produced a high-quality diploid MHC assembly of HG002, one of the GIAB benchmarking samples [49]. Future reference panels will take advantage of these technologies to realize more thorough MHC fine-mapping. Considering the current cost of the long-read sequencing technologies, the next step might be the flow of information from a relatively small number of long-read-sequenced samples to the refinement of HLA reference panels constructed by short-read sequencing, which could be used for HLA imputation from SNP array data [43].

Findings obtained from fine-mapping of the MHC region

In 2010, the association of HIV controllers was first mapped to specific amino acids of HLA class I genes through HLA imputation [50]. In 2012, as one of the first representative studies of MHC fine-mapping on an autoimmune disease using HLA imputation, the risk of rheumatoid arthritis (RA) in European populations was mapped to independent associations of amino acid alleles in HLA class I and II genes [51]. While the risk-HLA alleles had been conventionally reported for a set of amino acid positions 70–74 in HLA-DRβ1 (i.e., "shared epitope"), the study showed that the strongest association was fine-mapped to amino acid position 11 in HLA-DRβ1.

Since then, HLA imputation has successfully contributed to the fine-mapping in the MHC region on various autoimmune diseases. The risk amino acid positions of RA were replicated in East-Asian populations [31, 52], and additionally, the risk contributions of *HLA-DOA* were also identified [53]. The MHC risk of other autoimmune diseases was also fine-mapped, including systemic lupus erythematous (SLE) [52, 54], dermatomyositis [55], idiopathic inflammatory myositis [56], juvenile idiopathic arthritis [57], Sjögren's syndrome (SjS) [58], polyangiitis [59, 60], ankylosing spondylitis (AS) [61], psoriasis [33, 62, 63], celiac disease [64], T1D [8, 65], Graves' disease [32], inflammatory bowel diseases [66, 67], pulmonary alveolar proteinosis (PAP) [68], primary biliary cholangitis [69–71], and multiple sclerosis [72, 73]. Attempts to find novel insights by inter-ethnic comparison or integration by trans-ethnic fine-mapping have also been made for T1D [20] and ulcerative colitis [74].

For most common autoimmune diseases, the major risk-associated HLA loci were known from epidemiological studies; thus, fine-mapping studies have contributed to the confirmation of such associations at the level of specific variants and the additional identification of independent associations in different HLA loci. For instance, a fine-mapping study of AS revealed that not only the well-known HLA-B*27 alleles but also different *HLA-B* alleles were associated with the risk [61]. Interestingly, the stratified analysis demonstrated that a variant in *ERAP1*, which encodes a protein involved in peptide trimming in HLA class I presentation, was correlated with the risk in carriers of specific *HLA-B* alleles. In addition to *HLA-B*, variants in *HLA-A*, *-DPB1*, and *-DRB1* were independently associated with the risk of AS. In some diseases, the association with HLA itself has been proven by fine-mapping. PAP is a rare disease, in which autoimmunity to pulmonary surfactant contributed to the pathogenesis. An MHC find-mapping study first revealed that a specific *HLA-DRB1* allele confers its major genetic risk [68]. We summarized the current findings on risk-associated HLA loci and independent associations of specific HLA variants for autoimmune diseases obtained from MHC fine-mapping studies in Table 3. We note that these findings may be updated through refinement of reference panels as mentioned in the example of C4 [46] or functional fine-mapping as described later.

Not confined to so-called autoimmune diseases, MHC fine-mapping studies have successfully identified the risk HLA variants of different diseases, such as infectious diseases [75, 76], malignant tumors [77, 78], and neurological diseases [7, 79]. These studies could expand our knowledge of the involvement of autoimmunity in the progression of such diseases. For instance, in PD, a neurodegenerative disease characterized by the deposition of protein aggregates containing α-synuclein, GWAS, and fine-mapping studies have suggested an association of the HLA variants with the risk [80, 81]. Subsequently, an aberrant T cell response to α-synuclein associated with *HLA-DRB1* alleles was revealed in PD patients, significantly advancing the understanding of the role of acquired immunity in the pathogenesis of PD [82]. Furthermore, phenome-wide fine-mapping of the MHC region revealed a wide variety of associations and the relationships among different phenotypes [6, 76, 83].

Current procedure and challenges in fine-mapping

A typical procedure of fine-mapping of the MHC region enables both exploration of independent HLA loci and disentanglement of independently associated variants in the locus. For instance, in the MHC fine-mapping on RA [51], the strongest associations were mapped to the *HLA-DRB1* region, followed by the *HLA-B* and *HLA-DPB1* regions by step-wise conditional analysis, wherein the locus with the strongest association are successively conditioned on. Then, independently associated variant sets consisting of amino acid or HLA alleles were detected by step-wise conditional analysis in individual loci. Independent effects of single variants are evaluated in an additive model, in which the effects of the two alleles on a disease of interest are independent and combine linearly.

In contrast, non-additive effects statistically mean deviation from this linear relationship, which may arise from interactions between two alleles or individual alleles' inherent effects [84]. In terms of the function of HLA, an individual's two expressed HLA alleles with different antigen-binding repertoires are speculated to present a synergic effect on antigen-presentation ability, leading to an extraordinary disease risk. MHC fine-mapping analysis has also been used to elucidate non-additive and interaction effects. Lenz et al.

Table 3 Major findings on HLA associations in autoimmune diseases obtained from MHC fine-mapping studies

Disease	HLA class (top association)	HLA loci associated with risk	Independently associated HLA variants	Population	Reference
Rheumatoid arthritis	II	*HLA-DRB1*, *HLA-B*, *HLA-DPB1*	HLA-DRβ1 AA 11 and 13, 71, 74; HLA-B AA 9; HLA-DPβ1 AA 9	European	[51]
	II	*HLA-DRB1*, *HLA-B*, *HLA-DPB1*	HLA-DRβ1 AA 11 and 13, 57, 74; HLA-B AA 9; HLA-DPβ1 AA 9	East Asian	[31]
	II	*HLA-DRB1*	HLA-DRβ1 AA 11 and 13	East Asian	[52]
	II	*HLA-DRB1*, *HLA-DPB1*, *HLA-DOA*, *HLA-B*	HLA-DPβ1 AA 84; rs378352 (HLA-DOA); HLA-B*40:02	East Asian	[53]
Systemic lupus erythematous	II	*HLA-DRB1*	HLA-DRβ1 AA 11 and 13	East Asian	[52]
	II	*HLA-C*, *HLA-B*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*	HLA-DRB1∗15:03; HLA-DQB1∗02:02, 03:19; HLA-DQA1∗05:01 02:01, 05:05; HLA-B∗08:01; HLA-C∗17:01	African	[54]
	II	*HLA-DQB1*, *HLA-B*, *HLA-DRB3*, *HLA-DQA1*	HLA-DQB1∗02:01; HLA-B∗08:01, 18:01; HLA-DRB3∗02; HLA-DQA1∗01:02	European	[54]
Dermatomyositis	II	*HLA-DPB1*	HLA-DPB1*17	East Asian	[55]
Idiopathic inflammatory myositis	II	*HLA-DRB1*, *HLA-B*, *HLA-DQB1*	HLA-DRB1*03:01; HLA-B*08:01; HLA-DQβ1 AA 57; HLA-DQB1*04:02	European	[56]
Juvenile idiopathic arthritis	II	*HLA-DRB1*, *HLA-DPB1*, *HLA-A*, *HLA-B*	HLA-DRβ1 AA 13; HLA-DPB1*02:01; HLA-A AA 95; HLA-B AA 152	European	[57]
Sjögren's syndrome	II	*HLA-DQB1*	HLA-DQB1*0201	European	[58]
Granulomatosis with polyangiitis	II	*HLA-DPB1*	HLA-DPB1*04	European	[59]
Eosinophilic granulomatosis with polyangiitis	II	*HLA-DRB1*, *HLA-DQA1*, *HLA-DRB1*	HLA-DRB1*08:01; HLA-DQA1*02:01; HLA-DRB1*01:03	European	[60]
Ankylosing spondylitis	I	*HLA-B*, *HLA-A*, *HLA-DPB1*, *HLA-DRB1*	HLA-B*27, 07:02 and 57:01; HLA-A*02:01; rs1126513 (HLA-DPB1); HLA-DRB1*01:03	European	[61]
Psoriasis	I	*HLA-C*, *HLA-B*, HLA-DPB1	HLA-C*06:02, 07:04; HLA-B AA 9, 67; HLA-DPB1*05:01	East Asian	[33]
	I	*HLA-A*, *HLA-C*, *HLA-DQB1*	HLA-A*02:07; HLA-C*06:02; HLA-DQβ1 AA 57	East Asian	[62]
Celiac disease	II	*HLA-DQB1*, *HLA-DQA1*	HLA-DQβ1 AA 74, 57; HLA-DQα1 AA 47, 25	European	[63]
Type I diabetes	II	*HLA-DQB1*, *HLA-DRB1*, *HLA-B*, *HLA-A*	HLA-DQβ1 AA 57, HLA-DRβ1 AA 13, 71; HLA-B*39:06; HLA-DPB1*04:02; HLA-A AA 62	European	[8]
	II	*HLA-DQB1*, *HLA-DRB1*, *HLA-A*, *HLA-C*	rs1770 (HLA-DQB1); HLA-DRβ1 AA 74, 11; HLA-A AA 9; HLA-C AA 275	East Asian	[65]
	II	*HLA-DQB1*, *HLA-DRB1*, *HLA-B*, *HLA-A*	HLA-DQβ1 AA 185, 30, 70; HLA-DRβ1 AA 71, 74; HLA-B*54:01; HLA-A AA 62	East Asian, European	[20]
Grave's disease	II	*HLA-DPB1*, *HLA-A*, *HLA-B*, *HLA-DRB1*	HLA-DPβ1 AA 35, 9; HLA-A AA 9; HLA-B AA 45, 67; HLA-DRβ1 AA 74	East Asian	[32]

Table 3 (continued)

Disease	HLA class (top association)	HLA loci associated with risk	Independently associated HLA variants	Population	Reference
Crohn's disease	II	*HLA-DRB1*, *HLA-C*	HLA-DRB1*01:03; HLA-C*06:02	European	[66]
	II	*HLA-DRB1*, *HLA-DQB1*	HLA-DRβ1 AA 37, 57, HLA-DRB1*04:03	East Asian	[67]
Ulcerative colitis	II	*HLA-DQA1*, *HLA-DRB1*, *HLA-C*	rs6927022(HLA-DQA1); HLADRB1*01:03; HLA-C*12:02	European	[66]
Pulmonary alveolar proteinosis	II	*HLA-DRB1*	HLA-DRB1*08:03; HLA-DPβ1 AA 8	East Asian	[68]
Primary biliary cholangitis	II	*HLA-DRB1*, *HLA-DPB1*	HLA-DRB1*08 and 14; HLA-DPB1*03:01	European	[69]
	II	*HLA-DPB1*, *HLA-DRB1*	HLA-DPβ1 AA 11; HLA-DRβ1 AA 74; HLA-DQβ1 AA 57; HLA-C AA 155; HLA-DQα1 AA 13	European	[70]
	II	*HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*	HLA-DRβ1 AA 74; HLA-DQβ1 AA 55; HLA-DPβ1 AA 85, 55	East Asian	[71]
Multiple sclerosis	II	*HLA-DRB1*, *HLA-A*, *HLA-DPB1*, *HLA-B*	HLA-DRB1*15:01, 03:01, 13:03, 04:04, 04:01, 14:01; HLA-A*02:01; HLA-DPβ1 AA 65; rs2516489; HLA-B*37:01, 38:01	European	[72]
	II	*HLA-DRB1*, *HLA-A*, *HLA-DPB1*, *HLA-B*, *HLA-DQA1*, *HLA-DQB1*	HLA-DRB1*15:01, *03:01, *13:03, *08:01; HLA-A*02:01, rs9277565 (HLA-DPB1); HLA-B*44:02, *38:01, *55:01; rs2229029 (LTA)	European	[73]

AA amino acid

comprehensively evaluated the non-additive and interaction effects of several autoimmune diseases. They reported non-additive effects explained by interactions between specific HLA alleles in RA, T1D, and celiac disease [84]. Hu et al. reported that several combinations of haplotypes of *HLA-DRB1*, *-DQA1*, and *-DQB1* presented an association with the risk on T1D beyond an additive effect [8]. Interaction effects between MHC and non-MHC genes have also been reported in several autoimmune diseases, such as interaction with cytotoxic T lymphocyte antigen 4 (*CTLA4*) [85], several killer immunoglobulin receptor (KIR) genes [86, 87], and *ERAP1* and *2* [88, 89]. Variations in the KIR region can be imputed by a method similar to HLA imputation [90]. Thus, hybrid fine-mapping in the MHC region and KIR region would further our understanding on how they interactively associate with the risk of autoimmune diseases.

While focusing on the nominal statistical significance of all the variants in the MHC region is currently a standard approach, weighting or filtering of variants based on their functional annotations is an effective approach for fine-mapping in normal genomic regions [91]. Some variants have an eQTL effect on HLA genes [92, 93]. Furthermore, differential allelic expression of HLA genes has been reported in association with the etiology of several diseases [94, 95]. Considering these observations, functional fine-mapping would be helpful in further understanding the role of variations in the MHC region for disease etiology. Due to the difficulty in mapping short-reads of the highly polymorphic MHC loci and quantification of HLA gene expressions [96, 97], eQTL database is lacking, especially for different populations. The construction of an eQTL database for HLA genes based on a state-of-the-art NGS technology and mapping strategy would be expected [96–98].

Functional contribution of the HLA risk allele

MHC fine-mapping studies have shown that the amino acid alleles composing HLA alleles are likely to have stronger associations than the classical HLA alleles [8, 51]. Typically, risk-associated amino acid polymorphisms of autoimmune diseases are located in peptide-binding grooves of HLA molecules, which are considered to lead to the altered binding affinity of HLA molecules to the autoantigen peptides [32]. The altered binding affinity by the causal HLA alleles can be experimentally validated using HLA-peptide binding assay [99, 100]. Otherwise, if the epitope of an autoantigen for a disease of interest is already known, in silico prediction tools for HLA binding affinity could also be helpful to obtain a meaningful inference [7, 101]. Antigen peptides presented by HLA molecules are recognized by T cell receptors (TCRs), leading to antigen-specific immune responses. Then, the altered interaction between HLA, peptides, and TCRs by variations in HLA alleles is presumed to influence the immune response by two major mechanisms: thymic selection of T cells and peripheral T cell response [75, 102].

In the mechanism associated with thymic selection, specific MHC/peptide–TCR interactions, which MHC variants could alter, will determine the selection of the T cell repertoire during primary tolerance events, leading to differential susceptibility to disease progression. The antigen specificity of the TCR is determined by hyper-variable complementary determining region 3 (CDR3) [103]. During T cell development in the thymus, a highly diverse CDR3 repertoire is generated through random VDJ recombination in immature T cells. In the positive selection, thymic T cells that bind moderately to MHC complexes would survive. Conversely, T cells whose TCRs bind too strongly to MHC complexes, which are likely to be self-reactive, are killed as the process of negative selection. For instance, T1D risk-associated DQ molecules present weak binding to an epitope and are likely to escape from negative selection [104, 105]. Not limited to binding affinity, thymic escape due to the protein instability of DQ molecules is also suggested to associate with the risk of T1D [106, 107]. Recently, Ishigaki et al. investigated the association between HLA allelic variations and CDR3 amino acid features through CDR3 quantitative analysis (cdr3-QTL) [102]. In this study, the HLA amino acid position that explained the most variance in CDR3 composition was position 13 in HLA-DRβ1, which is the strongest association to RA risk. The effect sizes of multiple amino acids in this position were consistent between the risk of RA and cdr3-QTL, which supported the assumption that HLA risk for RA is mediated by TCR composition in some degree. Furthermore, they integrated the risk for several autoimmune diseases throughout the MHC region as an HLA risk score and identified multiple CDR patterns associated with the risk of the diseases. Considering the overlap between cdr3-QTL and risk-associated HLA variants in autoimmune diseases, the cdr3-QTL information might be utilized as an annotation for functional fine-mapping in the MHC region.

The MHC/peptide–TCR interaction in peripheral T cell immune responses would also be influenced by altered binding affinity dependent on MHC variants and associated with disease susceptibility. For instance, citrullinated self-peptides tend to bind to RA risk-associated HLA-DR molecules stronger than non-RA risk-associated HLA-DR molecules [99]. It is unclear whether T-cell selection in the thymus or the peripheral T-cell response is the primary contributor to the pathogenesis and how they are related to each other. An interesting example regarding their relationship is the neoantigen hypothesis of the association between RA and the risk-HLA alleles [108]. The conversion of electrically positive arginine to electrically neutral citrulline at the P4 position of peptides, which interacts with the SE, significantly increases the binding affinity of SE-containing HLA-DR

molecules [99]. This finding might suggest that SE-containing HLA-DR molecules fail to induce tolerance in thymic selection under non-inflammatory conditions because they cannot bind and present peptides with arginine residues at the P4 position. Then, P4-citrullinated self-peptides can be presented by SE-containing HLA-DR molecules and induce peripheral T cell response.

Conclusions

We have discussed current procedures, recent advances, and challenges in HLA imputation methods, along with topics regarding reference panels. Since no one method outperforms the others in all aspects, it is important to understand the advantages of each method and use or integrate different methods according to the situation. In general, newer reference panels contain more information covering wider variations and higher resolution of HLA typing. Therefore, HLA imputation methods should evolve with more learning capacity and higher computational performance. We have expectations of the high learning capacity of deep neural networks as one of such methods. We also reviewed the findings obtained from fine-mapping in the MHC region and the hypothetical mechanisms of how MHC variants affect the susceptibility of autoimmune diseases. An effective approach in this field is to compare the different risks among HLA alleles and their biochemical functions validated by experimental techniques. Thus, in this sense, reliable HLA imputation methods and informative fine-mapping would be essential for further understanding of the immunopathology of autoimmune diseases.

Data Availability Not applicable.

Code availability Not applicable.

Funding This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (19H01021 and 20K21834) and AMED (JP20km0405206, JP20km0405211, and JP20km0405217), Takeda Science Foundation, and Bioinformatics Initiative of Osaka University Graduate School of Medicine, Osaka University. T.N. was supported by JSPS KAKENHI (20J12189).

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Horton R, Wilming L, Rand V et al (2004) Gene map of the extended human MHC. Nat Rev Genet 5:889–899. https://doi.org/10.1038/nrg1489
2. Shiina T, Hosomichi K, Inoko H, Kulski JK (2009) The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 54:15–39. https://doi.org/10.1038/jhg.2008.5
3. Kennedy AE, Ozbek U, Dorak MT (2017) What has GWAS done for HLA and disease associations? Int J Immunogenet 44:195–211. https://doi.org/10.1111/iji.12332
4. Dendrou CA, Petersen J, Rossjohn J, Fugger L (2018) HLA variation and disease. Nat Rev Immunol 18:325–339. https://doi.org/10.1038/nri.2017.143
5. MacArthur J, Bowler E, Cerezo M et al (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45:D896–D901. https://doi.org/10.1093/nar/gkw1133
6. Hirata J, Hosomichi K, Sakaue S et al (2019) Genetic and phenotypic landscape of the major histocompatibility complex region in the Japanese population. Nat Genet 51:470–480. https://doi.org/10.1038/s41588-018-0336-0
7. Naito T, Satake W, Ogawa K et al (2021) Trans-ethnic fine-mapping of the major histocompatibility complex region linked to Parkinson's disease. Mov Disord 36:1805–1814. https://doi.org/10.1002/mds.28583
8. Hu X, Deutsch AJ, Lenz TL et al (2015) Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat Genet 47:898–905. https://doi.org/10.1038/ng.3353
9. DYC B, VRC A, Bitarello BD et al (2015) Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 Genomes Project Phase I Data. G3 Genes|Genomes|Genetics 5:931–941. https://doi.org/10.1534/g3.114.015784
10. Erlich H (2012) HLA DNA typing: Past, present, and future. Tissue Antigens 80:1–11. https://doi.org/10.1111/j.1399-0039.2012.01881.x
11. De Bakker PIW, McVean G, Sabeti PC et al (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 38:1166–1172. https://doi.org/10.1038/ng1885
12. Monsuur AJ, de Bakker PIW, Zhernakova A et al (2008) Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. PLoS One 3:1–6. https://doi.org/10.1371/journal.pone.0002270
13. Robinson J, Mistry K, McWilliam H et al (2011) The IMGT/HLA database. Nucleic Acids Res 39:D1171–D1176. https://doi.org/10.1093/nar/gkq998
14. Nunes E, Heslop H, Fernandez-Vina M et al (2011) Definitions of histocompatibility typing terms. Blood 118:e180–e183. https://doi.org/10.1182/blood-2011-05-353490
15. Dilthey A, Leslie S, Moutsianas L et al (2013) Multi-population classical HLA type imputation. PLoS Comput Biol 9:e1002877. https://doi.org/10.1371/journal.pcbi.1002877
16. Motyer A, Vukcevic D, Dilthey A et al (2016) Practical use of methods for imputation of HLA alleles from SNP genotype data. bioRxiv 091009. https://doi.org/10.1101/091009

17. Jia X, Han B, Onengut-Gumuscu S et al (2013) Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 8:e64683. https://doi.org/10.1371/journal.pone.0064683
18. Zheng X, Shen J, Cox C et al (2014) HIBAG - HLA genotype imputation with attribute bagging. Pharmacogenomics J 14:192–200. https://doi.org/10.1038/tpj.2013.18
19. Cook S, Choi W, Lim H et al (2021) Accurate imputation of human leukocyte antigens with CookHLA. Nat Commun 12:1264. https://doi.org/10.1038/s41467-021-21541-5
20. Naito T, Suzuki K, Hirata J et al (2021) A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. Nat Commun 12:1639. https://doi.org/10.1038/s41467-021-21975-x
21. Leslie S, Donnelly P, McVean G (2008) A statistical method for predicting classical HLA alleles from SNP data. Am J Hum Genet 82:48–56. https://doi.org/10.1016/j.ajhg.2007.09.001
22. Dilthey AT, Moutsianas L, Leslie S, McVean G (2011) HLA*IMP-an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics 27:968–972. https://doi.org/10.1093/bioinformatics/btr061
23. Li, Na (Department of Biostatistics, University of Washington, Seattle W 98195), Stephens, Matthew (Department of Statistics, University of Washington, Seattle W 98195) (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165:2213–2233
24. Karnes JH, Shaffer CM, Bastarache L et al (2017) Comparison of HLA allelic imputation programs. PLoS One 12:1–12. https://doi.org/10.1371/journal.pone.0172444
25. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097. https://doi.org/10.1086/521987
26. Luo Y, Kanai M, Choi W et al (2021) A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nat Genet 53:1504–1516. https://doi.org/10.1038/s41588-021-00935-7
27. Pasaniuc B, Zaitlen N, Shi H et al (2014) Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics 30:2906–2914. https://doi.org/10.1093/bioinformatics/btu416
28. Lim J, Bae S-C, Kim K (2019) Understanding HLA associations from SNP summary association statistics. Sci Rep 9:1337. https://doi.org/10.1038/s41598-018-37840-9
29. Degenhardt F, Wendorff M, Wittig M et al (2019) Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. Hum Mol Genet 28:20782092. https://doi.org/10.1093/hmg/ddy443
30. Kim K, Bang SY, Lee HS, Bae SC (2014) Construction and application of a Korean reference panel for imputing classical alleles and amino acids of human leukocyte antigen genes. PLoS One 9:9–13. https://doi.org/10.1371/journal.pone.0112546
31. Okada Y, Kim K, Han B et al (2014) Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum Mol Genet 23:6916–6926. https://doi.org/10.1093/hmg/ddu387
32. Okada Y, Momozawa Y, Ashikawa K et al (2015) Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat Genet 47:798–802. https://doi.org/10.1038/ng.3310
33. Zhou F, Cao H, Zuo X et al (2016) Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. Nat Genet 48:740–746. https://doi.org/10.1038/ng.3576
34. Ritari J, Hyvärinen K, Clancy J et al (2020) Increasing accuracy of HLA imputation by a population-specific reference panel in a FinnGen biobank cohort. NAR Genomics Bioinforma 2:1–9. https://doi.org/10.1093/nargab/lqaa030
35. Squire DM, Motyer A, Ahn R et al (2020) MHC*IMP - imputation of alleles for genes in the major histocompatibility complex. bioRxiv 2020.01.24.919191. https://doi.org/10.1101/2020.01.24.919191
36. Huang YH, Khor SS, Zheng X et al (2020) A high-resolution HLA imputation system for the Taiwanese population: a study of the Taiwan Biobank. Pharmacogenomics J 20:695–704. https://doi.org/10.1038/s41397-020-0156-3
37. Hosomichi K, Shiina T, Tajima A, Inoue I (2015) The impact of next-generation sequencing technologies on HLA research. J Hum Genet 60:665–673. https://doi.org/10.1038/jhg.2015.102
38. Carapito R, Radosavljevic M, Bahram S (2016) Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. Hum Immunol 77:1016–1023. https://doi.org/10.1016/j.humimm.2016.04.002
39. Dilthey A, Cox C, Iqbal Z et al (2015) Improved genome inference in the MHC using a population reference graph. Nat Genet 47:682–688. https://doi.org/10.1038/ng.3257
40. Dilthey AT, Mentzer AJ, Carapito R et al (2019) HLA∗LA - HLA typing from linearly projected graph alignments. Bioinformatics 35:4394–4396. https://doi.org/10.1093/bioinformatics/btz235
41. Lee H, Kingsford C (2018) Kourami: Graph-guided assembly for novel human leukocyte antigen allele discovery. Genome Biol 19:1–16. https://doi.org/10.1186/s13059-018-1388-2
42. Kim D, Paggi JM, Park C et al (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907–915. https://doi.org/10.1038/s41587-019-0201-4
43. Dilthey AT (2021) State-of-the-art genome inference in the human MHC. Int J Biochem Cell Biol 131:105882. https://doi.org/10.1016/j.biocel.2020.105882
44. Das S, Forer L, Schönherr S et al (2016) Next-generation genotype imputation service and methods. Nat Genet 48:1284–1287. https://doi.org/10.1038/ng.3656
45. Sekar A, Bialas AR, de Rivera H et al (2016) Schizophrenia risk from complex variation of complement component 4. Nature 530:177–183. https://doi.org/10.1038/nature16549
46. Kamitaki N, Sekar A, Handsaker RE et al (2020) Complement genes contribute sex-biased vulnerability in diverse disorders. Nature 582:577–581. https://doi.org/10.1038/s41586-020-2277-x
47. Koren S, Rhie A, Walenz BP et al (2018) De novo assembly of haplotype-resolved genomes with trio binning. Nat Biotechnol 36:1174–1182. https://doi.org/10.1038/nbt.4277
48. Jain M, Koren S, Miga KH et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36:338–345. https://doi.org/10.1038/nbt.4060
49. Chin CS, Wagner J, Zeng Q et al (2020) A diploid assembly-based benchmark for variants in the major histocompatibility complex. Nat Commun 11:1–9. https://doi.org/10.1038/s41467-020-18564-9
50. Pereyra F, Jia X, McLaren PJ et al (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330(80):1551–1557. https://doi.org/10.1126/science.1195271
51. Raychaudhuri S, Sandor C, Stahl EA et al (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 44:291–296. https://doi.org/10.1038/ng.1076
52. Kim K, Bang SY, Yoo DH et al (2016) Imputing variants in HLA-DR beta genes reveals that HLA-DRB1 is solely associated with rheumatoid arthritis and systemic lupus erythematosus. PLoS One 11:7–13. https://doi.org/10.1371/journal.pone.0150283

53. Okada Y, Suzuki A, Ikari K et al (2016) Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis. Am J Hum Genet 99:366–374. https://doi.org/10.1016/j.ajhg.2016.06.019
54. Hanscombe KB, Morris DL, Noble JA et al (2018) Genetic fine mapping of systemic lupus erythematosus MHC associations in Europeans and African Americans. Hum Mol Genet 27:3813–3824. https://doi.org/10.1093/hmg/ddy280
55. Zhang CE, Li Y, Wang ZX et al (2016) Variation at HLA-DPB1 is associated with dermatomyositis in Chinese population. J Dermatol 43:1307–1313. https://doi.org/10.1111/1346-8138.13397
56. Rothwell S, Cooper RG, Lundberg IE et al (2016) Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. Ann Rheum Dis 75:1558–1566. https://doi.org/10.1136/annrheumdis-2015-208119
57. Hinks A, Bowes J, Cobb J et al (2017) Fine-mapping the MHC locus in juvenile idiopathic arthritis (JIA) reveals genetic heterogeneity corresponding to distinct adult inflammatory arthritic diseases. Ann Rheum Dis 76:765–772. https://doi.org/10.1136/annrheumdis-2016-210025
58. Lessard CJ, Li H, Adrianto I et al (2013) Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. Nat Genet 45:1284–1292. https://doi.org/10.1038/ng.2792
59. Xie G, Roshandel D, Sherva R et al (2013) Association of granulomatosis with polyangiitis (Wegener's) with HLA-DPB1*04 and SEMA6A gene variants: evidence from genome-wide analysis. Arthritis Rheum 65:2457–2468. https://doi.org/10.1002/art.38036
60. Lyons PA, Peters JE, Alberici F et al (2019) Genome-wide association study of eosinophilic granulomatosis with polyangiitis reveals genomic loci stratified by ANCA status. Nat Commun 10:5120. https://doi.org/10.1038/s41467-019-12515-9
61. Cortes A, Pulit SL, Leo PJ et al (2015) Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat Commun 6:7146. https://doi.org/10.1038/ncomms8146
62. Okada Y, Han B, Tsoi LC et al (2014) Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am J Hum Genet 95:162–172. https://doi.org/10.1016/j.ajhg.2014.07.002
63. Hirata J, Hirota T, Ozeki T et al (2018) Variants at HLA-A, HLA-C, and HLA-DQB1 confer risk of psoriasis vulgaris in Japanese. J Invest Dermatol 138:542–548. https://doi.org/10.1016/j.jid.2017.10.001
64. Gutierrez-Achury J, Zhernakova A, Pulit SL et al (2015) Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. Nat Genet 47:577–578. https://doi.org/10.1038/ng.3268
65. Zhu M, Xu K, Chen Y et al (2019) Identification of novel T1D risk loci and their association with age and islet function at diagnosis in autoantibody-positive T1D individuals: based on a two-stage genome-wide association study. Diabetes Care 42:1414–1421. https://doi.org/10.2337/dc18-2023
66. Goyette P, Boucher G, Mallon D et al (2015) High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet 47:172–179. https://doi.org/10.1038/ng.3176
67. Han B, Akiyama M, Kim KK et al (2018) Amino acid position 37 of HLA-DRβ1 affects susceptibility to Crohn's disease in Asians. Hum Mol Genet 27:3901–3910. https://doi.org/10.1093/hmg/ddy285
68. Sakaue S, Yamaguchi E, Inoue Y et al (2021) Genetic determinants of risk in autoimmune pulmonary alveolar proteinosis. Nat Commun 12:1032. https://doi.org/10.1038/s41467-021-21011-y
69. Invernizzi P, Ransom M, Raychaudhuri S et al (2012) Classical HLA-DRB1 and DPB1 alleles account for HLA associations with primary biliary cirrhosis. Genes Immun 13:461–468. https://doi.org/10.1038/gene.2012.17
70. Darlay R, Ayers KL, Mells GF et al (2018) Amino acid residues in five separate HLA genes can explain most of the known associations between the MHC and primary biliary cholangitis. PLOS Genet 14:e1007833. https://doi.org/10.1371/journal.pgen.1007833
71. Wang C, Zheng X, Tang R et al (2020) Fine mapping of the MHC region identifies major independent variants associated with Han Chinese primary biliary cholangitis. J Autoimmun 107:102372. https://doi.org/10.1016/j.jaut.2019.102372
72. Patsopoulos NA, Barcellos LF, Hintzen RQ et al (2013) Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet 9:e1003926. https://doi.org/10.1371/journal.pgen.1003926
73. Moutsianas L, Jostins L, Beecham AH et al (2015) Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet 47:1107–1113. https://doi.org/10.1038/ng.3395
74. Degenhardt F, Mayr G, Wendorff M et al (2021) Trans-ethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals shared but also ethnicity-specific disease associations. Hum Mol Genet. https://doi.org/10.1093/hmg/ddab017
75. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol 18. https://doi.org/10.1186/s13059-017-1207-1
76. Tian C, Hromatka BS, Kiefer AK et al (2017) Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun 8. https://doi.org/10.1038/s41467-017-00257-5
77. Ferreiro-Iglesias A, Lesseur C, McKay J et al (2018) Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. Nat Commun 9:1–12. https://doi.org/10.1038/s41467-018-05890-2
78. Masuda T, Ito H, Hirata J et al (2020) Fine mapping of the major histocompatibility complex region and association of the HLA-B*52:01 allele with cervical cancer in Japanese women. JAMA Netw Open 3:e2023248. https://doi.org/10.1001/jamanetworkopen.2020.23248
79. Kunkle BW, Grenier-Boley B, Sims R et al (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat Genet 51:414–430. https://doi.org/10.1038/s41588-019-0358-2
80. Hamza TH, Zabetian CP, Tenesa A et al (2010) Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet 42:781. https://doi.org/10.1038/ng.642
81. Ahmed I, Tamouza R, Delord M et al (2012) Association between Parkinson's disease and the HLA-DRB1 locus. Mov Disord 27:1104–1110. https://doi.org/10.1002/mds.25035
82. Sulzer D, Alcalay RN, Garretti F et al (2017) T cells from patients with Parkinson's disease recognize α-synuclein peptides. Nature 546:656–661. https://doi.org/10.1038/nature22815
83. Karnes JH, Bastarache L, Shaffer CM et al (2017) Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. Sci Transl Med 9:1–14. https://doi.org/10.1126/scitranslmed.aai8708
84. Lenz TL, Deutsch AJ, Han B et al (2015) Widespread non-additive and interaction effects within HLA loci modulate the risk of

autoimmune diseases. Nat Genet 47:1085–1090. https://doi.org/10.1038/ng.3379
85. Hughes T, Adler A, Kelly JA et al (2012) Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. Arthritis Rheum 64:485–492. https://doi.org/10.1002/art.33354
86. Mahmoudi M, Fallahian F, Sobhani S et al (2017) Analysis of killer cell immunoglobulin-like receptors (KIRs) and their HLA ligand genes polymorphisms in Iranian patients with systemic sclerosis. Clin Rheumatol 36:853–862. https://doi.org/10.1007/s10067-016-3526-0
87. Machado-Sulbaran AC, Ramírez-Dueñas MG, Navarro-Zarza JE et al (2019) KIR/HLA gene profile implication in systemic sclerosis patients from Mexico. J Immunol Res 2019:1–11. https://doi.org/10.1155/2019/6808061
88. Kirino Y, Bertsias G, Ishigatsubo Y et al (2013) Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. Nat Genet 45:202–207. https://doi.org/10.1038/ng.2520
89. Vitulano C, Tedeschi V, Paladini F et al (2017) The interplay between HLA-B27 and ERAP1/ERAP2 aminopeptidases: from anti-viral protection to spondyloarthritis. Clin Exp Immunol 190:281–290. https://doi.org/10.1111/cei.13020
90. Vukcevic D, Traherne JA, Næss S et al (2015) Imputation of KIR types from SNP variation data. Am J Hum Genet 97:593–607. https://doi.org/10.1016/j.ajhg.2015.09.005
91. Schaid DJ, Chen W, Larson NB (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19:491–504. https://doi.org/10.1038/s41576-018-0016-z
92. Ting JP-Y, Trowsdale J (2002) Genetic control of MHC class II expression. Cell 109:S21–S33. https://doi.org/10.1016/S0092-8674(02)00696-7
93. Kaur G, Gras S, Mobbs JI et al (2017) Structural and regulatory diversity shape HLA-C protein expression levels. Nat Commun 8:15924. https://doi.org/10.1038/ncomms15924
94. Cauli A (2002) Increased level of HLA-B27 expression in ankylosing spondylitis patients compared with healthy HLA-B27-positive subjects: a possible further susceptibility factor for the development of disease. Rheumatology 41:1375–1379. https://doi.org/10.1093/rheumatology/41.12.1375
95. Kulkarni S, Qi Y, O'hUigin C et al (2013) Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease. Proc Natl Acad Sci 110:20705–20710. https://doi.org/10.1073/pnas.1312237110
96. Aguiar VRC, César J, Delaneau O et al (2019) Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. PLoS Genet 15:e1008091. https://doi.org/10.1371/journal.pgen.1008091
97. Gutierrez-Arcelus M, Baglaenko Y, Arora J et al (2020) Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. Nat Genet 52:247–253. https://doi.org/10.1038/s41588-020-0579-4
98. Yamamoto F, Suzuki S, Mizutani A et al (2020) Capturing differential allele-level expression and genotypes of all classical HLA loci and haplotypes by a new capture RNA-Seq method. Front Immunol 11:1–14. https://doi.org/10.3389/fimmu.2020.00941
99. Hill JA, Southwood S, Sette A et al (2003) Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J Immunol 171:538–541. https://doi.org/10.4049/jimmunol.171.2.538
100. Sidney J, Southwood S, Moore C et al (2013) Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. Curr Protoc Immunol 100. https://doi.org/10.1002/0471142735.im1803s100
101. Reynisson B, Alvarez B, Paul S et al (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 48:W449–W454. https://doi.org/10.1093/nar/gkaa379
102. Ishigaki K, Lagattuta K, Luo Y, et al (2020) HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. medRxiv 7:2020.11.08.20227983
103. Jung D, Alt FW (2004) Unraveling V(D)J Recombination: insights into gene regulation. Cell 116:299–311. https://doi.org/10.1016/S0092-8674(04)00039-X
104. Levisetti MG, Lewis DM, Suri A, Unanue ER (2008) Weak proinsulin peptide-major histocompatibility complexes are targeted in autoimmune diabetes in mice. Diabetes 57:1852–1860. https://doi.org/10.2337/db08-0068
105. James EA, Kwok WW (2008) Low-affinity major histocompatibility complex-binding peptides in type 1 diabetes. Diabetes 57:1788–1789. https://doi.org/10.2337/db08-0530
106. Ettinger RA, Liu AW, Nepom GT, Kwok WW (1998) Exceptional stability of the HLA-DQA1*0102/DQB1*0602 alpha beta protein dimer, the class II MHC molecule associated with protection from insulin-dependent diabetes mellitus. J Immunol 161:6439–6445
107. Miyadera H, Ohashi J, Lernmark Å et al (2015) Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. J Clin Invest 125:275–291. https://doi.org/10.1172/JCI74961
108. Busch R, Kollnberger S, Mellins ED (2019) HLA associations in inflammatory arthritis: emerging mechanisms and clinical implications. Nat Rev Rheumatol 15:364–381. https://doi.org/10.1038/s41584-019-0219-5