Next-Generation Sequencing Based HLA Typing: Deciphering Immunogenetic Aspects of Sarcoidosis

Amit Kishore and Martin Petrek*

Department of Pathological Physiology, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czechia

Unraveling of the HLA-related immunogenetic basis of several immune disorders is complex due to the extensive HLA polymorphism and strong linkage-disequilibrium between HLA loci. A lack of in phase sequence information, a relative deficiency of high resolution genotyping including non-coding regions and ambiguous haplotype assignment make it difficult to compare findings across association studies and to attribute a causal role to specific HLA alleles/haplotypes in disease susceptibility and modification of disease phenotypes. Earlier, historical antibody and DNA-based methods of HLA typing, primarily of low resolution at antigen/alellic group levels, yielded "indicative" findings which were partially improved by high-resolution DNA-based typing. Only recently, next-generation sequencing (NGS) approaches based on deep-sequencing of the complete HLA genes combined with bioinformatics tools began to provide the access to complete information at an allelic level. Analyzing HLA with NGS approaches, therefore, promises to provide further insight in the etiopathogenesis of several immune disorders in which HLA associations have been implicated. These range from coeliac disease and rheumatological conditions to even more complex disorders, such as type-1 diabetes, systemic lupus erythematosus and sarcoidosis. A systemic disease of unknown etiology, sarcoidosis has previously been associated with numerous HLA variants and also other gene polymorphisms, often in linkage with the HLA region. To date, the biological significance of these associations has only partially been defined. Therefore, more precise assignments of HLA alleles/haplotypes using NGS approaches could help to elucidate the exact role of HLA variation in the multifaceted etiopathogenesis of sarcoidosis, including epigenetic mechanisms. NGS-based HLA analyses may be also relevant for defining variable clinical phenotypes and for predicting the disease course or the response to current/plausible novel therapies.

Keywords: HLA, next-generation sequencing, genotyping, immune diseases, disease association, sarcoidosis, molecular pathophysiology

INTRODUCTION

The human leucocyte antigen (HLA, also known as the major histocompatibility complex 'MHC') region, a 3.6 Mb segment of the human genome at 6p21.3, encodes more than 220 genes of diverse function, including 20 immune protein-coding genes. The codominant expression of HLA genes regulates the acquired immune response as HLA-class I and HLA-class II genes interact with T cell receptors (TCRs) of CD8$^+$T$_c$ and CD4$^+$T$_h$ cells, through cell-mediated and humoral immune response mechanisms, respectively. HLA gene expression is further activated by cytokines (e.g., IFN-γ, IL-4) produced during both innate and adaptive immune responses. The HLA region has been associated with numerous diseases, primarily, of an autoimmune or infectious nature (Shiina et al., 2009).

The HLA region is the most polymorphic in the human genome and contains six classical HLA genes: class I *HLA-A, -B,* and *-C* and class II *HLA-DRB1, -DQB1,* and *-DPB1* (Shiina et al., 2009; Trowsdale and Knight, 2013; Robinson et al., 2015). Polymorphisms in the extended HLA region also exhibit linkage disequilibrium (LD), including with specific classical HLA alleles. At the protein level, sequence variations in the peptide-binding cleft region of HLA-class I (α1 and α2 helixes) and HLA-class II (β1 domain) molecules ("HLA interacting domains") interact with TCRs with differential antigen-binding properties, and thus may affect antigen presentation. The class I molecules are additionally sensed by receptor genes of NK cells (such as *killer-cell Ig-like receptors* 'KIR') and by cells of monocyte lineages (*leukocyte Ig-like receptor* 'LIR') (Shiina et al., 2009; Trowsdale and Knight, 2013). Thus, HLA and non-HLA genes, their intergenic variants and haplotypes within the HLA region exhibit polygenic loci and affect expression, peptide binding, and stability of the HLA-molecules.

PCR-based HLA typing using sequence-specific primers (SSP), sequence-specific oligonucleotide probes (SSOP), and Sanger sequencing-based typing (SBT) methods have significantly improved HLA typing resolution, however, these methods possess several limitations, including time-consuming protocols, low throughput, unphased data and ambiguity (Wittig et al., 2015). Single nucleotide polymorphism (SNP) based high-throughput DNA microarray studies tend to miss several HLA variants and alternatively use HLA imputation tools (HIBAG, SNP2HLA, HLA*IMP:02, and HLA-check) to infer the classical HLA alleles and amino acid changes using the LD information existing between HLA alleles and SNP markers in the HLA region (Jeanmougin et al., 2017; Karnes et al., 2017). However, higher resolution typing information, including synonymous and non-coding variants of HLA alleles as defined by the third and fourth fields of HLA nomenclature (**Figure 1**) remains undetected. Such HLA gene variants may be reported in two formats as (a) SNP rsID and (b) with four-field WHO nomenclature (**Supplementary Table S2**). The nomenclature with both formats (e.g., *HLA-B*08:01* rs4143332) (Fingerlin et al., 2015) may be helpful in designing the routine genotyping assays, such as MassARRAY and Taqman.

NEXT-GENERATION SEQUENCING BASED HLA TYPING

HLA polymorphism was first identified and characterized using alloantibodies (antisera) against leukocytes and the microlymphocytotoxicity test was developed for serological typing of HLA antigens (tissue types) in humans (Choo, 2007; Thorsby, 2009). However, since the advent of PCR techniques, DNA-based HLA typing techniques (Shiina et al., 2009; Hosomichi et al., 2015; Robinson et al., 2015; Carapito et al., 2016; Duke et al., 2016; Zhou et al., 2016; Gandhi et al., 2017) have advanced remarkably and rapid progress was driven by the fact that DNA-typing allowed higher HLA resolution and thus provides better diagnostic utility than serotyping. Nowadays, high-resolution HLA typing is the gold standard for HLA-based clinical applications, particularly for hematopoietic stem cell transplant (HSCT) matching and has been also helpful for scientific investigations of the role of HLA in disease pathophysiology.

NGS platforms commonly in use for high-throughput sequencing and massively parallel analysis, including for HLA encompass Ion torrent (Thermo Fisher), Illumina, SOLiD (ABI), PacBio, Oxford Nanopore. In addition to capital costs, they differ in various features, including sequence length and sequencing time; see **Supplementary Table S1** for details including the strength and limitations of each platform. The major limitations of NGS owing to cost, sequencing time and technical expertise are also outlined in **Supplementary Appendix S1**. High-throughput sequencing of the HLA region has been suggested as an integrated tool for accurate detection of SNPs, InDel, CNV, splice site variations, and typing of HLA genes to finer resolution (Shiina et al., 2009; Cao et al., 2013). The technical aspects of NGS based HLA typing have been recently reviewed in detail (Hosomichi et al., 2015; Carapito et al., 2016). Briefly, the steps involved in NGS include: (i) PCR amplification of the target region (complete HLA region, full-length genes, or only certain exons of HLA genes), quantification and enzymatic purification of amplicons; (ii) library preparation by ligation of amplicons to the indexed adaptors, bead based amplicon purification and size selection; (iii) sequencing using NGS platform; and (iv) data analysis using suitable software.

For NGS, the two main clonal techniques specifically targeting the HLA region include, (i) long/mid-range PCR based isolation, and (ii) hybridization-based capture methods; however, achievement of good allelic balance and homogenous coverage along all the target genes remains a major challenge (Hosomichi et al., 2015; Carapito et al., 2016). The combined approach of NGS based advanced methods for increased sequence lengths and advanced bioinformatics tools (**Table 1**) has potentially emerged as the gold standard for accurate typing of HLA variants (up to fourth field allele resolution) and generation of full-length HLA haplotypes without phase ambiguity. Sequencing of the complete HLA region using HLA sequence libraries is mostly needed to generate a population/ancestry based database and will be useful for the studies of HLA associated diseases (Zhou et al., 2016), whereas HLA typing of the full-length or selected exons of HLA genes is more appropriate for clinical
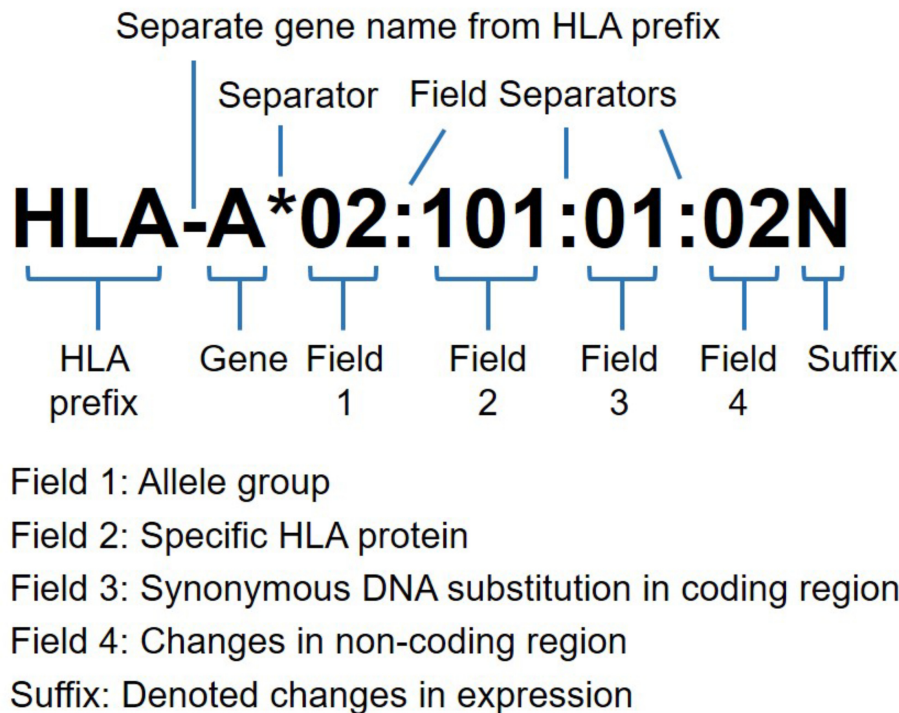
FIGURE 1 | HLA nomenclature with allele resolution at four fields (eight digits). Commonly, the HLA prefix followed by gene name and four fields of the allele. The suffix may be added to an allele to indicate its expression status as low cell surface expression "L", soluble secreted molecule, "S", present in the cytoplasm, "C", aberrant. "A", questionable, "Q" or not expressed as null alleles "N". The ambiguous allele typing for identical nucleotide sequences are coded as "G" after first three allele fields, and for identical protein sequences as "P" after two allele fields designation for exons encoding peptide binding domains (exon 2 and 3 for HLA-class I, and exon 2 only for HLA-class II alleles) (adapted from http://hla.alleles.org).

TABLE 1 | HLA typing analysis tools recently reported using next-generation sequencing data (*commercial).

Tools	Information (repository URL)	Reference
HISAT genotype	Assembles full-length sequences of HLA genes including exons and introns and performs HLA typing with alignment from IMGT/HLA database. More sensitive than other short-read aligners (http://ccb.jhu.edu/hisat-genotype)	Kim et al., 2018
xHLA	Fast and accurate HLA typing from short-read NGS data. Useful for protein level typing among both class I and II HLA genes (https://github.com/humanlongevity/HLA)	Xie et al., 2017
HLA-HD	Scans for variation through the entire sequence of HLA gene and accurately determines HLA alleles from NGS data with 6-digit precision. Compatible with low-quality NGS data. (http://www.genome.med.kyoto-u.ac.jp/HLA-HD)	Kawaguchi et al., 2017
HLAscan	An alignment-based program that determines haplotypes taking read distribution into account. Helpful for variant calling in highly polymorphic regions (http://www.genomekorea.com/display/tools/HLAscan)	(Ka et al., 2017)
*HLA twin (Omixon)	Combines two independent computational algorithms to ensure high confidence in allele calling. Consensus sequence and typing results reported as Histoimmunogenetics Markup Language (HML) format	–
*NGSengine (GenDx)	Single button analysis for NGS data provides genotyping results with minimal editing	–

purposes. Furthermore, information from NGS based HLA typing is expected to provide better insight into expression and regulation of HLA genes including epigenetic mechanisms and for an improved understanding of the "HLA-omics" of complex immune diseases (Hosomichi et al., 2015; Carapito et al., 2016).

With advancement of high-throughput HLA typing based on NGS, there has been an enormous expansion of the numbers of new HLA sequences and new alleles submitted to HLA databases (13,324 class I, 4857 class II and 182 other non-HLA alleles in the evolving Immuno Polymorphism Database-ImMunoGeneTics/HLA 'IPD-IMGT/HLA[1]' (Robinson et al., 2015; Marsh and WHO Nomenclature Committee for Factors

---

[1]www.ebi.ac.uk/ipd/imgt/hla/stats.html; http://hla.alleles.org/inc/images/graph_hires.png

of the HLA System, 2018). The common and well-documented 'CWD' catalog for HLA-class I and II alleles has been maintained to update the HLA alleles found at known frequencies (common 'C') or replicated using sequence-based typing and HLA haplotype data (well-documented 'WD')[2] (CWD v2.0.0) (Mack et al., 2013). The CWD categories also extend to 'G' and 'P' designations for different alleles with identical nucleotide and protein sequences, respectively (**Figure 1**). However, even with NGS HLA typing, challenges still remain with respect to phasing of polymorphisms leading to ambiguous results. The challenges could be compounded by PCR dropout or allelic imbalance, libraries with fragments sizes too small to achieve phasing and an inability to sequence selected HLA regions, such as exon 1 of the *DPB1* and *DRB1* genes (Carapito et al., 2016; Duke et al., 2016; Gandhi et al., 2017).

HLA AND DISEASE ASSOCIATIONS

The HLA region is associated with more than 100 multifactorial, complex diseases mainly of inflammatory and autoimmune pathogenesis. These diseases include coeliac disease, psoriasis, rheumatoid arthritis, type-1 diabetes, systemic lupus erythematosus, lung diseases (e.g., asthma and sarcoidosis), infectious diseases (e.g., HIV, hepatitis, leprosy, tuberculosis) and other conditions such as malignancies and neuropathies (Trowsdale and Knight, 2013; Carapito et al., 2016). In most cases it has been challenging to identify the principal genetic association, primarily for the following reasons: high-density of HLA genes, strong LD and effects of multiple HLA loci. However, use of NGS to fine map the HLA genes for unreported alleles and unambiguous haplotypes has already been helpful in resolving the complexity of the HLA-associated diseases and their clinical subtypes. For example, in psoriasis, new associations of *HLA-DPB1* and *BTNL2* genes and five loci among *HLA-C, -B, -DPB1,* and *BTNL2* were reported using NGS (Zhou et al., 2016). Distinct HLA associations were reported for the overall risk of psoriasis vulgaris and risk of its specific subphenotypes, such as psoriatic arthritis and cutaneous psoriasis (Okada et al., 2014).

In future, allelic resolution of HLA genes obtained by NGS will be helpful in studying the effect of variant interactions in disease risk, in selecting the participants for prevention or intervention trials (Zhou et al., 2016) and in deciphering antigen presentation mechanisms and T cell responses in chronic autoimmune disorders, such as rheumatoid arthritis (Sakurai et al., 2018). There are further important applications of NGS based HLA typing in biomedicine, namely (i) HLA-matching of donor-recipient pairs for HSCT, including HLA typing for registries of unrelated bone-marrow donors, and (ii) characterisation of individual response to drug therapy including pharmacogenetics. Further (iii) NGS will facilitate the novel strategies based on epitope matching in solid organ transplantation.

[2]http://igdawg.org/cwd.html

THE ROLE OF HLA IN SARCOIDOSIS

Sarcoidosis is a multi-systemic disease mainly affecting the lungs. It exhibits complex behavior, including the presence of multiple phenotypes, variable disease onset, and clinical course of the disease (Valeyre et al., 2014; Patterson and Chen, 2017). Sarcoidosis is believed to be the outcome of inefficient processing of yet unknown antigenic peptide(s) and their presentation by HLA with a subsequent aberrant immune response in the affected organs. This process – represented by the molecular triad (antigen presenting cells with HLA/antigen peptide/TCRs on T cells) – is further affected by the combined interaction of the genetic background of an individual and environmental pathogenic factors (Kishore and Petrek, 2013; Moller et al., 2017). Initial investigations of the role of HLA in sarcoidosis were performed by serological typing as exemplified by a report of HLA-B8 and -B13 associations with sarcoidosis in the Czech population (Levin et al., 2013). The concept of larger association studies investigating common risk HLA alleles for more than one population (ethnicity) has been demonstrated by a joint study with Czech and Italian patient cohorts, which also analyzed diverse sarcoidosis clinical phenotypes (Martinetti et al., 1995). Later, with the advancement of DNA-based HLA-typing, the extended alleles were determined as HLA-B*08:01, -B*08:04, -B*13:01, and -B*13:02 (Robinson et al., 2015; Marsh and WHO Nomenclature Committee for Factors of the HLA System, 2018); and other HLA variants implicated in sarcoidosis were identified in both European and U.S. studies (Casanova et al., 2015; Grunewald et al., 2016). The immunogenic determinants at 6p21.3, including the HLA region, reported in sarcoidosis using genome-wide association and candidate gene studies (Kishore and Petrek, 2013; Fingerlin et al., 2015; Moller et al., 2017) are briefly summarized in the following paragraph.

In the HLA region, candidate loci for sarcoidosis include HLA genes *HLA-A, -B, -DPB1, -DQB1, -DRB1, -DRB3*; and non-HLA genes *BTNL2, C4, C6orf10, HSPA1L, LTA, NOTCH4, TAP2, TNF,* and *VEGF*. The linkage of class III genes, such as *TNF* and *LTA* has been demonstrated with class I *HLA-B* and class II *HLA-DRB1*). These gene variants have been shown to associate with sarcoidosis phenotypes and sarcoidosis disease course (Grunewald et al., 2015). Association of variants *C6orf10**rs3129927 and *HLA-DRA**rs3135394 was reported for Löfgren's syndrome (LS) – a particular phenotype of more benign disease. By contrast, *BTNL2**rs2076530 and *HLA-DRA**rs3129882 was associated with the non-LS sarcoidosis (Rivera et al., 2016; Wolin et al., 2017) (**Supplementary Table S2**). To mention the further associations, *HLA-DRB1**01/*04* allelic groups have been associated with decreased sarcoidosis risk, *HLA-B**08* with increased risk, *HLA-DRB1**03:01* with LS and acute disease onset, *HLA-DRB1**12/*14* with non-LS, *HLA-DRB1**14:01* with chronic disease, HLA-*DRB1**04/*15* with the extrapulmonary involvement, *HLA-DQB1**0602* with disease progression, haplotype *HLA-DRB1**04:01-DPB1**04:01* with disease resolution (Foley et al., 2001; Kishore and Petrek, 2013; Fingerlin et al., 2015; Moller et al., 2017). For details please refer to **Supplementary Table S2**. In addition to HLA, other genes with apoptotic, enzymatic, regulatory, immune and inflammatory

functions have been implicated as candidate loci in sarcoidosis (Sato et al., 2010). These include e.g., *ANXA11, ACE, CCR2, CCR5, IL1A, IL23R, NOD2,* and *NRAMP1* (McGrath et al., 2001; Hutyrová et al., 2002; Petrek et al., 2002; Kishore and Petrek, 2013) (**Supplementary Table S3**), however, detailed discussion of these associations is beyond the scope of this mini-review.

To date, most information available concerning the involvement of the HLA sequence variation in sarcoidosis has been derived from low resolution HLA typing, to the first or maximally second field resolution. Higher resolution typing, which is now possible, particularly for non-coding sequences, will be helpful to pinpoint the specific contribution of particular HLA variants to disease susceptibility and development. In this regard, NGS based HLA typing promises to provide a valuable resource for the precise designation of HLA alleles and/or haplotypes. In near future it may be utilized for further dissecting the role of HLA variation in disease susceptibility and the development of distinct phenotypes, ranging from benign Löfgren's syndrome to progressing sarcoidosis. Thus, precise HLA typing by NGS should not only validate the previous findings and extend our knowledge of the genetic basis of sarcoidosis, but also promises to provide an insight into mechanisms of disease progression. Translational applications such as usage of HLA data obtained by NGS in clinical patient management, e.g., for prognosis, may be envisaged.

NGS BASED HLA TYPING IN POPULATION STUDIES AND ITS APPLICATION FOR INVESTIGATING HLA DISEASE ASSOCIATION

NGS based HLA typing has advantages for population studies, which are important not only for population genetics itself but also for the interpretation of disease association studies or utilization of transplantation donor registries. In all these applications, NGS can fill gaps in the existing data or could establish new population-specific HLA databases with improved accuracy (Yin et al., 2016). For example, population studies giving the prevalence of HLA alleles (CWD catalog) provide a reference to aid resolution of typing ambiguities in tissue transplantation or to determine the presence and prevalence of any HLA allele in different human populations (Mack et al., 2013; Sanchez-Mazas et al., 2017). This information can be updated and improved by NGS HLA typing. One of the first applications of NGS in population studies has been a report of a database of the complete HLA region in a Han Chinese population (Zhou et al., 2016). In a Saudi cohort of healthy individuals, the allele and haplotype frequencies of HLA genes *(HLA-A, -B, -C, -DRB1,* and *-DQB1)* were reported with improved resolution using NGS HLA typing (Hajeer et al., 2013). At targeted HLA gene level, NGS based *HLA-E* gene diversity was evaluated in West-Africans (Castelli et al., 2015). The current level of polymorphism in the HLA region characterized for both European and American populations is mostly at the second field level (Mack et al., 2013; Sanchez-Mazas et al., 2017). This is expected to be characterized with greater detail using NGS, as has recently been reported for

British and Argentinian populations (Davey et al., 2017; Hurley et al., 2018).

In the context of disease association studies, it has been suggested that risk conferred by HLA can differ between distinct populations (Ortiz-Fernández et al., 2014; Martínez-Ojinaga et al., 2018). A population-wise descriptive list of *HLA* alleles (at low resolution) as increased or decreased risk variants in disease phenotypes and disease progression in sarcoidosis has been outlined (Kishore and Petrek, 2013; Fingerlin et al., 2015). Coeliac disease represents another example of a HLA associated disease with various European populations. This disease associates with *HLA-DQA* and *-DQB* genes and about 96% of European coeliac patients carry the *HLA-DQ2.5* heterodimer *(DQA1*05-DQB1*0201)* (Karell et al., 2003; Martínez-Ojinaga et al., 2018; Núñez et al., 2018) and in some reports differed among non-Europeans, for e.g., in Brazilians and Iranians, -DQ2.2 was not associated with coeliac disease (Rostami-Nejad et al., 2014; Selleski et al., 2018). Also, the differential HLA basis of the *GIMAP* locus, which encoded proteins involved in the development of T and B lymphocytes, and Behçet's disease was described in Asian and European populations (Ortiz-Fernández et al., 2014).

An improved knowledge of HLA population genetics is therefore of great importance for investigations of particular HLA allele and/or haplotype disease associations across different ethnic groups. For this purpose, the HLA databases such as allelefrequency.net (González-Galarza et al., 2015) or specific allele and/or haplotype catalogs providing data on common alleles and/or haplotypes and genetic architecture of HLA in the population should be exploited. Their update with allele resolution HLA typing by NGS will provide valuable information including previously unmapped regions (introns, expanded exons, and other gene regulatory sequences) and will be helpful for further understanding the role of the implicated HLA variants in disease susceptibility and modification of clinical outcome. In this context, general guidelines such as STrengthening the REporting of Genetic Association studies (STREGA) and the STrengthening the REporting of Immunogenomic Studies (STREIS) statement could also be applied for analyses of HLA associations (Hollenbach et al., 2011). Further efforts in this area are underway, greatly stimulated by the activities of the 17th International HLA and Immunogenetics Workshop 'IHIW,' which evolved around NGS HLA typing and its applications, with a centralized approach for data analysis and management, including data collection, sharing and storage (Chang et al., 2017). The details for ongoing projects and updates related to HLA population genetics, such as EFI-PG, EUROSTAM, HLA-NET, and AHPD can be accessed at https://hla-net.eu/projects/.

REGULATION OF HLA GENES BY EPIGENETIC MECHANISMS

The rapidly evolving field of epigenetics deciphers the complex crosstalk between genetic factors and environmental exposure.

The main epigenetic mechanisms include DNA methylation, histone modifications and non-coding RNA (microRNA, miRNA). In the context of HLA, non-coding sequences from HLA regions, particularly the regulatory enhancers, promoter and untranslated regions (5'- and 3'-UTRs) could form a repertoire for non-genetic factors acting as epigenetic regulators of HLA genes. Primarily, the regulatory elements have differential CpG methylation and miRNA binding sites, while histone modifications are derived by specific enzymes or protein complexes. Interestingly, HLA was among the first regions explored for methylation profiling, the HLA region can be analyzed for both genetic and epigenetic features (Tomazou et al., 2008). The evidence for involvement of DNA methylation in HLA gene regulation is exemplified by the observation that high methylation of the *HLA-A* promoter region correlated with a reduced level of gene expression (Ramsuran et al., 2015). Methylation also exerts long gene distance impacts, as three CpG sites in the *HLA-DQB1* coding region are associated (non-linkage) with the promoter region of *HLA-F* (Esterhuyse et al., 2015). Furthermore, the hypermethylated *HLA-DQB1* gene in tuberculosis patients was associated with reduced expression of HLA-class II genes (Esterhuyse et al., 2015). Histone acetylation and methylation can activate *HLA-DRA* gene expression through multiple complexes of histone lysine acetyltransferase (KAT) and histone lysine methyltransferase (KMT) respectively, in a class II transactivator (CIITA) independent or dependent manner (Choi and Boss, 2012). The UTRs sequence variation may be a predictor of the genetic predisposition of an individual to express different levels of *HLA-G* as reported in several studies. In this context, *HLA-G 5'-UTR* methylation was associated with *HLA-G* gene silencing (Carosella et al., 2015) and *HLA-G 3'-UTR* with a common SNP rs1063320 was reported to modulate binding of miRNA-148/152 family and suppress soluble *HLA-G* expression in asthma (Naidoo et al., 2015).

Finally, in the context of HLA there have been reports of the involvement of yet another epigenetic mechanism – short-sequence non-coding miRNAs. Firstly, miRNA expression itself may be influenced by the presence of SNPs (miR-SNP) in the seed domain, which eventually modulates the target gene expression (Kishore et al., 2014). Similarly, genetic variation in the distal *HLA-A 3'-UTR* (100 bp insertion) affecting RNA binding protein Syncrip was shown to modulate the post-transcriptional expression of the HLA-A protein (Kulkarni et al., 2017). A concerted effect of methylation and genomic variation (14 bp insertion) in *HLA-G 3'-UTR* was also demonstrated for *HLA-G* regulation (Verloes et al., 2017). Secondly, an expanded role for HLA genes in regulating other immune responses via miRNA interaction was demonstrated by inhibition of *HLA-B* intron-encoded miR-6891-5p that impacts the expression of several transcripts, including I gA, affecting a large number of metabolic and immune response pathways (Chitnis et al., 2017). It is conceivable that investigations of miRNA in the context of HLA will continue, as exemplified e.g., by sequence characterization of

novel and haplotype-specific miRNA transcripts for complete *HLA* region using deep RNA sequencing (Clark et al., 2018).

CONCLUSION

HLA typing has already benefited and will continue to benefit from NGS applications. Comprehensive and unambiguous characterisation of extensive polymorphisms in the HLA region by NGS will contribute to progress in the area of HLA and disease associations by precise identification of these associations at the allelic level and by utilizing exact data from NGS-based population genetic studies. In sarcoidosis, the translation of expected developments to clinical practice will aim at defining new genetic markers for particular disease phenotypes and at using knowledge on the nature of yet unknown peptides presented by HLA molecules for understanding disease mechanisms, resulting in development of specifically tailored therapies. In our opinion, dynamic developments in NGS methodologies and linked bioinformatic analyses will ultimately expand clinical applications of HLA typing beyond traditional areas of transplantation and immunogenetics. Understanding the molecular mechanisms of sarcoidosis using these novel approaches represents just a small piece at the start of these endeavors.

AUTHOR CONTRIBUTIONS

MP and AK contributed to the conception and writing of this review. Both authors have reviewed and approved the submitted work.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00503/full#supplementary-material

REFERENCES

Cao, H., Wu, J., Wang, Y., Jiang, H., Zhang, T., Liu, X., et al. (2013). An integrated tool to study MHC region: accurate SNV detection and HLA genes typing in human MHC region using targeted high-throughput sequencing. *PLoS One* 8:e69388. doi: 10.1371/journal.pone.0069388

Carapito, R., Radosavljevic, M., and 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. doi: 10.1016/j.humimm.2016.04.002

Carosella, E. D., Rouas-Freiss, N., Tronik-Le Roux, D., Moreau, P., and LeMaoult, J. (2015). HLA-G: an immune checkpoint molecule. *Adv. Immunol.* 127, 33–144. doi: 10.1016/bs.ai.2015.04.001

Casanova, N., Zhou, T., Knox, K. S., and Garcia, J. G. N. (2015). Identifying novel biomarkers in sarcoidosis using genome-based approaches. *Clin. Chest. Med.* 36, 621–630. doi: 10.1016/j.ccm.2015.08.005

Castelli, E. C., Mendes-Junior, C. T., Sabbagh, A., Porto, I. O. P., Garcia, A., Ramalho, J., et al. (2015). HLA-E coding and 3' untranslated region variability determined by next-generation sequencing in two West-African population samples. *Hum. Immunol.* 76, 945–953. doi: 10.1016/j.humimm.2015.06.016

Chang, C. J., Osoegawa, K., Milius, R. P., Maiers, M., Xiao, W., Fernandez-Viña, M., et al. (2017). Collection and storage of HLA NGS genotyping data for the 17th International HLA and immunogenetics workshop. *Hum. Immunol.* 79, 77–86. doi: 10.1016/j.humimm.2017.12.004

Chitnis, N., Clark, P. M., Kamoun, M., Stolle, C., Brad Johnson, F., and Monos, D. S. (2017). An expanded role for HLA genes: HLA-B encodes a microRNA that regulates IgA and other immune response transcripts. *Front. Immunol.* 8:583. doi: 10.3389/fimmu.2017.00583

Choi, N. M., and Boss, J. M. (2012). Multiple histone methyl and acetyltransferase complex components bind the HLA-DRA gene. *PLoS One* 7:e37554. doi: 10.1371/journal.pone.0037554

Choo, S. Y. (2007). The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* 48, 11–23. doi: 10.3349/ymj.2007.48.1.11

Clark, P. M., Chitnis, N., Shieh, M., Kamoun, M., Johnson, F. B., and Monos, D. (2018). Novel and haplotype specific microRNAs encoded by the major histocompatibility complex. *Sci. Rep.* 8:3832. doi: 10.1038/s41598-018-19427-6

Davey, S., Ord, J., Navarrete, C., and Brown, C. (2017). HLA-A, -B and -C allele and haplotype frequencies defined by next generation sequencing in a population of 519 English blood donors. *Hum. Immunol.* 78, 397–398. doi: 10.1016/j.humimm.2017.04.001

Duke, J. L., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Gasiewski, A., et al. (2016). Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA* 87, 141–152. doi: 10.1111/tan.12736

Esterhuyse, M. M., Weiner, J., Caron, E., Loxton, A. G., Iannaccone, M., Wagman, C., et al. (2015). Epigenetics and proteomics join transcriptomics in the quest for tuberculosis biomarkers. *mBio* 6, 1187–1115. doi: 10.1128/mBio.01187-15

Fingerlin, T. E., Hamzeh, N., and Maier, L. A. (2015). Genetics of Sarcoidosis. *Clin. Chest Med.* 36, 569–584. doi: 10.1016/j.ccm.2015.08.002

Foley, P. J., McGrath, D. S., Puscinska, E., Petrek, M., Kolek, V., Drabek, J., et al. (2001). Human leukocyte antigen-DRB1 position 11 residues are a common protective marker for sarcoidosis. *Am. J. Respir. Cell Mol. Biol.* 25, 272–277. doi: 10.1165/ajrcmb.25.3.4261

Gandhi, M. J., Ferriola, D., Huang, Y., Duke, J. L., and Monos, D. (2017). Targeted next-generation sequencing for human leukocyte antigen typing in a clinical laboratory metrics of relevance and considerations for its successful implementation. *Arch. Pathol. Lab. Med.* 141, 806–812. doi: 10.5858/arpa.2016-0537-RA

González-Galarza, F. F., Takeshita, L. Y. C., Santos, E. J. M., Kempson, F., Maia, M. H. T., Da Silva, A. L. S., et al. (2015). Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res.* 43, D784–D788. doi: 10.1093/nar/gku1166

Grunewald, J., Kaiser, Y., Ostadkarampour, M., Rivera, N. V., Vezzi, F., Lötstedt, B., et al. (2016). T-cell receptor-HLA-DRB1 associations suggest specific antigens in pulmonary sarcoidosis. *Eur. Respir. J.* 47, 898–909. doi: 10.1183/13993003.01209-2015

Grunewald, J., Spagnolo, P., Wahlström, J., and Eklund, A. (2015). Immunogenetics of disease-causing inflammation in Sarcoidosis. *Clin. Rev. Allergy Immunol.* 49, 19–35. doi: 10.1007/s12016-015-8477-8

Hajeer, A. H., Al Balwi, M. A., Aytul Uyar, F., Alhaidan, Y., Alabdulrahman, A., Al Abdulkareem, I., et al. (2013). HLA-A, -B, -C, -DRB1 and -DQB1 allele and haplotype frequencies in Saudis using next generation sequencing technique. *Tissue Antigens* 82, 252–258. doi: 10.1111/tan.12200

Hollenbach, J. A., Mack, S. J., Gourraud, P.-A., Single, R. M., Maiers, M., Middleton, D., et al. (2011). A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of immunogenomic studies statement. *Tissue Antigens* 78, 333–344. doi: 10.1111/j.1399-0039.2011.01777.x

Hosomichi, K., Shiina, T., Tajima, A., and Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. *J. Hum. Genet.* 60, 665–673. doi: 10.1038/jhg.2015.102

Hurley, C. K., Hou, L., Lazaro, A., Gerfen, J., Enriquez, E., Galarza, P., et al. (2018). Next generation sequencing characterizes the extent of HLA diversity in an Argentinian registry population. *HLA* 91, 175–186. doi: 10.1111/tan.13210

Hutyrová, B., Pantelidis, P., Drábek, J., Zůurková, M., Kolek, V., Lenhart, K., et al. (2002). Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 165, 148–151. doi: 10.1164/ajrccm.165.2.2106004

Jeanmougin, M., Noirel, J., Coulonges, C., and Zagury, J. F. (2017). HLA-check: evaluating HLA data from SNP information. *BMC Bioinformatics* 18:334. doi: 10.1186/s12859-017-1746-1

Ka, S., Lee, S., Hong, J., Cho, Y., Sung, J., Kim, H. N., et al. (2017). HLAscan: genotyping of the HLA region using next-generation sequencing data. *BMC Bioinformatics* 18:258. doi: 10.1186/s12859-017-1671-3

Karell, K., Louka, A. S., Moodie, S. J., Ascher, H., Clot, F., Greco, L., et al. (2003). HLA types in celiac disease patients not carrying the DQA1 *05-DQB1 *02 (DQ2) heterodimer: results from the European genetics cluster on celiac disease. *Hum. Immunol.* 64, 469–477. doi: 10.1016/S0198-8859(03)00027-2

Karnes, J. H., Shaffer, C. M., Bastarache, L., Gaudieri, S., Glazer, A. M., Steiner, H. E., et al. (2017). Comparison of HLA allelic imputation programs. *PLoS One* 12:e0172444. doi: 10.1371/journal.pone.0172444

Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R., and Matsuda, F. (2017). HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. *Hum. Mutat.* 38, 788–797. doi: 10.1002/humu.23230

Kim, D., Paggi, J. M., and Salzberg, S. (2018). HISAT-genotype: next generation genomic analysis platform on a personal computer. *bioRxiv* [Preprint]. doi: 10.1101/266197

Kishore, A., Borucka, J., Petrkova, J., and Petrek, M. (2014). Novel insights into miRNA in lung and heart inflammatory diseases. *Mediators Inflamm.* 2014:259131. doi: 10.1155/2014/259131

Kishore, A., and Petrek, M. (2013). Immunogenetics of saarcoidosis. *Int. Trends Immun.* 1, 43–53.

Kulkarni, S., Ramsuran, V., Rucevic, M., Singh, S., Lied, A., Kulkarni, V., et al. (2017). Posttranscriptional regulation of HLA-A protein expression by alternative polyadenylation signals involving the RNA-binding protein syncrip. *J. Immunol.* 199, 3892–3899. doi: 10.4049/jimmunol.1700697

Levin, A. M., Iannuzzi, M. C., Montgomery, C. G., Trudeau, S., Datta, I., McKeigue, P., et al. (2013). Association of ANXA11 genetic variation with sarcoidosis in African Americans and European Americans. *Genes Immun.* 14, 13–18. doi: 10.1038/gene.2012.48

Mack, S. J., Cano, P., Hollenbach, J. A., He, J., Hurley, C. K., Middleton, D., et al. (2013). Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens* 81, 194–203. doi: 10.1111/tan.12093

Marsh SGE, and WHO Nomenclature Committee for Factors of the HLA System (2018). Nomenclature for factors of the HLA system, update March 2018. *Hum. Immunol.* 79, 516–525. doi: 10.1016/j.humimm.2018.04.007

Martinetti, M., Tinelli, C., Kolek, V., Cuccia, M., Salvaneschi, L., Pasturenzi, L., et al. (1995). "The sarcoidosis map": a joint survey of clinical and immunogenetic findings in two European countries. *Am. J. Respir. Crit. Care Med.* 152, 557–564. doi: 10.1164/ajrccm.152.2.7633707

Martínez-Ojinaga, E., Molina, M., Polanco, I., Urcelay, E., and Núñez, C. (2018). HLA-DQ distribution and risk assessment of celiac disease in a Spanish center. *Rev. Esp. Enferm. Dig.* 110, 421–426. doi: 10.17235/reed.2018.5399/2017

McGrath, D. S., Foley, P. J., Petrek, M., Izakovicova-Holla, L., Kolek, V., Veeraraghavan, S., et al. (2001). ACE Gene I/D polymorphism and sarcoidosis pulmonary disease severity. *Am. J. Respir. Crit. Care Med.* 164, 197–201. doi: 10.1164/ajrccm.164.2.2011009

Moller, D. R., Rybicki, B. A., Hamzeh, N. Y., Montgomery, C. G., Chen, E. S., Drake, W., et al. (2017). Genetic, immunologic, and environmental basis of sarcoidosis. *Ann. Am. Thorac. Soc.* 14, S429–S436. doi: 10.1513/AnnalsATS.201707-565OT

Naidoo, D., Wu, A. C., Brilliant, M. H., Denny, J., Ingram, C., Kitchner, T. E., et al. (2015). A polymorphism in HLA-G modifies statin benefit in asthma. *Pharmacogenomics J.* 15, 272–277. doi: 10.1038/tpj.2014.55

Núñez, C., Garrote, J. A., Arranz, E., Bilbao, J. R., Fernández Bañares, F., Jiménez, J., et al. (2018). Recommendations to report and interpret HLA genetic findings in coeliac disease. *Rev. Esp. Enferm. Dig.* 110, 458–461. doi: 10.17235/reed.2018.5269/2017

Okada, Y., Han, B., Tsoi, L. C., Stuart, P. E., Ellinghaus, E., Tejasvi, T., et al. (2014). Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am. J. Hum. Genet.* 95, 162–172. doi: 10.1016/j.ajhg.2014.07.002

Ortiz-Fernández, L., Conde-Jaldón, M., García-Lozano, J. R., Montes-Cano, M. A., Ortego-Centeno, N., Castillo-Palma, M. J., et al. (2014). GIMAP and Behçet disease: no association in the European population. *Ann. Rheum. Dis.* 73, 1433–1434. doi: 10.1136/annrheumdis-2013-205156

Patterson, K. C., and Chen, E. S. (2017). The pathogenesis of pulmonary sarcoidosis and implications for treatment. *Chest* 153, 1432–1442. doi: 10.1016/j.chest.2017.11.030

Petrek, M., Gibejova, A., Drabek, J., Mrazek, F., Kolek, V., Weigl, E., et al. (2002). CC chemokine receptor 5 (CCR5) mRNA expression in pulmonary sarcoidosis. *Immunol. Lett.* 80, 189–193.

Ramsuran, V., Kulkarni, S., O'huigin, C., Yuki, Y., Augusto, D. G., Gao, X., et al. (2015). Epigenetic regulation of differential HLA-A allelic expression levels. *Hum. Mol. Genet.* 24, 4268–4275. doi: 10.1093/hmg/ddv158

Rivera, N. V., Ronninger, M., Shchetynsky, K., Franke, A., Nöthen, M. M., Müller-Quernheim, J., et al. (2016). High-density genetic mapping identifies new susceptibility variants in sarcoidosis phenotypes and shows genomic-driven phenotypic differences. *Am. J. Respir. Crit. Care Med.* 193, 1008–1022. doi: 10.1164/rccm.201507-1372OC

Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., and Marsh, S. G. E. (2015). The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* 43, D423–D431. doi: 10.1093/nar/gku1161

Rostami-Nejad, M., Romanos, J., Rostami, K., Ganji, A., Ehsani-Ardakani, M. J., Bakhshipour, A. R., et al. (2014). Allele and haplotype frequencies for HLA-DQ in Iranian celiac disease patients. *World J. Gastroenterol.* 20, 6302–6308. doi: 10.3748/wjg.v20.i20.6302

Sakurai, K., Ishigaki, K., Shoda, H., Nagafuchi, Y., Tsuchida, Y., Sumitomo, S., et al. (2018). HLA-DRB1 shared epitope alleles and disease activity are correlated with reduced T cell receptor repertoire diversity in CD4 + T cells in rheumatoid arthritis. *J. Rheumatol.* 45, 905–914. doi: 10.3899/jrheum.170909

Sanchez-Mazas, A., Nunes, J. M., Middleton, D., Sauter, J., Buhler, S., McCabe, A., et al. (2017). Common and well-documented HLA alleles over all of Europe and within Europe sub-regions: a catalogue from the European federation for immunogenetics. *HLA* 89, 104–113. doi: 10.1111/tan.12956

Sato, H., Woodhead, F. A., Ahmad, T., Grutters, J. C., Spagnolo, P., van den Bosch, J. M. M., et al. (2010). Sarcoidosis HLA class II genotyping distinguishes differences of clinical phenotype across ethnic groups. *Hum. Mol. Genet.* 19, 4100–4111. doi: 10.1093/hmg/ddq325

Selleski, N., Almeida, L. M., Almeida, F. C., de Pratesi, C. B., Nóbrega, Y. K., de, M., et al. (2018). Prevalence of celiac disease predisposing genotypes, including HLA-DQ2.2 variant, in Brazilian children. *Arq. Gastroenterol.* 55, 82–85. doi: 10.1590/S0004-2803.201800000-16

Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* 54, 15–39. doi: 10.1038/jhg.2008.5

Thorsby, E. (2009). A short history of HLA. *Tissue Antigens* 74, 101–116. doi: 10.1111/j.1399-0039.2009.01291.x

Tomazou, E. M., Rakyan, V. K., Lefebvre, G., Andrews, R., Ellis, P., Jackson, D. K., et al. (2008). Generation of a genomic tiling array of the human major histocompatibility complex (MHC) and its application for DNA methylation analysis. *BMC Med. Genomics* 1:19. doi: 10.1186/1755-8794-1-19

Trowsdale, J., and Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annu. Rev. Genomics Hum. Genet.* 14, 301–323. doi: 10.1146/annurev-genom-091212-153455

Valeyre, D., Prasse, A., Nunes, H., Uzunhan, Y., Brillet, P.-Y., and Müller-Quernheim, J. (2014). Sarcoidosis. *Lancet* 383, 1155–1167. doi: 10.1016/S0140-6736(13)60680-7

Verloes, A., Spits, C., Vercammen, M., Geens, M., LeMaoult, J., Sermon, K., et al. (2017). The role of methylation, DNA polymorphisms and microRNAs on HLA-G expression in human embryonic stem cells. *Stem Cell Res.* 19, 118–127. doi: 10.1016/j.scr.2017.01.005

Wittig, M., Anmarkrud, J. A., Kässens, J. C., Koch, S., Forster, M., Ellinghaus, E., et al. (2015). Development of a high-resolution NGS-based HLA-typing and analysis pipeline. *Nucleic Acids Res.* 43:e70. doi: 10.1093/nar/gkv184

Wolin, A., Lahtela, E. L., Anttila, V., Petrek, M., Grunewald, J., van Moorsel, C. H. M., et al. (2017). SNP variants in major histocompatibility complex are associated with sarcoidosis susceptibility-A joint analysis in four European populations. *Front. Immunol.* 8:422. doi: 10.3389/fimmu.2017.00422

Xie, C., Yeo, Z. X., Wong, M., Piper, J., Long, T., Kirkness, E. F., et al. (2017). Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8059–8064. doi: 10.1073/pnas.1707945114

Yin, Y., Lan, J. H., Nguyen, D., Valenzuela, N., Takemura, P., Bolon, Y.-T., et al. (2016). Application of high-throughput next-generation sequencing for HLA typing on buccal extracted DNA: results from over 10,000 donor recruitment samples. *PLoS One* 11:e0165810. doi: 10.1371/journal.pone.0165810

Zhou, F., Cao, H., Zuo, X., Zhang, T., Zhang, X., Liu, 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. doi: 10.1038/ng.3576