High-Throughput *MICA/B* Genotyping of Over Two Million Samples: Workflow and Allele Frequencies

Anja Klussmeier[1]*, Carolin Massalski[1], Kathrin Putke[1], Gesine Schäfer[1], Jürgen Sauter[2], Daniel Schefzyk[2], Jens Pruschke[2], Jan Hofmann[2], Daniel Fürst[3,4], Raphael Carapito[5], Seiamak Bahram[5], Alexander H. Schmidt[1,2] and Vinzenz Lange[1]

[1] DKMS Life Science Lab, Dresden, Germany, [2] DKMS, Tübingen, Germany, [3] Institute of Clinical Transfusion Medicine and Immunogenetics Ulm, German Red Cross Blood Transfusion Service, Baden Wuerttemberg – Hessen, and University Hospital Ulm, Ulm, Germany, [4] Institute of Transfusion Medicine, University of Ulm, Ulm, Germany, [5] Laboratoire d'ImmunoRhumatologie Moléculaire, Plateforme GENOMAX, INSERM UMR_S 1109, LabEx TRANSPLANTEX, Université de Strasbourg, Strasbourg, France

MICA and MICB are ligands of the NKG2D receptor and thereby influence NK and T cell activity. *MICA/B* gene polymorphisms, expression levels and the amount of soluble MICA/B in the serum have been linked to autoimmune diseases, infections, and cancer. In hematopoietic stem cell transplantation, *MICA* matching between donor and patient has been correlated with reduced acute and chronic graft-vs.-host disease and improved survival. Hence, we developed an extremely cost-efficient high-throughput workflow for genotyping *MICA/B* for newly registered potential stem cell donors. Since mid-2017, we have genotyped over two million samples using NGS amplicon sequencing for *MICA/B* exons 2–5. In donors of German origin, *MICA*008* is the most common *MICA* allele with a frequency of 42.3%. It is followed by *MICA*002* (11.7%) and *MICA*009* (8.8%). The three most common *MICB* alleles are *MICB*005* (43.9%), *MICB*004* (21.7%), and *MICB*002* (18.9%). In general, *MICB* is less diverse than *MICA* and only 6 alleles, instead of 15, account for a cumulative allele frequency of 99.5%. In 0.5% of the samples we observed at least one allele of *MICA* or *MICB* which has so far not been reported to the IPD/IMGT-HLA database. By providing *MICA/B* typed voluntary donors, clinicians become empowered to include *MICA/B* into their donor selection process to further improve unrelated hematopoietic stem cell transplantation.

Keywords: MICA, MICB, hematopoietic stem cell transplantation, allele, genotyping, next generation sequencing, NGS, high-throughput

INTRODUCTION

The *MICA* (MHC class I polypeptide-related sequence A) and *MICB* (MHC class I polypeptide-related sequence B) genes are located between the MHC class I and class III genes inside the human major histocompatibility complex (MHC) (1). Although being highly similar to the classical human leukocyte antigen (*HLA*) genes, they do not present peptides and are not expressed at the surface of human leukocytes but on endothelial cells, fibroblasts, epithelial cells, and tumor cells (2). There they act as ligands for the NKG2D receptor which plays an important role in immune surveillance

by activating NK cells and co-stimulating T cell subsets (3, 4). Therefore, the expression of NKG2D ligands is highly regulated and induced by cellular stress (e.g., infection, oxidative stress, transformation).

MICA and *MICB* are highly similar and share around 91% of their coding sequence (1). Exon 1 encodes the leader peptide, exons 2, 3, and 4 the three extracellular domains, exon 5 the transmembrane domain and exon 6 the cytoplasmic tail (1, 2, 5). Even though *MICA* and *MICB* do not seem to be as diverse as the conventional *HLA* genes, a large number of distinct alleles have been described: release 3.37.0 of the IPD-IMGT/HLA database contains 109 *MICA* and 47 *MICB* alleles (6). *MICA*008* has been reported to be the most common *MICA* allele with frequencies ranging from 25 to 55% depending on the population. Frequencies above 5% were observed for *MICA*002*, *MICA*009*, *MICA*004*, *MICA*010*, and *MICA*007* in Europeans. In Chinese cohorts, the alleles *MICA*019*, *MICA*027*, and *MICA*045* are also common (7–11). The less diverse *MICB* gene has been predominantly studied in Asian populations. There, the allele *MICB*005* is the most common allele with frequencies of over 50%. It is followed by *MICB*002* and *MICB*004* with frequencies over 10% and *MICB*008* and the null allele *MICB*009N* with frequencies over 5% (10–13).

The most frequent *MICA* allele *MICA*008* differs substantially from most other alleles since it lacks the transmembrane domain due to a frameshift in exon 5. Alleles sharing this feature are also referred to as "A5.1" alleles (14). Their products are bound to the cellular membrane by a GPI-anchor and are frequently released into exosomes thereby triggering a systemic downregulation of the NKG2D receptor on effector cells. Other *MICA* and *MICB* alleles do this to a lesser extent using a soluble form caused by a proteolytic shedding mechanism (15, 16). Since high levels of both forms of soluble MICA and MICB (sMICA/B) have been found in various cancers, the release of MIC proteins is thought to be one cause for cancer immune escape. sMICA/B are therefore considered promising targets for immunotherapy (17–20).

Several studies looked into the general impact of *MICA/B* polymorphisms on different diseases. Especially the *MICA*-129Met/Val dimorphism encoded by the SNP rs1051792 has received attention because it separates the different *MICA* alleles into NKG2D-receptor low (Val)- and high (Met)-affinity binding alleles (21). Health risk associations have been shown for several autoimmune diseases, cancer and viral infections (22–27). Furthermore, matching of *MICA*, including the *MICA*-129 dimorphism, between donor and patient has been correlated with improved outcome of unrelated hematopoietic stem cell transplantation and reduced acute and chronic graft-vs.-host disease (28–32). Because *MICA* is in strong linkage disequilibrium with *HLA-B*, over 90% of 10/10 *HLA*-matched donor/patient pairs are also matched for *MICA* (8, 30). In partially matched cases, in particular in *HLA-B* mismatch situations, *MICA* mismatches are more frequent.

To facilitate further studies on *MICA* and/or *MICB* matching in unrelated hematopoietic stem cell transplantation, we included both genes into our high-throughput genotyping workflow for newly registered potential stem cell donors in 2017. This workflow was initially developed for the six classical *HLA* genes *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQB1,* and *HLA-DPB1* and was then gradually extended to also include *CCR5*, the blood groups *ABO* and *Rh* as well as the several *KIR* genes and *HLA-E* (33–37). Today, this workflow has been applied to genotype over seven million donors, among them more than two million including *MICA* and *MICB*.

MATERIALS AND METHODS

Samples
Volunteers from Germany, Poland, UK, USA, Chile and India provided over two million samples to DKMS for their registration as potential stem cell donors between August 2017 and October 2019. We determined *MICA* and *MICB* allele frequencies based on 1,201,896 samples of donors from DKMS Germany who declared to be of German descent. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The described genotyping is within the scope of the consent forms signed at recruitment and performed as genotyping service.

DNA Isolation and Quantification
The vast majority of samples were provided as buccal swabs (Copan, Brescia, Italy). Few samples were provided as blood. DNA was isolated using the chemagic™ Blood/Swab Kits (PerkinElmer chemagen Technologie GmbH, Baesweiler, Germany) and quantified by fluorescence as described before (36).

PCR Amplification
MICA and *MICB* were amplified in one multiplexed PCR reaction targeting exons 2, 3, and 4/5. The resulting amplicons had lengths between 417 and 480 bp (**Figure 1**). Exons 2 and 3 were amplified as separate amplicons and were completely covered. In contrast, exons 4 and 5 were amplified together as one joined amplicon with primers inside the exons. Therefore, 65 bases at the beginning of exon 4 and 13 bases at the end of exon 5 were not covered. The 8 μl PCR reactions were performed in 384-well plates using FastStart™ Taq DNA Polymerase (Roche, Basel, Switzerland) in its associated buffer system. After amplification, products were pooled with other amplicons of the same sample and subjected to a barcoding/indexing PCR as described previously (33–37).

Library Preparation and Sequencing
After indexing PCR, 384 barcoded samples were pooled together and purified using SPRIselect beads (BeckmanCoulter, Brea, USA) with a ratio of 0.6:1 beads to DNA and subsequently quantified by qPCR. Equimolar amounts of 10 pools were then combined to a final sequencing library which contained all amplicons from 3,840 donors. The library was denatured and diluted as recommended by Illumina (MiSeq Reagent Kit V2-Reagent Preparation Guide) and loaded at 12.5 pM onto HiSeq flow cells. Paired-end sequencing was performed at 2 × 249

bp using HiSeq Rapid SBS Kits V2 (500 cycles) on HiSeq2500 instruments (Illumina, San Diego, USA) (33–37).

Genotyping

The neXtype software was extended to support *MICA* and *MICB* genotyping (33, 36). It uses a decision-tree-based algorithm to match the generated *MICA/B* amplicons to known alleles from the official IPD/IMGT-HLA database. Since no known *MICA* amplicon sequence matches a known *MICB* amplicon sequence, reads could be unambiguously assigned to either *MICA* or *MICB*. For more than 95% of the samples neXtype generated correct results with only minor requirements for user interaction. In case of insufficient read coverage, rare or questionable results, a new PCR reaction was initiated from the original DNA. If a low read coverage was limited to exons 4 and 5, trained analysts could decide to generate a result based on exons 2 and 3 only. Genotyping results were finally exported using the GL string format (38).

Frequency Analysis of *MICA* and *MICB* Alleles

MICA and *MICB* genotyping results of 1,201,896 samples of German origin were analyzed based on the first field, which identifies the unique MICA and MICB proteins. Homozygous genotyping results were counted as two alleles. Allele groups which could not be distinguished due to missing sequencing information were reported by a representative allele which was marked with a hash symbol (#) (**Table 1**). For samples with phasing ambiguities, the probability of each possible result was calculated based on the allele frequencies of unambiguously typed samples. According to these probabilities, counts were added to the different alleles. To verify rare allele calls, all alleles observed <50 times were reconfirmed in at least two samples.

RESULTS

High-Throughput *MICA/B* Genotyping

Assay Validation and Performance

For assay validation, we exchanged DNA from 95 samples with two labs with established workflows for *MICA* or *MICB* genotyping (*MICA*: Institute of Clinical Transfusion Medicine and Immunogenetics Ulm, Germany; *MICB*: Laboratoire d'ImmunoRhumatologie Moléculaire, Strasbourg, France). For *MICA*, we additionally used the UCLA *MICA* Panel Set (UCLA Immunogenetic Center, USA), which consists of 24 samples with diverse combinations of *MICA* alleles. The results obtained from our newly established workflow were 100% concordant with the reference genotypes for both *MICA* and *MICB*

(**Supplementary File 1**). Subsequently, *MICA/B* genotyping was included into our standard genotyping workflow in August 2017 and applied for all newly registered donors. So far, we have generated *MICA/B* genotyping data for over two million samples, on average more than 20,000 samples per week. Because *MICA/B* amplicons are pooled with the *HLA* amplicons directly after the initial PCR, additional costs for genotyping *MICA/B* are minor and reflect the costs for one 8 μl PCR reaction, sequencing and data analysis. We are targeting an average coverage of 1,000 reads per locus and exon corresponding to a total of 6,000 reads for *MICA/B* with associated costs of about 10 cents per sample for sequencing. This efficient strategy makes it feasible to genotype every newly registered donor for *MICA/B*.
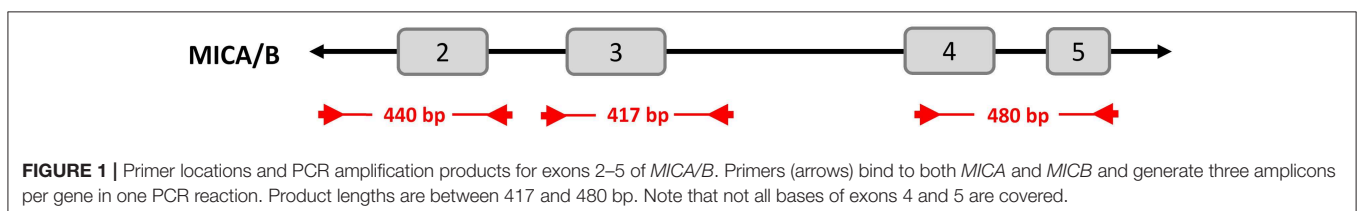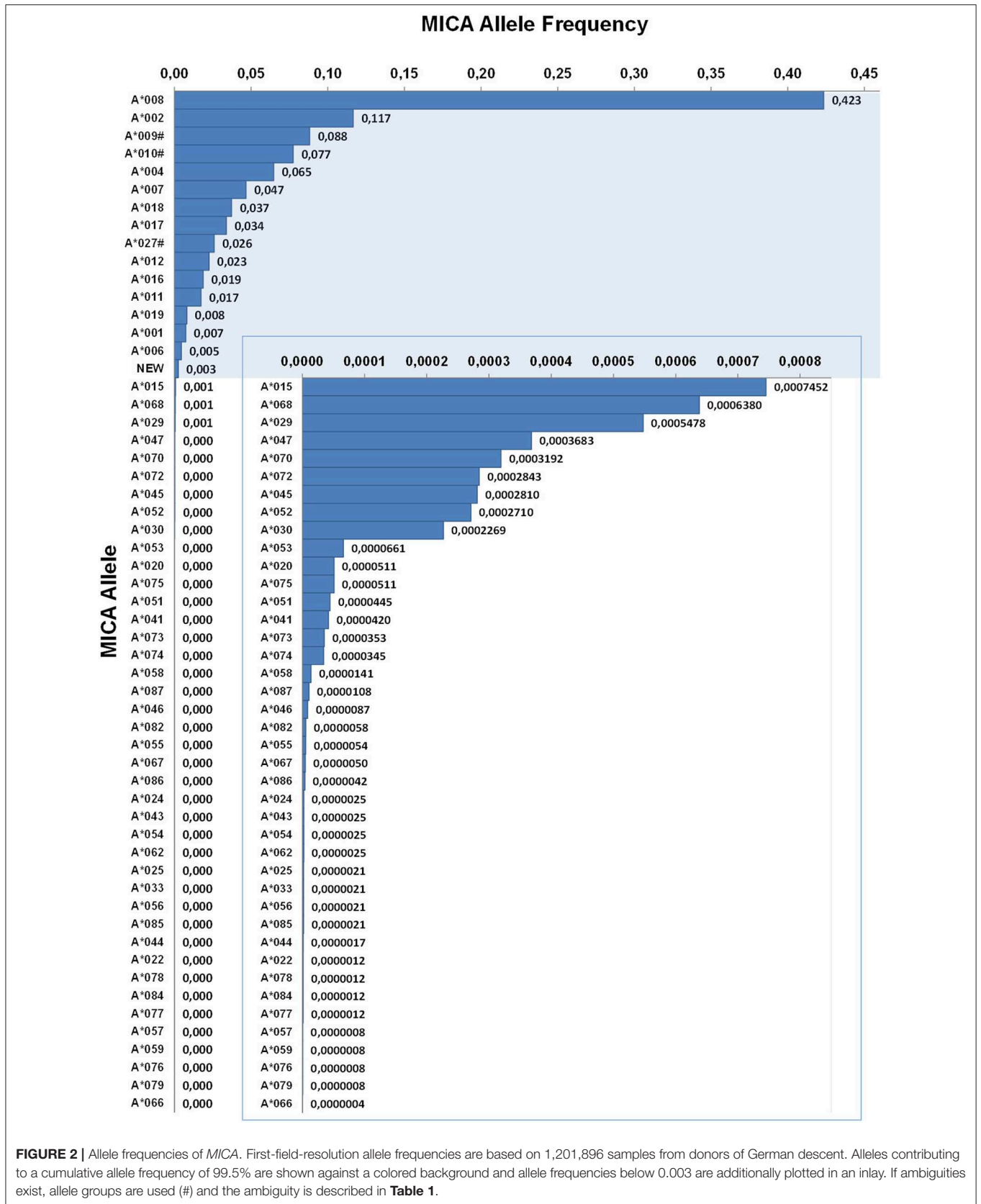
Resolution and Ambiguities

Our *MICA/B* genotyping workflow targets and amplifies exons 2 and 3 separately and most of exons 4 and 5 using a combined amplicon (**Figure 1**). Consequently, exons 1 and 6 and 78 bases of exons 4 and 5 are not sequenced. This amplification strategy promised a good genotyping resolution while being highly cost-efficient. *MICA/B* exons 2, 3, and 5 were considered mandatory because they encode the receptor-interacting domains or define *MICA*008-like alleles. Expansion of the exon 5 amplicon made it possible to also include most of exon 4. Exons 1 and 6 encode a leader peptide and the cytoplasmic tail. As these regions do not encode extracellular domains of the proteins and are characterized by a lower diversity they were not included in the genotyping strategy. However, some alleles may only be differentiated by sequence features within one of the not covered regions. For example, SNPs in exon 6 are the only way to distinguish *MICA*010 from *MICA*069 or *MICA*009:01 from *MICA*049. *MICA*009:02, on the other hand, can be unambiguously genotyped because it differs from *MICA*049 and

TABLE 1 | Overview of ambiguous genotyping results.

Allele group	Alleles
MICA	
MICA*009#	MICA*009, MICA*049
MICA*010#	MICA*010, MICA*065, MICA*069
MICA*027#	MICA*027, MICA*048
MICB	
MICB*004#	MICB*004, MICB*028
MICB*005#	MICB*003, MICB*005, MICB*006, MICB*010
MICB*014#	MICB*014, MICB*015

Alleles which cannot be distinguished from each other by the workflow are combined in an allele group marked with a hash symbol (#).

FIGURE 1 | Primer locations and PCR amplification products for exons 2–5 of *MICA/B*. Primers (arrows) bind to both *MICA* and *MICB* and generate three amplicons per gene in one PCR reaction. Product lengths are between 417 and 480 bp. Note that not all bases of exons 4 and 5 are covered.

FIGURE 2 | Allele frequencies of *MICA*. First-field-resolution allele frequencies are based on 1,201,896 samples from donors of German descent. Alleles contributing to a cumulative allele frequency of 99.5% are shown against a colored background and allele frequencies below 0.003 are additionally plotted in an inlay. If ambiguities exist, allele groups are used (#) and the ambiguity is described in **Table 1**.

its synonymous allele *MICA*009:01* in exon 3 (**Table 1**) (14). Due to the primer location inside exon 4 our workflow also cannot distinguish between *MICA*10* and *MICA*065*.

For *MICB*, the most common allele *MICB*005:02* cannot be distinguished from *MICB*003*, *MICB*006*, and *MICB*010*, while other variants of *MICB*005* can be distinguished. Likewise, the pairs *MICB*004* and *MICA*028* or *MICB*014* and *MICA*015* cannot be resolved (**Table 1**).

In addition to the ambiguities caused by missing sequence information, we encounter phasing ambiguities. They occur because the sequences of short amplicons cannot be phased if the targeted regions are not overlapping. As a consequence, some observed sequence combinations can be explained by more than one allele pair. In our workflow, phasing ambiguities occur in 3% of *MICA* and 24% of *MICB* samples. In over 99.9% of those cases, however, one possibility can statistically be ruled out since the combination of two rare alleles would be highly unlikely if the other option includes two common alleles. This is in contrast to *HLA* genotyping where some important phasing ambiguities cannot be solved statistically. For example, the most common *MICB* phasing ambiguity result is either the combination *MICB*002* and *MICB*005#* or the combination *MICB*018* and *MICB*019* [GL-String notation: *MICB*002+MICB*005#|MICB*018+MICB*019* (38)]. Based on the allele frequencies determined in this study, the likelihood of the allele combination *MICB*002+MICB*005#* is 0.039. In contrast, the likelihood of *MICB*018+MICB*019* is only 1.9 × 10⁻⁸. Hence, *MICB*018+MICB*019* would be expected to occur only once in 2.08 million samples with the given phasing result. In our dataset of 1,201,896 samples, 182,383 samples have the result *MICB*002+MICB*005#|MICB*018+MICB*019*. Now, by claiming that *MICB*002+MICB*005#* is always the correct result, we are making only one wrong call in 13.7 million genotyped samples. Therefore, we have disregarded the highly unlikely combinations of rare alleles in our allele frequency calculations. This is not expected to introduce a relevant error. In contrast, disregarding all samples with phasing results altogether would substantially distort the results since the phasing events predominantly involve certain alleles.

Novel Alleles

We encounter novel *MICA* or *MICB* alleles in 0.5% of the samples, resulting in the observation of ∼100 novel alleles per week (recurrences included). They are automatically flagged by the genotyping software and trigger a new PCR reaction from the original sample for verification. In general, the novel alleles fall into two categories: Novel sequences or novel combinations of previously reported exonic sequences. The task to characterize them in full length and submit the sequences to IPD/IMGT-HLA is currently in progress.

MICA Allele Frequencies

MICA allele frequencies were calculated on 1,201,896 samples of German descent (**Figure 2**). These samples represent more than 50% of our genotyped samples and were therefore the largest ethnically defined population available. With a frequency of 42.3%, the allele *MICA*008* is the most frequent *MICA* allele

TABLE 2 | *MICA/B* alleles described in IPD/IMGT-HLA release 3.37.0, but never observed in our cohort of over two million samples.

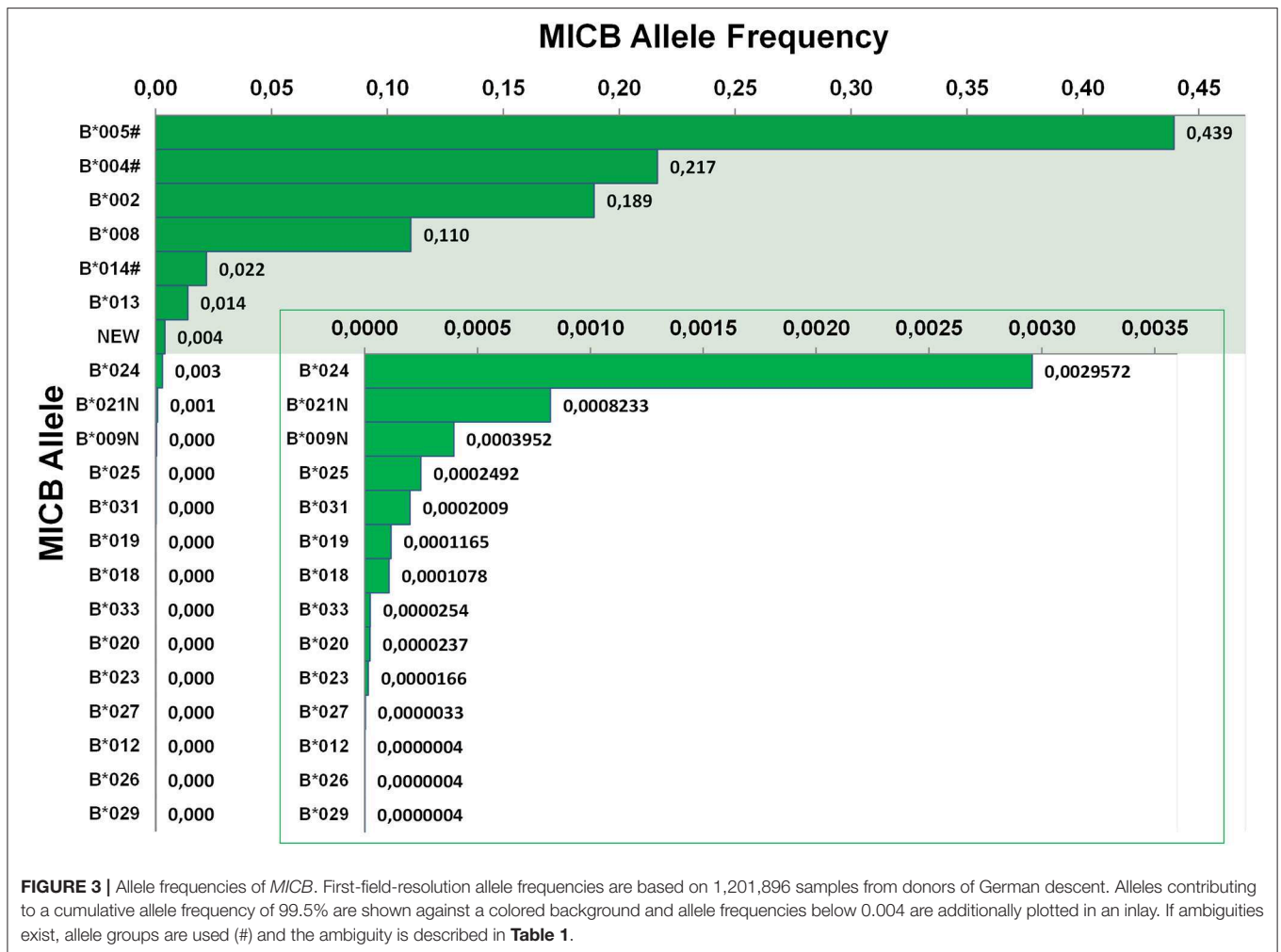|	
MICA	*MICA*005*, *MICA*013*, *MICA*014*, *MICA*023*, *MICA*026*, *MICA*028*, *MICA*031*, *MICA*032*, *MICA*034*, *MICA*036*, *MICA*039*, *MICA*042*, *MICA*050*, *MICA*061*, *MICA*063N*, *MICA*065*, *MICA*081*, *MICA*083*
MICB	*MICB*001*, *MICB*011*, *MICB*016*, *MICB*022*, *MICB*030*, *MICB*032*

in Germany. It is followed by the alleles *MICA*002* (11.7%), *MICA*009#* (8.8%), *MICA*010#* (7.7%), and *MICA*004* (6.5%). The 15 most common alleles account for a cumulative allele frequency of 99.5%. The other 41 alleles observed in the German dataset account for the remaining 0.5%. We further identified six *MICA* alleles (*MICA*035*, *MICA*037*, *MICA*038*, *MICA*040*, *MICA*060*, and *MICA*064N*) with very low frequencies in samples not of German origin. Despite the huge sample size, we have never observed the remaining 18 alleles contained in the IPD-IMGT/HLA database (release 3.37.0) (**Table 2**).

MICB Allele Frequencies

MICB allele frequencies were calculated based on the same sample cohort used for *MICA* (**Figure 3**). With a frequency of 43.9%, *MICB*005#* is by far the most frequent allele in Germany. However, since our workflow cannot distinguish all *MICB*005* variants from *MICB*003*, *MICB*006*, and *MICB*010*, the true frequency of *MICB*005* might be lower (**Table 1**). In our samples, *MICB*005#* is followed by *MICB*004#*, *MICB*002*, and *MICB*008*, having frequencies of 21.7 18.9, and 11.0%, respectively. Together with *MICB*014#* (2.2%) and *MICB*013* (1.4%) they account for a cumulative allele frequency of 99.5%. 14 other alleles have been detected in the German cohort. *MICB*007* has only been identified in a few samples of non-German origin. We have never observed the six remaining alleles described in the IPD-IMGT/HLA database (release 3.37.0) (**Table 2**).

DISCUSSION

The regulation of NK/T cell activation is an elaborate interplay between several receptors and their associated ligands. To further add another layer of complexity, receptors like KIR or ligands like MICA/B exist in a variety of distinct alleles with varying effects on NK/T cell activity (6, 39). A comprehensive sequencing study of the MHC complex indicated that the sequence of *MICA* is more diverse than the sequence of *HLA-DQB1* or *HLA-DPB1*, but the number of named *MICA* alleles is much lower (6, 10). And even though MICA/B do not present antigenic peptides like the classical *HLA* class I genes, matching of *MICA/B* between patient and donor has been reported to improve outcome and reduce acute and chronic graft-vs.-host disease in hematopoietic stem cell transplantation, especially in partially matched scenarios (30, 31, 40). Translation of these findings into clinical practice is, amongst others, hampered by the lack of *MICA/B* genotyping data. Hence, we present a workflow to genotype both *MICA* and *MICB* with a mean throughput of over 20,000 samples

FIGURE 3 | Allele frequencies of *MICB*. First-field-resolution allele frequencies are based on 1,201,896 samples from donors of German descent. Alleles contributing to a cumulative allele frequency of 99.5% are shown against a colored background and allele frequencies below 0.004 are additionally plotted in an inlay. If ambiguities exist, allele groups are used (#) and the ambiguity is described in **Table 1**.

per week. To date, we have processed more than two million donor samples.

Based on 1.2 million samples of German origin we identified *MICA*008* as the most common *MICA* allele (42.3%), followed by *MICA*002* (11.7%) and *MICA*009#* (8.8%). This is concordant to previous studies which present allele frequencies between 43 and 55% for *MICA*008*, 8–14% for *MICA*002* and 4–8% for *MICA*009* in European/American populations (7–9). Although *MICA*008* is also the most common allele in China, with a frequency of about 25% it is far less abundant than in European/American populations (10, 11, 41). Since *MICA*008* and other rare alleles bearing the A5.1 microsatellite marker are more prone to produce sMICA than other alleles, they are more effective in inactivating NKG2D and NK/T cell activity (15). Therefore, these alleles might contribute to the disease prevalence in different populations. Indeed, A5.1-carriers have been associated with an increased risk for several types of cancer and higher levels of sMICA seem to have a negative prognostic value for tumor patient survival (18, 27, 42–44). To reactivate a patient's NK cells, the reduction of soluble NKG2D ligands is a promising approach. Current strategies comprise the inhibition

of enzymes responsible for shedding as well as blocking the cleavage sites with therapeutic antibodies. Most likely, the efficacy of some of these new drugs will be limited to certain *MICA/B* alleles which increases the need for reliable genotyping (18, 45).

MICB is less diverse than *MICA*. The most common allele *MICB*005#* was detected at 43.9% allele frequency in the German population. However, given the incomplete sequence coverage, our workflow cannot distinguish *MICB*003*, *MICB*005*, *MICB*006,* and *MICB*010*. Studies on Asian cohorts report allele frequencies of at least 55% for *MICB*005*, 3% for *MICB*003* and no observations of *MICB*006* or *MICB*010* (10, 11, 13). Limited full gene analysis of 51 samples with *MICB*005#* pre-typing results indicated a similar distribution in our dataset (data not shown).

The *MICB*003/005:02* ambiguity with its distinguishing bases at the beginning of exon 4 and in exon 6 is one case in which our workflow cannot differentiate between two presumably common alleles. However, an amplicon of at least 530 bp would be necessary to include the SNP at the beginning of exon 4 and to not lose sequencing information for the microsatellite region in *MICA* exon 5. Since this exceeds Illumina's 2 × 250 bp read

length, bases at the end of exon 4 would not be sequenced, thereby creating other ambiguities. Consequently, to clearly distinguish between *MICB*005:02* and *MICB*003* a separate fourth PCR amplicon would be required. But given the lack of clinical data for the relevance of regions outside exons 2, 3, and 5, one might wonder if a higher resolution for *MICA/B* genotyping is necessary. In *HLA* genotyping transplantation compatible allele groups have been defined (G or P Codes) combining all alleles harboring the same sequence across the antigen recognition domain (2, 46, 47). For *MICA/B*, there is no similar system yet. Consequently, we do not think that it is proportionate to increase the sequencing costs for all samples without further evidence of the clinical importance of remaining ambiguities. For individual samples, genotyping results with three-field resolution can be generated using long-read sequencing technologies (48). Moreover, our amplicon strategy does not include the 5′ and 3′ UTRs of *MICA/B* which contain additional polymorphic positions (49, 50). Some of them influence (s)MICA/B expression which varies between different alleles (18, 51–53). However, to the best of our knowledge, there are no studies, which address the effects of donor *MICA/B* variations outside the exons in hematopoietic stem cell transplantation.

Although we genotyped over two million samples, we have not encountered some of the *MICA/B* alleles described in the IPD/IMGT-HLA database (**Table 2**). This may be due to several reasons. First of all, the majority of our samples are of European origin. Therefore, we might lack rare alleles occurring predominantly in other ethnicities. One example is *MICB*032* which was originally isolated from an Uyghur individual (54). In other cases, initial submissions to IPD/IMGT-HLA could be erroneous. This might especially be true for the alleles that have never been independently confirmed. For example, all heterozygous positions defining *MICA*005* or *MICA*013* also occur in one of the two most common alleles *MICA*008* and *MICA*002*. If those positions were not correctly phased during Sanger sequence analysis, *MICA*005* and *MICA*013* could have been erroneously reported. However, the sequencing of cloned PCR fragments should have prevented such errors (1, 55). Other not observed alleles, like *MICA*081*, *MICB*011*, *MICB*016,* or *MICB*022*, differ from more common alleles in only one position (56, 57). While this may reflect sequencing errors, it is more likely that the more recent submissions represent very low frequency observations as we discover on a daily basis. However, for the individual allele this may only be resolved by resequencing the original DNA which is often no longer available.

In conclusion, our workflow demonstrates that upfront *MICA/B* genotyping for potential stem cell donors can be performed with only minor increases in expenses and workload.

So far, *MICA/B* informed donor selection has not yet found widespread application in clinical practice. Clearly, additional confirmatory studies would be worthwhile. However, the availability of genotyping information remains a major hurdle for the translation of new markers into clinical practice. With the *MICA/B* genotyping of millions of donors we provide that data to facilitate *MICA/B* informed donor selection.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KP developed and tested the primer set. JS, DS, JP, and JH developed and implemented the genotyping algorithm. DF, RC, and SB performed genotyping for reference samples. AK, CM, GS, and JS analyzed frequency data. AK prepared figures and tables. AK and VL wrote the manuscript. AK, AS, and VL conceived and supervised the work. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.00314/full#supplementary-material

REFERENCES

1. Bahram S, Bresnahan M, Geraghty DE, Spies T. A second lineage of mammalian major histocompatibility complex class I genes. *Proc Natl Acad Sci USA.* (1994) 91:6259–63. doi: 10.1073/pnas.91.14.6259

2. Risti M, Bicalho MD. MICA and NKG2D: is there an impact on kidney transplant outcome? *Front Immunol.* (2017) 8:179. doi: 10.3389/fimmu.2017.00179

3. Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, et al. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science.* (1999) 285:727–9. doi: 10.1126/science.285.5428.727

4. Glienke J, Sobanov Y, Brostjan C, Steffens C, Nguyen C, Lehrach H, et al. The genomic organization of NKG2C, E, F, and D receptor genes in the human natural killer gene complex. *Immunogenetics.* (1998) 48:163–73. doi: 10.1007/s002510050420

5. Li P, Morris DL, Willcox BE, Steinle A, Spies T, Strong RK. Complex structure of the activating immunoreceptor NKG2D and its MHC class I–like ligand MICA. *Nat Immunol*. (2001) 2:443–51. doi: 10.1038/87757

6. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. *Nucl Acids Res*. (2015) 43:D423-31.

7. Petersdorf EW, Shuler KB, Longton GM, Spies T, Hansen JA. Population study of allelic diversity in the human MHC class I-related MIC-A gene. *Immunogenetics*. (1999) 49:605–12. doi: 10.1007/s002510050655

8. Gao X, Single RM, Karacki P, Marti D, O'Brien SJ, Carrington M. Diversity of MICA and linkage disequilibrium with HLA-B in two North American populations. *Hum Immunol*. (2006) 67:152–8. doi: 10.1016/j.humimm.2006.02.009

9. Ahmad T, Marshall SE, Mulcahy-Hawes K, Orchard T, Crawshaw J, Armuzzi A, et al. High resolution MIC genotyping: design and application to the investigation of inflammatory bowel disease susceptibility. *Tissue Antigens*. (2002) 60:164–79. doi: 10.1034/j.1399-0039.2002.600207.x

10. Zhou F, Cao H, Zuo X, Zhang T, Zhang X, Liu X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet*. (2016) 48:740–6. doi: 10.1038/ng.3576

11. Wang W, Tian W, Zhu F, Li L, Cai J, Wang F, et al. MICA Gene Deletion in 3411 DNA Samples from Five Distinct Populations in Mainland China and Lack of Association with Nasopharyngeal Carcinoma (NPC) in a Southern Chinese Han population. *Ann Hum Genet*. (2016) 80:319–26. doi: 10.1111/ahg.12175

12. Ying Y, He Y, Tao S, Han Z, Wang W, Chen N, et al. Distribution of MICB diversity in the Zhejiang Han population: PCR sequence-based typing for exons 2–6 and identification of five novel MICB alleles. *Immunogenetics*. (2013) 65:485–92. doi: 10.1007/s00251-013-0699-4

13. Cha CH, Sohn YH, Oh HB, Ko SY, Cho MC, Kwon OJ. MICB polymorphisms and haplotypes with MICA and HLA alleles in Koreans. *Tissue Antigens*. (2011) 78:38–44. doi: 10.1111/j.1399-0039.2011.01694.x

14. Frigoul A, Lefranc M-P. MICA: standardized IMGT allele nomenclature, polymorphisms and diseases. In: Pandalai SG, editor. *Recent Research Developments in Human Genetics*, Vol. 3. Trivandrum: Research Signpost (2005). p. 95–145.

15. Ashiru O, Boutet P, Fernández-Messina L, Agüera-González S, Skepper JN, Valés-Gómez M, et al. Natural killer cell cytotoxicity is suppressed by exposure to the human NKG2D ligand MICA*008 that is shed by tumor cells in exosomes. *Cancer Res*. (2010) 70:481–9. doi: 10.1158/0008-5472.CAN-09-1688

16. Ashiru O, López-Cobo S, Fernández-Messina L, Pontes-Quero S, Pandolfi R, Reyburn HT, et al. A GPI anchor explains the unique biological features of the common NKG2D-ligand allele MICA*008. *Biochem J*. (2013) 454:295–302. doi: 10.1042/BJ20130194

17. Nückel H, Switala M, Sellmann L, Horn PA, Dürig J, Dührsen U, et al. The prognostic significance of soluble NKG2D ligands in B-cell chronic lymphocytic leukemia. *Leukemia*. (2010) 24:1152–9. doi: 10.1038/leu.2010.74

18. Schmiedel D, Mandelboim O. NKG2D ligands-critical targets for cancer immune escape and therapy. *Front Immunol*. (2018) 9:2040. doi: 10.3389/fimmu.2018.02040

19. Duan S, Guo W, Xu Z, He Y, Liang C, Mo Y, et al. Natural killer group 2D receptor and its ligands in cancer immune escape. *Mol Cancer*. (2019) 18:29. doi: 10.1186/s12943-019-0956-8

20. de Andrade LF, Tay RE, Pan D, Luoma AM, Ito Y, Badrinath S, et al. Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell–driven tumor immunity. *Science*. (2018) 359:1537–42. doi: 10.1126/science.aao0505

21. Steinle A, Li P, Morris DL, Groh V, Lanier LL, Strong RK, et al. Interactions of human NKG2D with its ligands MICA, MICB, and homologs of the mouse RAE-1 protein family. *Immunogenetics*. (2001) 53:279–87. doi: 10.1007/s002510100325

22. Zuo J, Mohammed F, Moss P. The Biological influence and clinical relevance of polymorphism within the NKG2D ligands. *Front Immunol*. (2018) 9:1820. doi: 10.3389/fimmu.2018.01820

23. Pollock RA, Chandran V, Pellett FJ, Thavaneswaran A, Eder L, Barrett J, et al. The functional MICA-129 polymorphism is associated with skin but not joint manifestations of psoriatic disease independently of HLA-B and HLA-C. *Tissue Antigens*. (2013) 82:43–7. doi: 10.1111/tan.12126

24. Tong HV, Toan NL, Song LH, Bock CT, Kremsner PG, Velavan TP. Hepatitis B virus-induced hepatocellular carcinoma: functional roles of MICA variants. *J Viral Hepat*. (2013) 20:687–98. doi: 10.1111/jvh.12089

25. Isernhagen A, Malzahn D, Bickeböller H, Dressel R. Impact of the MICA-129Met/Val dimorphism on NKG2D-mediated biological functions and disease risks. *Front Immunol*. (2016) 7:588. doi: 10.3389/fimmu.2016.00588

26. Chen E, Chen C, Chen F, Yu P, Lin L. Positive association between MIC gene polymorphism and tuberculosis in Chinese population. *Immunol Lett*. (2019) 213:62–9. doi: 10.1016/j.imlet.2019.07.008

27. Carapito R, Gottenberg JE, Kotova I, Untrau M, Michel S, Naegely L, et al. A new MHC-linked susceptibility locus for primary Sjögren's syndrome: MICA. *Hum Mol Genet*. (2017) 26:2565–76. doi: 10.1093/hmg/ddx135

28. Isernhagen A, Malzahn D, Viktorova E, Elsner L, Monecke S, von Bonin F, et al. The MICA-129 dimorphism affects NKG2D signaling and outcome of hematopoietic stem cell transplantation. *EMBO Mol Med*. (2015) 7:1480–502. doi: 10.15252/emmm.201505246

29. Parmar S, Del Lima M, Zou Y, Patah PA, Liu P, Cano P, et al. Donor-recipient mismatches in MHC class I chain-related gene A in unrelated donor transplantation lead to increased incidence of acute graft-versus-host disease. *Blood*. (2009) 114:2884–7. doi: 10.1182/blood-2009-05-223172

30. Fuerst D, Neuchel C, Niederwieser D, Bunjes D, Gramatzki M, Wagner E, et al. Matching for the MICA-129 polymorphism is beneficial in unrelated hematopoietic stem cell transplantation. *Blood*. (2016) 128:3169–76. doi: 10.1182/blood-2016-05-716357

31. Carapito R, Jung N, Kwemou M, Untrau M, Michel S, Pichot A, et al. Matching for the nonconventional MHC-I MICA gene significantly reduces the incidence of acute and chronic GVHD. *Blood*. (2016) 128:1979–86. doi: 10.1182/blood-2016-05-719070

32. Carapito R, Aouadi I, Ilias W, Bahram S. Natural Killer Group 2, Member D/NKG2D ligands in hematopoietic cell transplantation. *Front Immunol*. (2017) 8:368. doi: 10.3389/fimmu.2017.00368

33. Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics*. (2014) 15:63. doi: 10.1186/1471-2164-15-63

34. Schöfl G, Lang K, Quenzel P, Böhme I, Sauter J, Hofmann JA, et al. 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. *BMC Genomics*. (2017) 18:161. doi: 10.1186/s12864-017-3575-z

35. Lang K, Wagner I, Schöne B, Schöfl G, Birkner K, Hofmann JA, et al. ABO allele-level frequency estimation based on population-scale genotyping by next generation sequencing. *BMC Genomics*. (2016) 17:374. doi: 10.1186/s12864-016-2687-1

36. Wagner I, Schefzyk D, Pruschke J, Schöfl G, Schöne B, Gruber N, et al. Allele-Level KIR genotyping of more than a million samples: workflow, algorithm, and observations. *Front Immunol*. (2018) 9:2843. doi: 10.3389/fimmu.2018.02843

37. Solloch UV, Lang K, Lange V, Böhme I, Schmidt AH, Sauter J. Frequencies of gene variant CCR5-Δ32 in 87 countries based on next-generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. *Hum Immunol*. (2017) 78:710–7. doi: 10.1016/j.humimm.2017.10.001

38. Milius RP, Mack SJ, Hollenbach JA, Pollack J, Heuer ML, Gragert L, et al. Genotype List String: a grammar for describing HLA and KIR genotyping results in a text string. *Tissue Antigens*. (2013) 82:106–12. doi: 10.1111/tan.12150

39. Marsh SGE, Parham P, Dupont B, Geraghty DE, Trowsdale J, Middleton D, et al. Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Hum Immunol*. (2003) 64:648–54. doi: 10.1016/S0198-8859(03)00067-3

40. Carapito R, Jung N, Untrau M, Michel S, Pichot A, Giacometti G, et al. Matching of MHC Class I chain-related genes a and B is associated with reduced incidence of severe acute Graft-Versus-Host disease after unrelated hematopoietic stem cell transplantation. *Blood*. (2014) 124:664. doi: 10.1182/blood.V124.21.664.664

41. Tian W, Cai JH, Wang F, Li LX. MICA polymorphism in a northern Chinese Han population: the identification of a new MICA allele, MICA*059. *Hum Immunol*. (2010) 71:423–7. doi: 10.1016/j.humimm.2010.01.025

42. Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, et al. Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst*. (2013) 105:624–33. doi: 10.1093/jnci/djt051

43. Jiang X, Zou Y, Huo Z, Yu P. Association of major histocompatibility complex class I chain-related gene A microsatellite polymorphism and hepatocellular carcinoma in South China Han population. *Tissue Antigens*. (2011) 78:143–7. doi: 10.1111/j.1399-0039.2011.01693.x

44. Onyeaghala G, Lane J, Pankratz N, Nelson HH, Thyagarajan B, Walcheck B, et al. Association between MICA polymorphisms, s-MICA levels, and pancreatic cancer risk in a population-based case-control study. *PLoS ONE*. (2019) 14:e0217868. doi: 10.1371/journal.pone.0217868

45. Lombana TN, Matsumoto ML, Berkley AM, Toy E, Cook R, Gan Y, et al. High-resolution glycosylation site-engineering method identifies MICA epitope critical for shedding inhibition activity of anti-MICA antibodies. *MAbs*. (2019) 11:75–93. doi: 10.1080/19420862.2018.1532767

46. HLA Nomenclature @ hla.alleles.org [Internet]. Available online at: http://hla.alleles.org/alleles/g_groups.html (accessed November 29, 2019).

47. HLA Nomenclature @ hla.alleles.org [Internet]. Available online at: http://hla.alleles.org/alleles/p_groups.html (accessed November 29, 2019).

48. Albrecht V, Zweiniger C, Surendranath V, Lang K, Schöfl G, Dahl A, et al. Dual redundant sequencing strategy: full-length gene characterisation of 1056 novel and confirmatory HLA alleles. *HLA*. (2017) 90:79–87. doi: 10.1111/tan.13057

49. Cox ST, Madrigal JA, Saudemont A. Diversity and characterization of polymorphic 5′ promoter haplotypes of MICA and MICB genes. *Tissue Antigens*. (2014) 84:293–303. doi: 10.1111/tan.12400

50. Cox ST, Hernandez D, Danby R, Turner TR, Madrigal JA. Diversity and characterisation of polymorphic 3′ untranslated region haplotypes of MICA and MICB genes. *HLA*. (2018) 92:392–402. doi: 10.1111/tan.13434

51. Rodríguez-Rodero S, González S, Rodrigo L, Fernández-Morera JL, Martínez-Borra J, López-Vázquez A, et al. Transcriptional regulation of MICA and MICB: a novel polymorphism in MICB promoter alters transcriptional regulation by Sp1. *Eur J Immunol*. (2007) 37:1938–53. doi: 10.1002/eji.200737031

52. Lo PHY, Urabe Y, Kumar V, Tanikawa C, Koike K, Kato N, et al. Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk. *PLoS ONE*. (2013) 8:e61279. doi: 10.1371/journal.pone.0061279

53. Shi C, Li H, Couturier JP, Yang K, Guo X, He D, et al. Allele specific expression of MICA variants in human fibroblasts suggests a pathogenic mechanism. *Open Rheumatol J*. (2015) 9:60–4. doi: 10.2174/1874312901409010060

54. Alleles Report < IMGT/HLA < IPD < EMBL-EBI [Internet]. [cited 2019 Dec 17]. Available online at: https://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/get_allele.cgi?MICB^\ast032 (accessed November 29, 2019).

55. Fodil N, Laloux L, Wanner V, Pellet P, Hauptmann G, Mizuki N, et al. Allelic repertoire of the humanMHC class IMICA gene. *Immunogenetics*. (1996) 44:351–7. doi: 10.1007/BF02602779

56. Schroeder M, Elsner HA, Kim TD, Blasczyk R. Eight novel MICB alleles, including a null allele, identified in gastric MALT lymphoma patients. *Tissue Antigens*. (2004) 64:276–80. doi: 10.1111/j.1399-0039.2004.00286.x

57. Visser CJT, Tilanus MGJ, Schaeffer V, Tatari Z, Tamouza R, Janin A, et al. Sequencing-based typing reveals six novel MHC class I chain-related gene B (MICB) alleles. *Tissue Antigens*. (1998) 51:649–52. doi: 10.1111/j.1399-0039.1998.tb03008.x