Characterization of Over 400 Novel MICA Alleles and Their Impact on MICA Allele Frequencies in the German Population

Viviane Albrecht[1,2], Christin Paech[1], Jürgen Sauter[2], Vinzenz Lange[1], Anja Klussmeier[1]

[1] DKMS Life Science Lab, Dresden, Germany
[2] DKMS Group, Tübingen, Germany

Introduction

In 2020, we published MICA allele frequencies from a cohort of over one million German individuals (1). This study identified MICA*008 (42%), MICA*002 (12%), and MICA*009 (9%) as the most common MICA alleles at protein resolution. Furthermore, we discovered novel alleles with a cumulative frequency of 0.3%. To decrease this fraction of unknown sequences, we aimed to fully characterize the most frequent novel alleles using both long- and short-read sequencing.

Novel Alleles

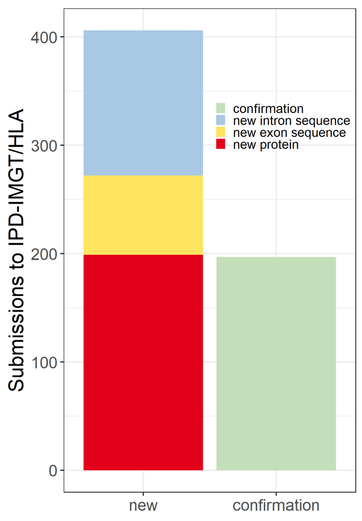
Sample Selection

Novel MICA alleles were identified by our high-throughput genotyping workflow for potential stem cell donors. This workflow sequences exons 2 and 3 and larger parts of exons 4 and 5 of MICA (1). For novel allele characterization, we selected samples that harbored the most frequent variations, e.g., some of the most frequent novel sequence at that time were identified more than 1000 times by our workflow. Other variations, e.g., intronic variations, were not targeted, but were characterized if identified by chance.

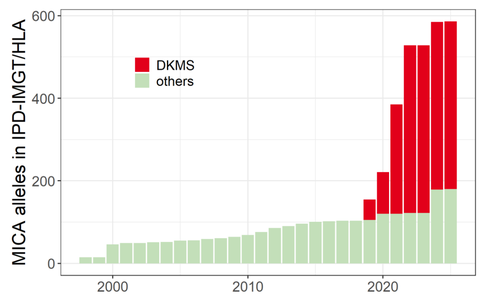
PCR & Sequencing

Buccal swab-isolated DNA of selected samples was amplified in two separate PCR reactions with primers (forward: CTGCTTGAGCCGCTGAGAGG, reverse: GATCCAGGCAGGGAATTGAATCCC and GAGATCCAGGCAGGGAATTCAATTCC) that amplify MICA in full-length (12 kb) using previously described protocols (2). Amplicons from one PCR reaction were sequenced with nanopores (Oxford Nanopore Technologies, Oxford, UK), the other was sequenced on a MiSeq instrument (Illumina, San Diego, USA). Sequencing reads were analyzed using both NGSengine (GenDx, Utrecht, The Netherlands) and DR2S (3). Finally, consensus sequences of novel alleles were submitted to IPD-IMGT/HLA using TypeLoader2 (4). If possible, sequences of two different samples with the same novel allele were submitted for confirmation.

Submissions to IPD-IMGT/HLA



In total, 603 sequences were submitted to IPD-IMGT/HLA. Among them were 406 distinct novel alleles and 197 sequence confirmations. 199 of the novel alleles code for novel MICA proteins, 73 carry synonymous exon variations and the remaining 134 novel alleles harbor only intron variations.



These 406 distinct novel allele submissions largely contribute to the total amount of described MICA alleles today. The current IPD-IMGT/HLA Database release (3.60) contains 586 MICA alleles, among them 278 distinct MICA proteins and 9 null alleles (5).

Conclusion

We successfully characterized and submitted 406 distinct novel MICA alleles, among them 199 novel MICA proteins, to the IPD-IMGT/HLA Database. As a result, we could reduce the number of reported novel MICA alleles during genotyping substantially from 0.3% to 0.03% (German population). Furthermore, MICA allele frequency analyses for the German and South African Black population revealed major differences, including alleles that were almost exclusively found in the South African Black population.

MICA Allele Frequencies

German Population

We analyzed MICA allele frequencies in a cohort of Germans (n=48,618), which were genotyped from 2022 to 2023 by our high-throughput workflow (1).

In general, the frequencies from this novel, independent cohort confirm the results from our first cohort (1) (Table 1). As expected, after novel allele submission, the cumulative frequency of novel alleles decreased substantially from 0.3% to 0.03% in this new cohort.

The most frequent sequences that were newly submitted to IPD-IMGT/HLA were MICA*107N (0.02%), MICA*141 (0.01%), and MICA*089 (0.01%) (Table 1).

Out of the 198 novel MICA proteins that were characterized and submitted to IPD-IMGT/HLA, we could re-identify 70 in the novel cohort of 48,618 Germans (Table 1, marked in yellow). Another 49 were re-identified in samples from other populations (total cohort size: 93,814 samples). The remaining 79 novel MICA proteins were not re-identified and are presumably rare.

allele	ambiguity
MICA*002#	MICA*002/110
MICA*009#	MICA*009/049
MICA*010#	MICA*010/065/069
MICA*027#	MICA*027/048
MICA*047#	MICA*047/101

Table 2: Hash symbols (#) mark alleles that contain ambiguities that cannot be resolved in our high-throughput workflow due to variations in non-sequenced exons.

MICA allele	frequency DE		MICA allele	frequency DE
MICA*008	0.43643		MICA*109	0.00003
MICA*002#	0.11496		MICA*077	0.00003
MICA*009#	0.08733		MICA*210	0.00003
MICA*010#	0.07777		MICA*114	0.00002
MICA*004	0.06501		MICA*236	0.00002
MICA*007	0.04845		MICA*185	0.00002
MICA*018	0.03602		MICA*131	0.00002
MICA*017	0.03321		MICA*224	0.00002
MICA*012	0.02145		MICA*130	0.00002
MICA*016	0.01886		MICA*222N	0.00002
MICA*011	0.01846		MICA*073	0.00002
MICA*027#	0.01577		MICA*137	0.00002
MICA*019	0.00838		MICA*203	0.00002
MICA*001	0.00778		MICA*053	0.00002
MICA*006	0.00425		MICA*090	0.00002
MICA*015	0.00079		MICA*171Q	0.00002
MICA*029	0.00069		MICA*231	0.00002
MICA*068	0.00054		MICA*229	0.00002
MICA*047#	0.00031		MICA*162	0.00002
MICA*072	0.00031		MICA*095	0.00002
MICA*045	0.00028		MICA*252	0.00002
NEW	0.00026		MICA*098	0.00002
MICA*070	0.00026		MICA*180	0.00001
MICA*052	0.00018		MICA*271	0.00001
MICA*030	0.00018		MICA*051	0.00001
MICA*107N	0.00016		MICA*186	0.00001
MICA*141	0.00014		MICA*291	0.00001
MICA*089	0.00012		MICA*067	0.00001
MICA*119	0.00010		MICA*113	0.00001
MICA*136	0.00009		MICA*189	0.00001
MICA*247	0.00007		MICA*148	0.00001
MICA*075	0.00006		MICA*178	0.00001
MICA*041	0.00005		MICA*214	0.00001
MICA*168	0.00004		MICA*195N	0.00001
MICA*267	0.00004		MICA*177	0.00001
MICA*074	0.00004		MICA*087	0.00001
MICA*167	0.00004		MICA*125	0.00001
MICA*146	0.00004		MICA*198	0.00001
MICA*181	0.00004		MICA*024	0.00001
MICA*020	0.00003		MICA*082	0.00001
MICA*208	0.00003		MICA*153	0.00001
MICA*257	0.00003		MICA*190	0.00001
MICA*232	0.00003		MICA*155	0.00001
MICA*165	0.00003		MICA*055	0.00001
MICA*116	0.00003		MICA*207	0.00001
MICA*160	0.00003		MICA*043	0.00001
MICA*135	0.00003		MIC*A243	0.00001
MICA*245	0.00003		MIC*A049	0.00001

Table 1: MICA allele frequencies in German population. Alleles marked in yellow were newly characterized and submitted to IPD-IMGT/HLA.

South African Black Population

MICA allele	frequency ZA	frequency DE / frequency ZA
MICA*008	0.2769	1.58
MICA*004	0.2447	0.26
MICA*002#	0.2248	0.51
MICA*019	0.0543	0.15
MICA*015	0.0463	0.02
MICA*009#	0.0454	1.93
MICA*018	0.0287	1.26
MICA*012	0.0184	1.17
MICA*011	0.0161	1.14
MICA*001	0.0135	0.58
MICA*068	0.0091	0.06
MICA*030	0.0076	0.02
MICA*041	0.0038	0.01
MICA*007	0.0022	21.95
MICA*016	0.0020	9.62
NEW	0.0015	0.17
MICA*258	0.0015	not in DE
MICA*046	0.0007	not in DE
MICA*045	0.0005	0.56
MICA*010#	0.0005	158.58
MICA*027#	0.0004	42.87
MICA*044	0.0003	not in DE
MICA*029	0.0003	2.80
MICA*017	0.0003	135.46
MICA*043	0.0003	0.04
MICA*151	0.0001	not in DE

Table 3: MICA allele frequencies in the South African Black population and comparison to the respective frequencies in the German population. Alleles marked in yellow were newly characterized and submitted to IPD-IMGT/HLA. Some alleles were not identified in the German population ('not in DE').

Some of our submitted novel alleles were exclusively found in non-European populations. As an example, we calculated allele frequencies for the South African Black population (n=4,085) and identified major differences to the German population.

Even though MICA*008 is the most frequent MICA allele in South African Blacks as well, its frequency is roughly half of the German frequency. In contrast, MICA*004 is much more frequent in the South African Blacks (24% versus 7% in Germans), as are MICA*019 (5% versus 0.8%) and MICA*015 (5% versus 0.08%). MICA*010 and MICA*017, which are common alleles in Germany (8% and 3%), are only rarely identified in South African Blacks (0.05% and 0.03%).

The submitted novel allele MICA*258 was identified in 26 samples from South Africa (all self-assessed as 'Black' or 'Colored'), but not in any other population (n=93,814). Similarly, the novel exon variation MICA*008:28 was almost restricted to the South African Black population (n=45, one individual from India) (data not shown in tables).

References

1. Klussmeier A, Massalski C, Putke K, Schäfer G, Sauter J, Schefzyk D, Pruschke J, Hofmann J, Fürst D, Carapito R, Bahram S, Schmidt AH and Lange V (2020) High-Throughput *MICA/B* Genotyping of Over Two Million Samples: Workflow and Allele Frequencies. *Front.Immunol.* 11:314. doi: 10.3389/fimmu.2020.00314
2. Putke, K. et al. (2024). Full-Length Characterization of Novel HLA-DRB1 Alleles for Reference Database Submission. In: Boegel, S. (eds) HLA Typing. Methods in Molecular Biology, vol 2809. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-3874-3_10
3. Klasberg, S., Schmidt, A.H., Lange, V. et al. DR2S: an integrated algorithm providing reference-grade haplotype sequences from heterozygous samples. BMC Bioinformatics 22, 236 (2021). https://doi.org/10.1186/s12859-021-04153-0
4. Schöne, B., Fuhrmann, M., Surendranath, V., Schmidt, A.H., Lange, V., Schöfl, G. (2024). Submitting Novel Full-Length HLA, MIC, and KIR Alleles with TypeLoader2. In: Boegel, S. (eds) HLA Typing. Methods in Molecular Biology, vol 2809. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-3874-3_11
5. Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, Marsh SGE. The IPD-IMGT/HLA Database. Nucleic Acids Research (2023) 51:D1053-60

DKMS
WE DELETE BLOOD CANCER

DKMS Life Science Lab
St. Petersburger Str. 2
01069 Dresden. Germany
www.dkms-lab.de