



## Hap-E Search as search engine for WMDA Search & Match

Christine Urban<sup>1</sup>, Denis Flaig<sup>1</sup>, Mark Melchers<sup>2</sup>, Alexander H. Schmidt<sup>1</sup>, Jan A. Hofmann<sup>1</sup>

<sup>1</sup> DKMS Group, Kressbach 1, 72072 Tübingen, Germany

<sup>2</sup> World Marrow Donor Association, Schipholweg 57, 1<sup>st</sup> floor, 2316 ZL Leiden, The Netherlands

### Introduction

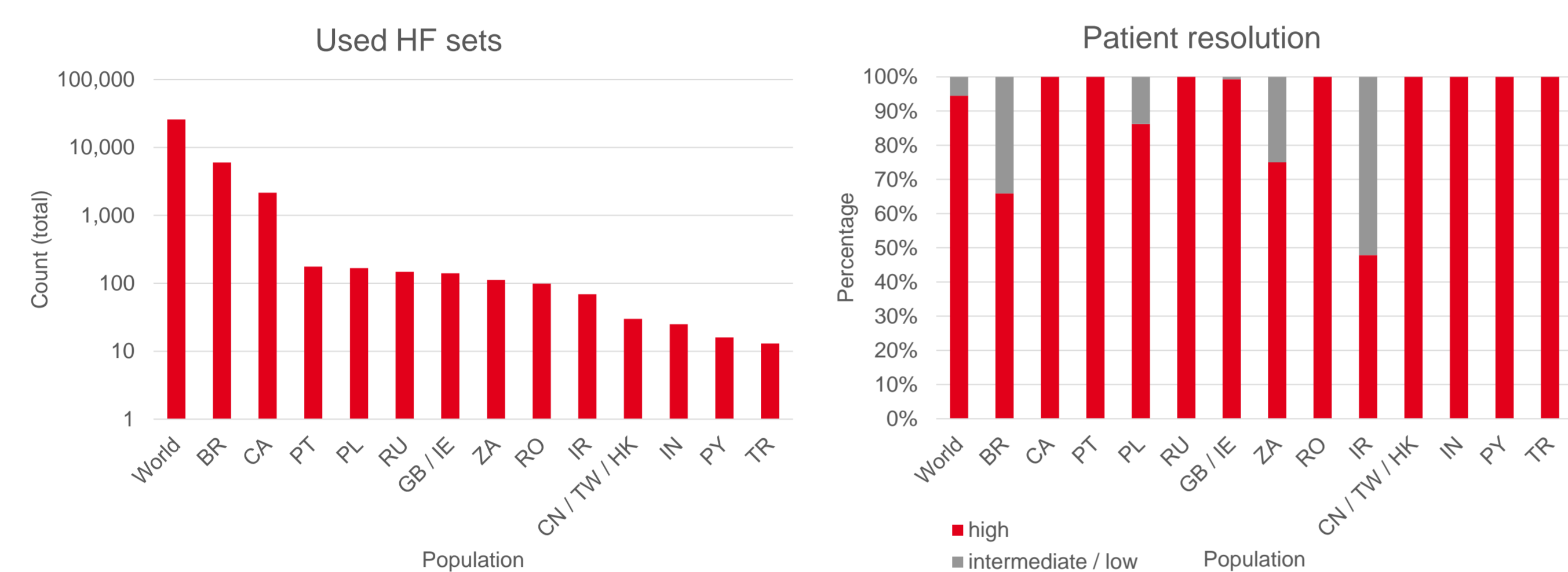
After running for more than four years as internal search engine for the DKMS Registry, Hap-E Search has been embedded in the WMDA Search & Match environment. It was made available in a pilot phase to the community in May 2022 and became fully operational on July 5<sup>th</sup>, 2022.

Since then, more than 47,000 searches have been processed with Hap-E Search on the WMDA Search & Match productive system. By providing Hap-E search, a state-of-the-art search algorithm, in the WMDA setup, we contribute to the community-driven approach of the WMDA initiative of Search, Match & Connect.

### Features

- Use of population-specific HLA haplotype frequencies.
- Search categories based on:
  - HLA-A, -B, -C, -DRB1, -DQB1 or HLA-A, -B, -C, -DRB1 or HLA-A, -B, -DRB1
  - HLA-A, -B low, HLA-DRB1 high for CB
- Support of up to two mismatches for all categories, and up to four mismatches for high-resolution based CB searches.
- Daily search result updates are provided.
- The algorithm is located on Oracle cloud server infrastructure (OCI).

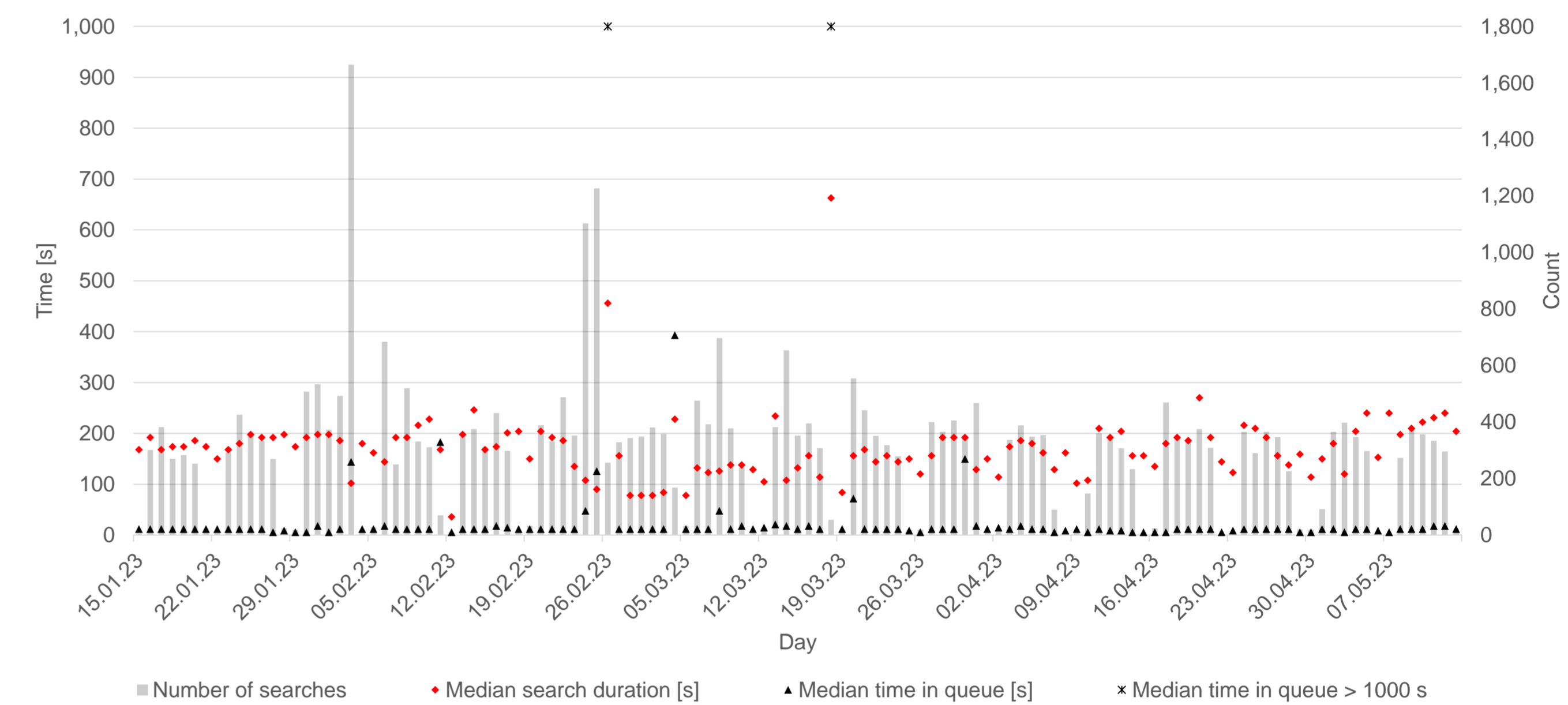
### Used HF sets



BR	Brazil	GB / IE	Great Britain / Ireland	CN / TW / HK	China / Taiwan / Hongkong
CA	Canada	ZA	South Africa	IN	India
PT	Portugal	RO	Romania	PY	Paraguay
PL	Poland	IR	Iran	TR	Turkey
RU	Russia				

- Haplotype frequency sets for patients used in at least 10 searches are shown.
- Searches without high resolution patient typing benefit from the use of multiple HF sets that allow for better matching predictions.
- A large number of searches (1,433) without high resolution typing still uses the unspecific "world" HF set. These searches could benefit from more specific HF sets.

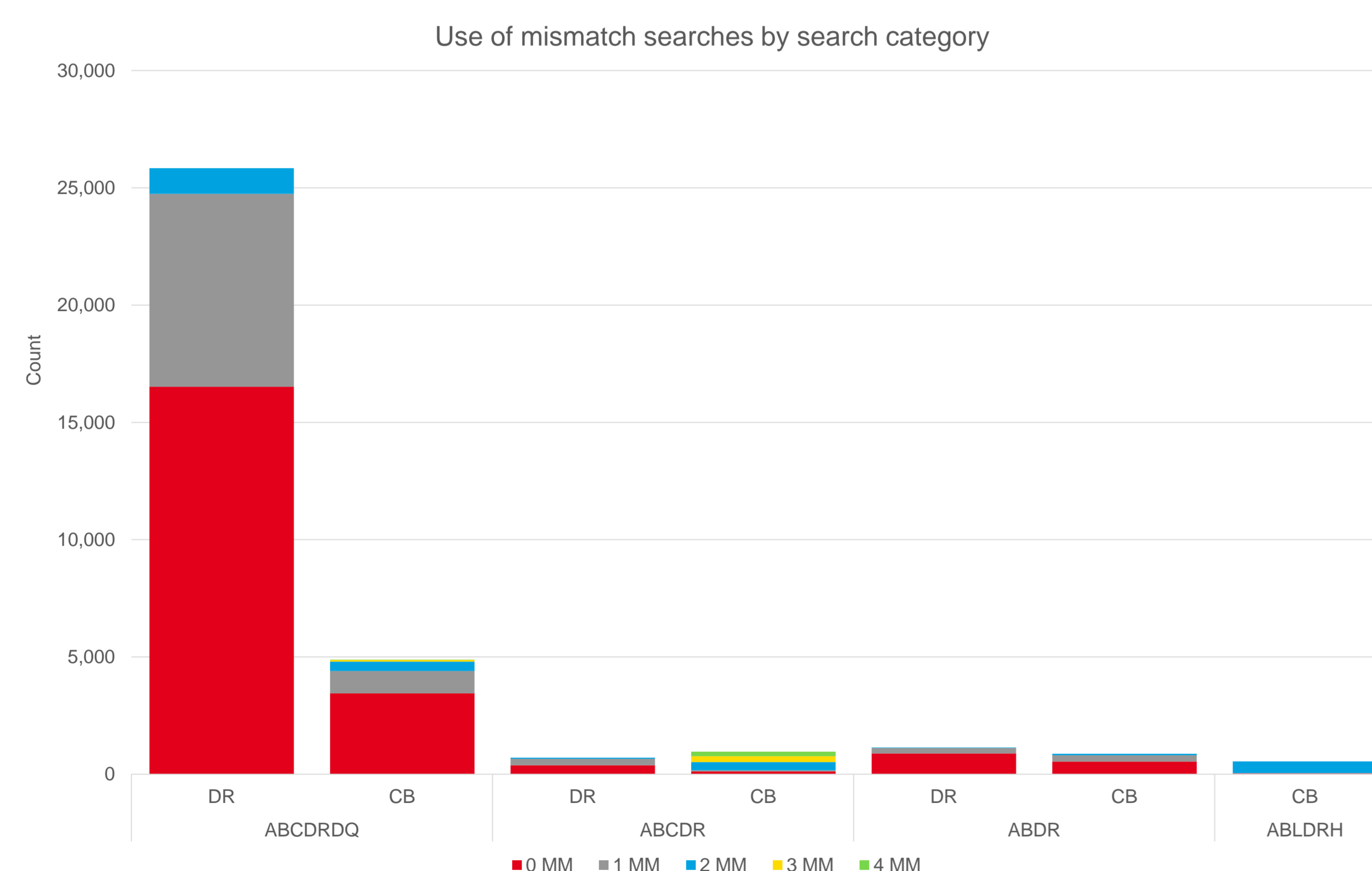
### Results – number of searches and duration



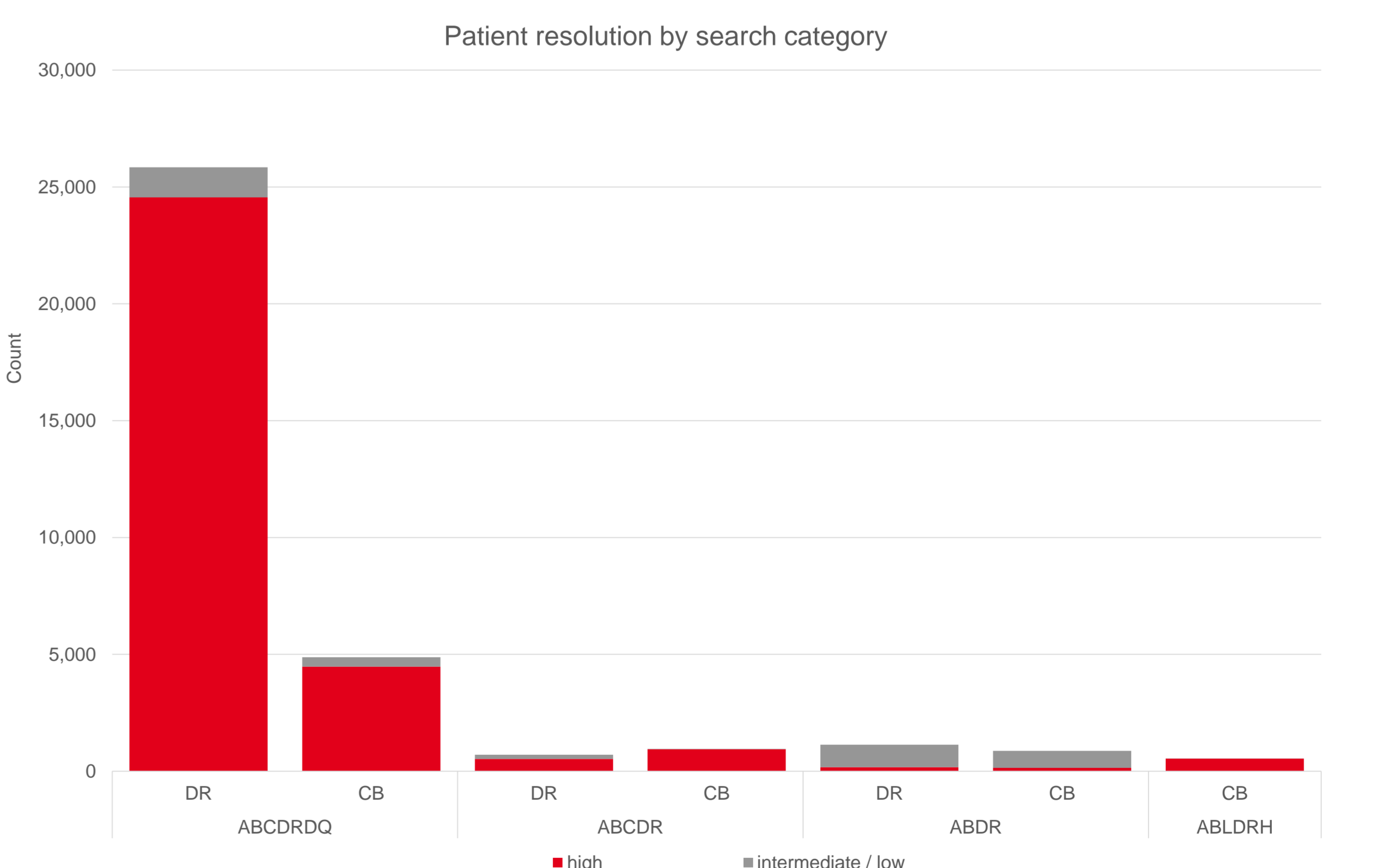
- The median search duration for different mismatch categories is:
 

Number of mismatches	0	1	2	3	4	overall
Median search duration [s]	138	216	300	294	618	168
- The two runaway values with median time in queue above 1000 s (26.02.23 and 18.03.23) are due to application downtime on those weekends.
- Number of searches shows a stable weekly pattern.
- Performance of the algorithm is independent of the number of searches.

### Results – used search categories



- Most searches with multiple mismatches are run for ABCDR CB searches.
- Distribution reflects different search strategies and API usage.



- Dominating part of searches is performed for patients with high resolution HLA typing.
- Major exception: in the ABDR-category, most patients have low resolution typing.
- If the patient has no typing results for DQB1 and / or C, only ABCDR / ABDR searches are allowed, respectively.
- Probably, low resolution ABDR searches are performed to get a first glance for patients with result from family typing, before high resolution typing results are obtained.

### Conclusion

- Since its launch in July 2022, the Hap-E Search algorithm has provided stable and consistent search durations for WMDA Search & Match.
- Queuing time of searches is minimized by scalability of resources.
- The major part of searches run at WMDA Search & Match are 0 mismatch ABCDRDQ searches for patients with five-locus high resolution typing.