**Title: Using Patient HLA Genotype Frequency Can Predict the Difficulty of an Unrelated Donor Search in the Be The Match Registry**

Kimberly Wadsworth, Ph.D.1, Mark Albrecht, MStat MBA1, Rachel Fonstad, ??2, Stephen Spellman, M.B.S.??2, Martin Maiers, B.A.1, Jason Dehn, M.P.H.1

1National Marrow Donor Program, Minneapolis, MN, USA

2Center for International Blood and Marrow Transplant Research, Minneapolis, MN, USA

**Corresponding Author:**

Kimberly Wadsworth, Ph.D.

National Marrow Donor Program

3001 Broadway St NE, Suite 100, Minneapolis, MN 55413

Phone: 612 627 5876; Fax: 612 362 3488; Email: kwadswor@nmdp.org

**Key words:**

Marrow and Stem Cell Transplantation, HLA Matching, Donor Registry, Match Likelihood, Search Prognosis

**Abstract**

Identifying the best possible Human Leukocyte Antigen (HLA) match for a searching patient (Pt) in need of a bone marrow transplant in a short amount of time is critical to transplant success. Having a simple scoring system that can provide a quick search prognosis at the onset of an unrelated donor (URD) search could be a useful tool for transplant physicians to understand the difficulty of identifying a fully matched or suitably mismatched URD. In this study, we aimed to determine if a Pt HLA genotype frequency (GF) could be used as a surrogate measure of whether or not the Pt can identify a potential 10/10 and/or 9/10 URD in the Be The Match Registry®. GF was assigned using the haplostats.org application on a training dataset of 1307 Pt with intermediate/high-resolution 5-locus typing (HLA-A, -B, -C, -DRB1, -DQB1) that had searched the Be The Match Registry using the transplantation center’s reporting of Pt race/ethnicity: White (WH), African-American/Black (AFA), Asian/Pacific Islander (API), Hispanic (HIS), or unknown (UNK). Using the National Marrow Donor Program’s Traxis® application, the number of 10/10 and 9/10 URD for each Pt with a HapLogicSM predicted allele match of >50% was recorded. GF ranges were established for each race group using a proportional odds model to correlate with defined search prognosis categories: “good” (> 2 10/10 and > 2 9/10), “fair” (< 3 10/10 and > 2 9/10), or “poor” (No 10/10 and < 3 9/10). To validate the precision of using GF to predict search prognosis, a second cohort (n= 3515) was used to calculate the overall concordance for each race group: WH: 83%, AFA: 77%, API: 74%, HIS: 70%, and UNK: 77%. We present here a categorization scheme that uses Pt GF as a predictive measure for the productivity of an URD search. Although not a replacement for an actual URD search, the GF may offer a quick way for transplant physicians to get an indication of the likely search outcome, engage HLA expertise earlier in a Pt search process, and guide early clinical consideration of non-fully matched URD alternative stem cell source options.

**Introduction**

A variety of malignant and non-malignant hematological diseases and inherited metabolic disorders can be cured using allogeneic hematopoietic stem cell transplantation. Only 30% of searching patients (Pt) can identify a Human Leukocyte Antigen (HLA) identical related donor. When a related match cannot be identified, one option for Pt is to find an unrelated HLA-matched adult donor (URD) from one of the many worldwide bone marrow registries [1].

HLA matching between the URD and the Pt is critical to increasing overall survival and reducing transplant-related mortality and graft vs host disease (GvHD) [2-5]. Currently, the ideal matching level considered by many transplanting centers is a 10/10 allele match between the Pt and URD matching at the following loci: HLA-A, -B, -C, -DRB1, and -DQB1, although matching at DQB1 has not shown to have a significant impact on clinical outcome [2, 6]. A suitable match is usually considered to be at least a 7/8 match (matching on HLA-A, -B, -C, and –DRB1). The impact of matching at other loci (e.g. DPB1, DRB3/4/5) has also been shown to have a positive impact on overall survival [7-9].When a full match adult URD cannot be found, other alternatives include using a mismatched URD, cord blood unit(s), or haploidentical related donor [10, 11]. The ability for a Pt to efficiently find a suitably matched URD can significantly impact the Pt search time and costs.

Multiple studies have been performed in the past that use individual haplotype commonalities to help predict the likelihood of finding a matched URD at the onset of the Pt search. Initial studies used individual HLA-A, -B, -DRB1 haplotype commonalities to help determine the probability of finding an HLA-matched URD [12]. Later studies correlated the probability of finding a URD with individual HLA-A, -B, -C, -DRB1, and DQB1 haplotype commonalities as well as the presence of rare alleles and other factors [13-15]. Most of these studies were performed on a predominately White-European population and weren’t always straightforward or easy to use.

We have developed a simple scoring system that uses a Pt genotype frequency (GF) to help predict whether a Pt is likely to have a potential 10/10 and/or 9/10 URD in the National Marrow Donor Program’s (NMDP) Be The Match Registry®, which is comprised of approximately half of the more than 24 million URD worldwide. This scoring system was devised for each broad race and ethnic (hereafter “race”) group separately– White (WH), African American/Black (AFA), Asian/Pacific Islander (API), and Hispanic (HIS), to account for the HLA diversity between groups. [16]. This scoring system could prove useful to transplant physicians at the beginning of a Pt search to understand the likelihood of whether the Pt will find a 10/10 and/or 9/10 matched URD or no such URD.

**Materials and Methods**

NMDP Institutional Review Board approval for human subject research was obtained for all patients.

**Patient GF Estimation:**

Pt GF information was obtained using the NMDP’s publicly available haplostats.org application, which provides analysis of HLA typing using NMDP Bioinformatics HLA haplotype frequencies by race [17, 18], on newly entered Be The Match Registry Pt from 01/07/2013 - 12/10/2013. Only Pt with intermediate/high- resolution HLA-A, B, C, DRB1, and DQB1 typing were included. The A~C~B~DRBX(i.e. DRB3/4/5)~DRB1~DQB1 haplotype frequency tables were used. GF was calculated using the product of a Pt projected most likely individual haplotype frequencies, within a Pt self-identified broad race group. In the cases where the Pt race was unknown (UNK), WH GF data was used and analyzed as a separate group. Last, if the genotype frequency was undefined (due to the inability to find a haplotype pair consistent with the patient’s HLA typing) that patient’s GF was defined as the minimum GF observed across all populations.

**Search Productivity Determination:**

Each search was run in the NMDP Traxis® application and the number of potential 9/10 and 10/10 URD with a HapLogic predicted allele match >50% was recorded. Patients were classified as having a “Good”, “Fair”, or “Poor” prognosis based upon the following: “Good” was defined as a search with at least 3 potential 10/10 URD with a HapLogic predicted match of >50% and any number of 9/10s; “Fair” was defined as a search that had 1-2 potential 10/10s and any number of 9/10 URD or a search with no 10/10s, but at least 3 potential 9/10 URD all with a HapLogic predicted match of >50%; and, “Poor” was defined as a search that had no potential 10/10s and less than 3 potential 9/10s with a HapLogic predicted match of >50%.

**Proportional Odds Classification Model and Validation:**

A proportional odds model (i.e. cumulative logit [X]) was fit to a random 50% training split (n=2410) of the data with Search Productivity (“Good”, “Fair”, “Poor”) as the response; model predictors were genotype frequency and race entered into the model as additive plus first order interaction terms - such a model specification results in separate classification models for each race group. Within each race group, observations were weighted by class to adjust for response proportion imbalance, thus each class had an equal misclassification cost. Model predictions were used to establish classification boundaries as a function of genotype frequency and race based upon the most probable class. These classification boundaries were then applied to a random 50% validation split (n=2411) and a confusion matrix was generated comparing predicted to actual search productivity class.

**Clinical Validation on URD Search Cohort:**

As an additional validation, the model was used to score (i.e. classify) an independent dataset of “pseudo-patients” used in a previous study [19] which estimated the URD match rate for each of the four broad race groups. Searches were performed on the pseudo-patients and potential URD were HLA typed as necessary to determine whether a potential 10/10 (or 8/8) URD from the Be The Match Registry could be identified. The number of patients having at least one 10/10 URD was then tallied by the model-predicted search prognosis category.

**Results**

Table 1 displays the patient race distribution of the training and validation datasets utilized in the study.

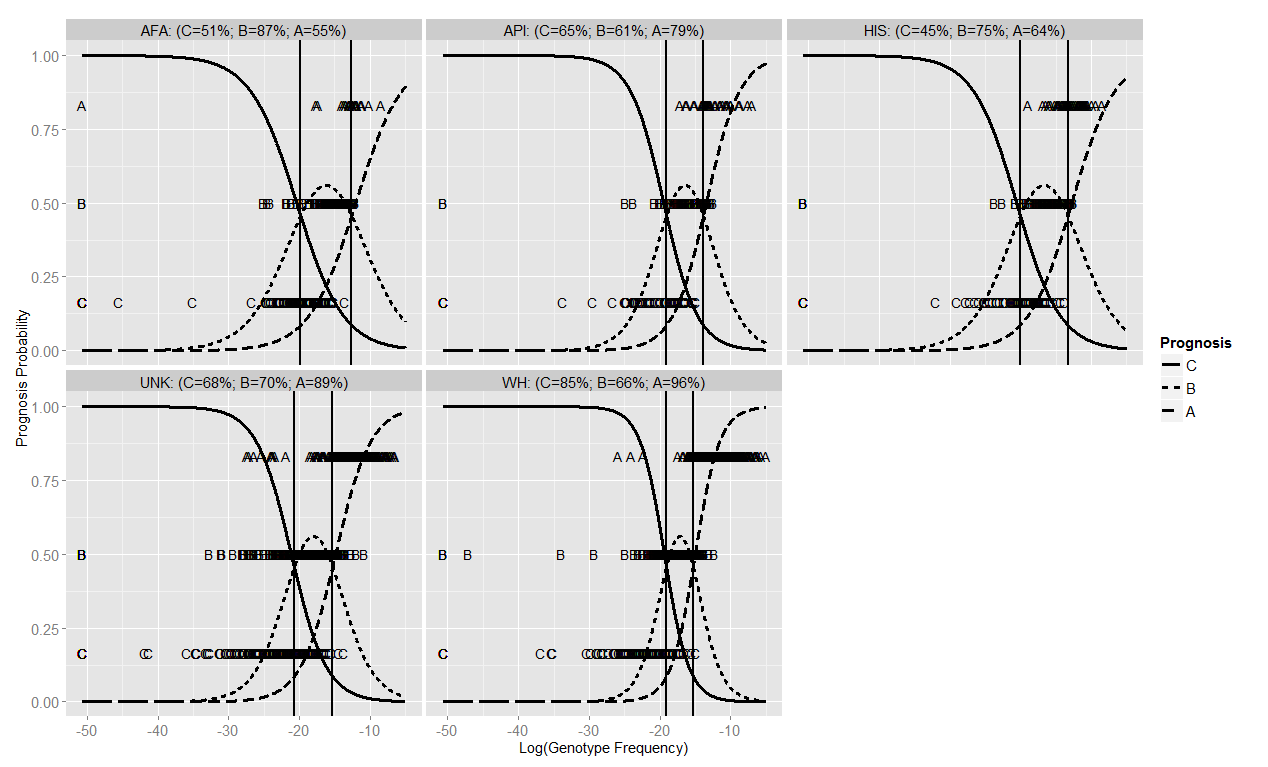
**Table 1:** Productivity distribution by race for the training and validation data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Race Group** | **Training Data N** | | | | **Validation Data N** | | | |
| Good | Fair | Poor | Total | Good | Fair | Poor | Total |
| White | 411 | 287 | 128 | 826 | 394 | 330 | 139 | 863 |
| African American/Black | 22 | 122 | 158 | 302 | 36 | 102 | 144 | 282 |
| Asian/Pacific Islander | 33 | 62 | 68 | 163 | 36 | 53 | 61 | 150 |
| Hispanic | 59 | 84 | 98 | 241 | 42 | 86 | 87 | 215 |
| Unknown | 253 | 384 | 241 | 878 | 394 | 330 | 139 | 901 |

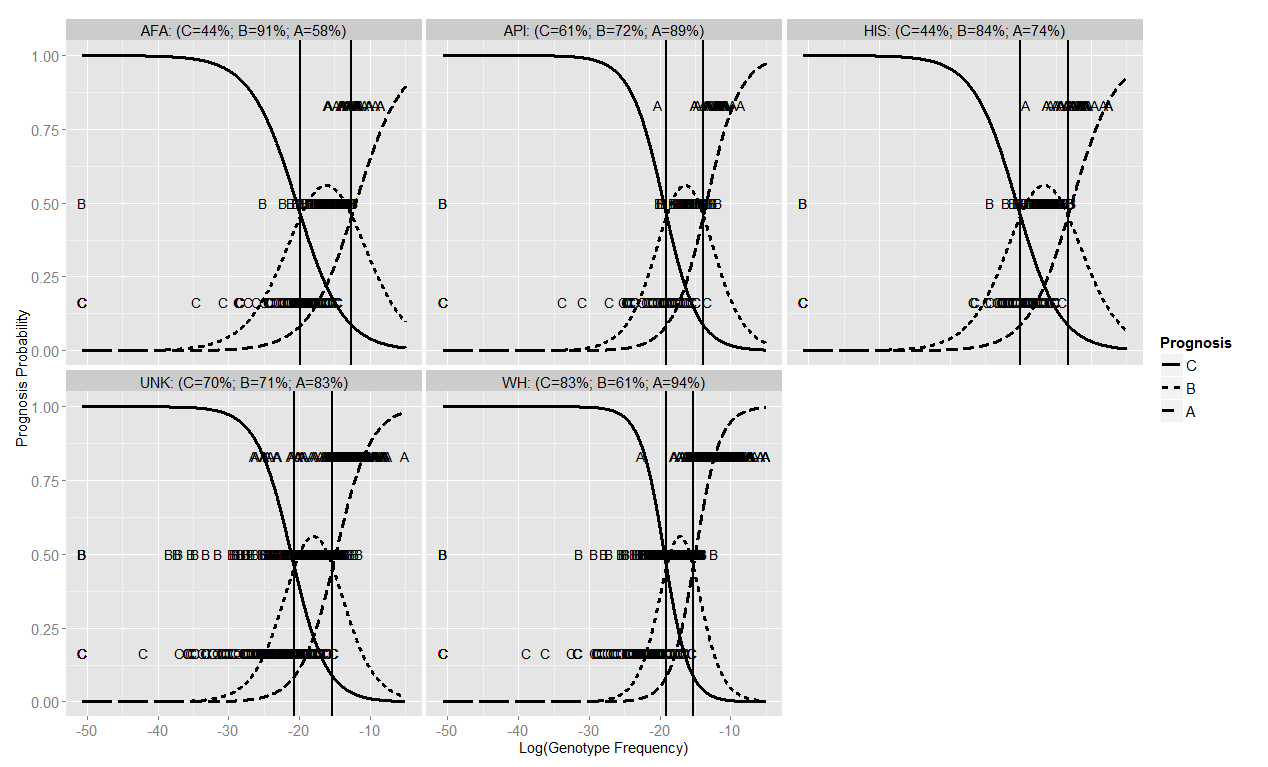
The GF classification boundaries derived from the model fit to the training data are shown in Table 2 for all race groups. For example, in the left column is the breakdown of the GF classification ranges for the WH race group. All GF greater than or equal to 2 x 10-7 are classified as “good” meaning these Pt likely have three or more potential 10/10 URD with a HapLogic predicted match of >50%. Fair is defined as having a GF between 1.99x10-7 and 6x10-10 and “poor” corresponds to GF less than 6 x 10-10.

**Table 2:** Proportional odds model classification boundaries

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Search Prognosis**  **Category** | **White** | **African American/Black** | **Asian/Pacific Islander** | **Hispanic** | **Unknown** |
| **Good** | GF ≥ 2.2x10-7 | GF ≥ 3.3x10-6 | GF ≥ 9.4x10-7 | GF ≥ 1.7x10-6 | GF ≥ 2.3x10-7 |
| **Fair** | 2.2x10-7 > GF ≥ 5.1x10-9 | 3.3x10-6 > GF ≥ 2.5x10-9 | 9.4x10-7 > GF ≥ 5.2x10-9 | 1.7x10-6 > GF ≥ 2.0x10-9 | 2.3x10-7 > GF ≥ 1.0x10-9 |
| **Poor** | 5.1x10-9 > GF | 2.5x10-9 > GF | 5.2x10-9 > GF | 2.0x10-9 > GF | 1.0x10-9 > GF |



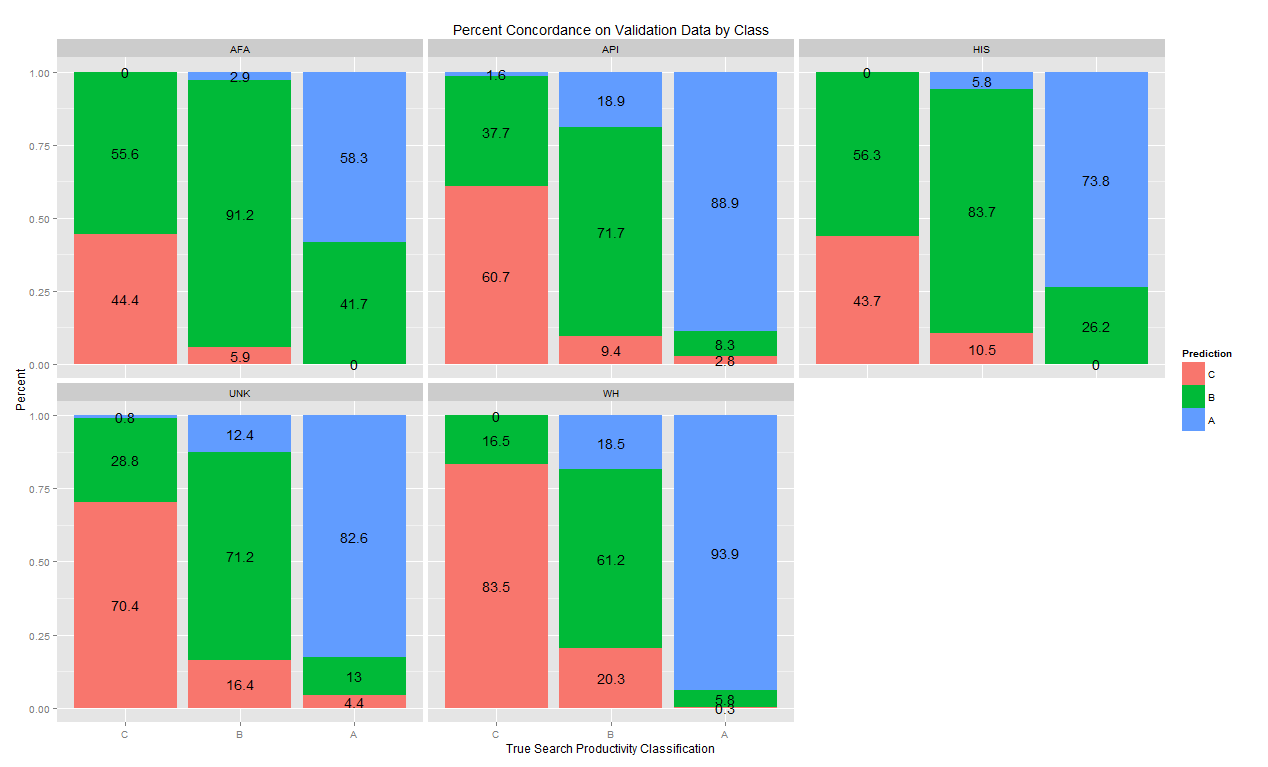
**Figure X: Model classification probabilities as a function of GF and Race for the Training data.** The vertical black lines denote the classification boundaries listed in Table X. Above each panel is the concordance (% correctly classified) of the model fit on the training data.



**Figure X: Model classification probabilities as a function of GF and Race for the Validation data.** The vertical black lines denote the classification boundaries listed in Table X. Above each panel is the concordance (% correctly classified) of the model fit on the validation data.

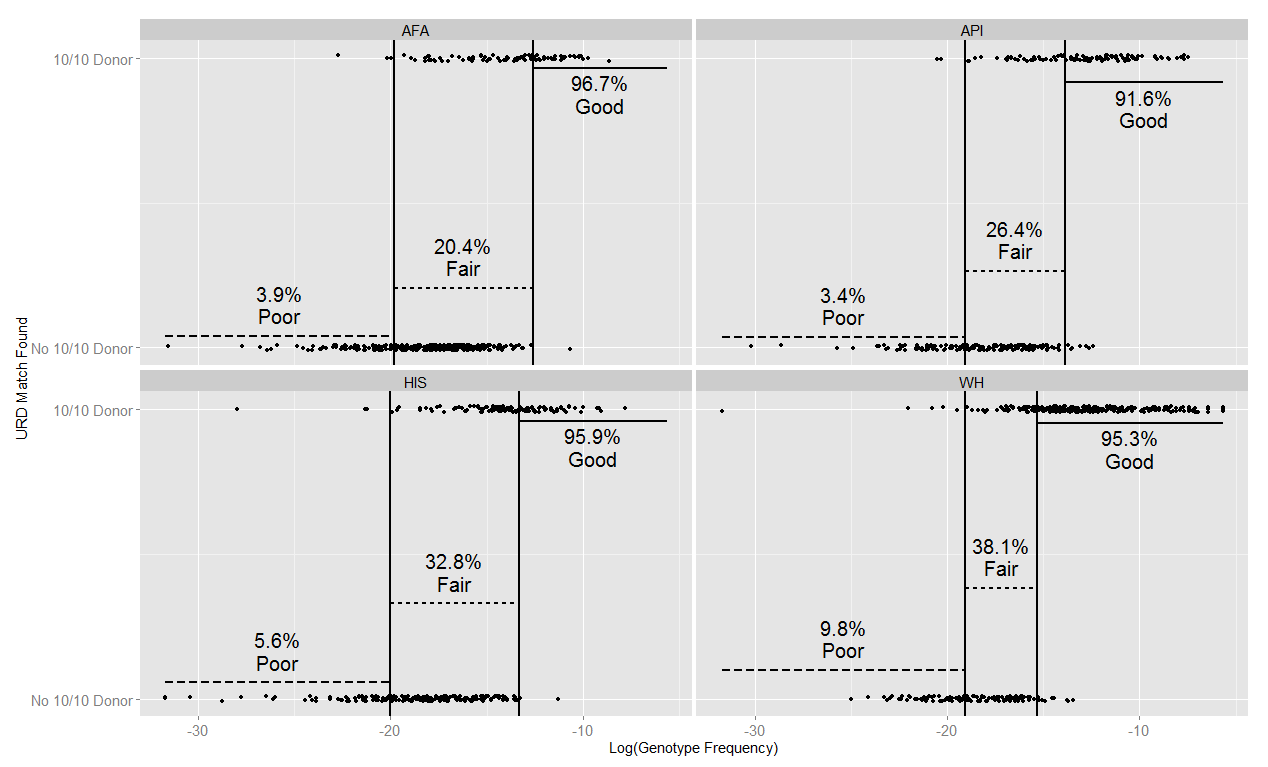
Figure 1 shows the results of the validation dataset’s concordance with their corresponding search prognosis category. The thick black border highlights groups where the model classification and search prognosis were concordant. When the model classification did not correspond with the search prognosis category, it was most often in a bordering category (97%). For example, 87% of WH searches with a GF in the good category had a corresponding search prognosis category of “good”, 13% had a search prognosis category of “fair”, and less than one percent were in the poor search prognosis category. The overall concordance by race for the validation dataset was WH: 83%, AFA: 77%, API: 74%, HIS: 70%, and UNK: 77% (Table 3).

Here is the data you’ll need to update the figure below. This is the validation data split out by race and truth versus predicted.

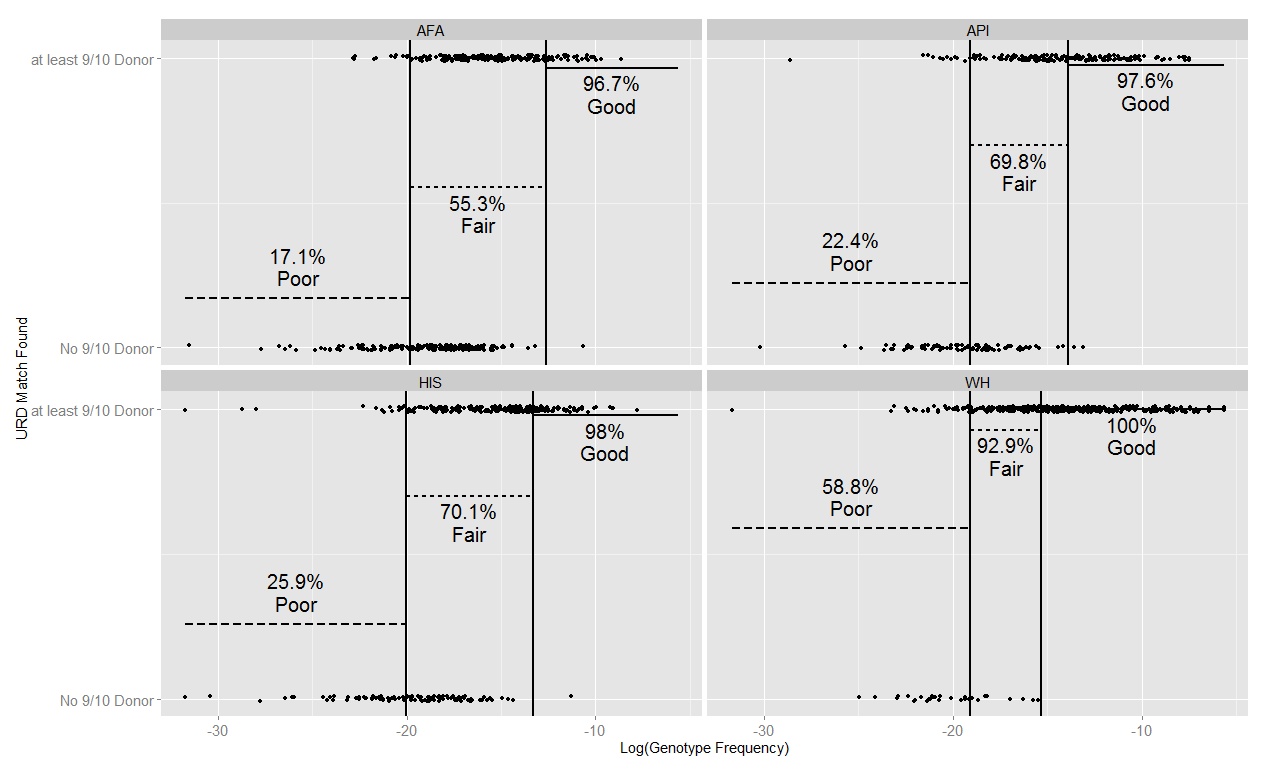


**Figure 1: Breakdown of the genotype frequency categories with their corresponding search prognosis category for the validation dataset.** The thick black borders indicate where the model classification and search prognosis were concordant. For example, in the API data, 91% of those Pt with a model classification of good had a corresponding search prognosis category of good. GF = Genotype Frequency, WH = White, AFA = African American/Black, API = Asian/Pacific Islander, HIS = Hispanic, UNK = Unknown

Finally, application of the predictive model to the independent dataset of pseudo-patients used in a previous study [19] resulted in 95% of WH pseudo-patients classified to the good GF category having a 10/10 URD, 95% in the AFA population, 93% in API, and 98% in HIS (Figure 2). For those classified to the fair GF category, 29% of WH pseudo-patients had a 10/10 URD, 38% in AFA, 36% in API, and 47% in HIS. Finally, of those classified to the poor GF category 10% of WH pseudo-patients had a 10/10 URD, 8% in AFA, 7% in API, and 11% in HIS.



**Figure 2: Percentage of pseudo-patients that could identify a 10/10 donor by predicted model class and race.** Model classification boundaries are shown along with individual results (slightly jittered to avoid over-plotting) to illustrate the genotype frequency effect within model classes with regard to finding a 10/10 URD.



**Figure 2: Percentage of pseudo-patients that could identify at least a 9/10 donor by predicted model class and race.** Model classification boundaries are shown along with individual results (slightly jittered to avoid over-plotting) to illustrate the genotype frequency effect within model classes with regard to finding at least a 9/10 URD.

**Discussion**

Our study presents a simple scoring system to help provide a quick prognosis of URD search productivity. This scoring system classifies a search into one of three search prognosis categories: good, fair, and poor. Those searches in the good category are very likely to have a potential 10/10 URD. Those searches in the fair category may have a 10/10 URD, but are likely to have a potential 9/10 URD. Finally, those searches in the poor category are less likely to identify a potential 10/10 URD, but may have a 9/10 URD. Even though the poor category is less likely to find a potential 10/10 URD, it does not mean one cannot be found or even that there are none available. To define these categories, we only included those URD with a HapLogic prediction of >50%. This may rule out any potential URD with B-C or DRB1-DQB1 associations that have some variability, but are still both common associations (e.g. DRB1\*07:01 with either DQB1\*02:02 or DQB1\*03:03 or DRB1\*04:01 with either DQB1\*03:01 or DQB1\*03:02). In some cases a 10/10 will be found, despite the HapLogic prediction is <50%. In addition, the Be The Match Registry only contains about half of the worldwide URD pool, so some URD may be found in registries not evaluated on the HapLogic listing.

These GF categories produced in this study could predict whether or not there was a potential 10/10 or 9/10 URD with a HapLogic prediction of >50% with good precision. The concordance for the WH group was the highest of the races evaluated, which may be the result of high numbers of subjects evaluated for HLA frequencies and more comprehensive reference data. The GF ranges for the AFA, API, and HIS race groups were also considerably different than the WH ranges, likely reflecting the diversity of HLA in these populations as well as the fact that the racial composition of URD on the Be The Match Registry are reported to be x% WH. The UNK GF ranges are most similar to the WH GF ranges, likely a result of the majority of Pt classified as UNK race being from European countries. The GF ranges for AFA, API, and HIS were similar to each other.

When the GF category and the search prognosis category were not concordant, it was most often in a bordering category (97%). The good GF category corresponded the most with the search prognosis category, and the fair category had the lowest level of precision. This may be because the fair GF category most closely correlates with predicting whether or not a 9/10 URD can be found. The mechanism of using the commonality of Pt HLA haplotypes demonstrated in this study can fail to adequately project potential donor haplotypes when one allele on an uncommon haplotype is exchanged for another allele (i.e. single mismatch with the Pt) which occurs as a frequent haplotype. This is less likely to occur when both haplotypes are uncommon, which is the case in the poor prognosis category.

Applying the GF ranges to the prior pseudo-patient dataset [19] was done for a few reasons. First, it allowed us to validate these categories on a completely independent dataset. It also allowed us to determine whether or not a potential 10/10 URD could be found without having the variability of transplant center activation and typing of potential URD. Real Pt searches do not progress for a multitude of reasons, but by applying these GF ranges to this dataset removed that variability because either there was already a 10/10 allele match URD, there was no potential 10/10 match, or URD were typed to resolve the matching status. More than 95% of pseudo-patients in the good GF category identified a 10/10 URD, slightly more than a third in the fair GF category, and only 8% in the poor GF category could identify a 10/10 URD. The results of this data fit nicely with the predicted GF ranges.

Although a reported Pt race GF should best reflect the Pt search, there are times where another race may more accurately reflect the actual URD list. Adding to the complexity is unstandardized practices on the collection of this information on Pt; from the potential of determinations being done by center staff using Pt name or observations of the Pt to self-reporting directly by the Pt. This is an area of ongoing research to identify how to best characterize HLA typing based on a genetic single race assignment or blend of haplotypes from multiple race groups. That could further optimize the UNK group or those Pt that could not identify a pair of haplotypes in a single race group haplotype frequency data.

Previous studies have developed similar search prognosis tools, but some were slightly more complicated and the tools developed were for predominately WH populations [12-15]. The method developed in this study uses one indicator – a Pt GF – that can easily be calculated using the haplostats.org application. It can also be applied to any broad race group (WH, AFA, API, or HIS) or when a Pt race is UNK. There is a potential for a tool to be developed that can automatically convert a Pt GF into a GF category to support transplant center processes and provide early information that can support clinical decision making. In addition to its use as a prognostic indicator for prospective Pt searches, this surrogate measure could also be applied to retrospective studies in need of an indicator of the productivity of the search.

The results of this study help provide some indication at the onset of a Pt search of whether or not a potential 10/10 or 9/10 URD can be identified. Although not a replacement for an actual URD search, the GF may offer a quick way for transplant physicians to get an indication of the likely search outcome, engage HLA expertise earlier in a Pt search process and guide early clinical consideration of non-fully matched URD alternative stem cell source options.

**References**

1. Petersdorf, E.W., *The World Marrow Donor Association: 20 years of international collaboration for the support of unrelated donor and cord blood hematopoietic cell transplantation.* Bone Marrow Transplant, 2010. **45**(5): p. 807-10.

2. Lee, S.J., et al., *High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation.* Blood, 2007. **110**(13): p. 4576-83.

3. Flomenberg, N., et al., *Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome.* Blood, 2004. **104**(7): p. 1923-30.

4. Woolfrey, A., et al., *HLA-C antigen mismatch is associated with worse outcome in unrelated donor peripheral blood stem cell transplantation.* Biol Blood Marrow Transplant, 2011. **17**(6): p. 885-92.

5. Horan, J., et al., *Evaluation of HLA matching in unrelated hematopoietic stem cell transplantation for nonmalignant disorders.* Blood, 2012. **120**(14): p. 2918-24.

6. Fernandez-Vina, M.A., et al., *Multiple mismatches at the low expression HLA loci DP, DQ, and DRB3/4/5 associate with adverse outcomes in hematopoietic stem cell transplantation.* Blood, 2013. **121**(22): p. 4603-10.

7. Burt, C., et al., *In a 12-allele analysis HLA-DPB1 matching is associated with improved OS in leukaemic and myelodysplastic patients receiving myeloablative T-cell-depleted PBSCT from unrelated donors.* Bone Marrow Transplant, 2014. **49**(5): p. 657-63.

8. Pidala, J., et al., *Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation.* Blood, 2014. **124**(16): p. 2596-606.

9. Fleischhauer, K., et al., *Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study.* Lancet Oncol, 2012. **13**(4): p. 366-74.

10. Aversa, F., et al., *Hematopoietic stem cell transplantation from alternative donors for high-risk acute leukemia: the haploidentical option.* Curr Stem Cell Res Ther, 2007. **2**(1): p. 105-12.

11. Anasetti, C., F. Aversa, and C.G. Brunstein, *Back to the future: mismatched unrelated donor, haploidentical related donor, or unrelated umbilical cord blood transplantation?* Biol Blood Marrow Transplant, 2012. **18**(1 Suppl): p. S161-5.

12. Pedron, B., et al., *Common genomic HLA haplotypes contributing to successful donor search in unrelated hematopoietic transplantation.* Bone Marrow Transplant, 2003. **31**(6): p. 423-7.

13. Pedron, B., et al., *Contribution of HLA-A/B/C/DRB1/DQB1 common haplotypes to donor search outcome in unrelated hematopoietic stem cell transplantation.* Biol Blood Marrow Transplant, 2011. **17**(11): p. 1612-8.

14. Joris, M.M., et al., *The impact of frequent HLA haplotypes in high linkage disequilibrium on donor search and clinical outcome after unrelated haematopoietic SCT.* Bone Marrow Transplant, 2013. **48**(4): p. 483-90.

15. Tiercy, J.M., et al., *The probability of identifying a 10/10 HLA allele-matched unrelated donor is highly predictable.* Bone Marrow Transplant, 2007. **40**(6): p. 515-22.

16. Pidala, J., et al., *Race/ethnicity affects the probability of finding an HLA-A, -B, -C and -DRB1 allele-matched unrelated donor and likelihood of subsequent transplant utilization.* Bone Marrow Transplant, 2013. **48**(3): p. 346-50.

17. Gragert, L., et al., *Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry.* Hum Immunol, 2013. **74**(10): p. 1313-20.

18. Maiers, M., L. Gragert, and W. Klitz, *High-resolution HLA alleles and haplotypes in the United States population.* Hum Immunol, 2007. **68**(9): p. 779-88.

19. Dehn, J., et al., *8/8 and 10/10 High-resolution match rate for the Be The Match Unrelated Donor Registry.* Biol Blood Marrow Transplant, 2014.