# What is the likelihood of finding a suitable stem cell donor

At present, there are close to 29 million potential stem cell donors in the Bone Marrow Donors Worldwide registry [4]. Though the number of donors continues to grow worldwide, there are significant resource implications in donor recruitment and HLA typing. Therefore, the challenge of thoughtful donor recruitment strategy becomes increasingly relevant. These include recruitment efforts focused on young male donors [5] or on relatives of registered donors with rare human leukocyte antigen (HLA) phenotypes [6], minority donor recruitment programs [7-10], and regional priority setting of recruitment activities based on HLA frequency differences [11-14].

The decisive question of " What is the likelihood of finding a suitable matched adult donor in their registry?" definitely warrants registries strategy planning. Recently, Schmidt, et al [15] reported that population-specific matching probabilities (MP) are a key parameter to assess the benefits of unrelated stem cell donor registries and the need for further donor recruitment efforts. The authors described a general framework for MP estimations of specific and mixed patient populations under consideration of international stem cell donor exchange. Calculations were based on HLA-A, -B, -C, -DRB1 loci high-resolution haplotype frequencies (HF) of up to 21 populations. Based on the existing donor numbers, the largest MP increases in addition of 500, 000 same-population donors were observed for patients from Greece (+0. 21) while only small MP increases occurred for European Americans (+0. 004) and Germans (+0. 01). Due to the large Chinese population, the optimal distribution of 5, 000, 000 new donors worldwide included 3. 9 million Chinese donors [15]. Nevertheless, the

authors observed the need for same-population donor recruitment in order to increase population-specific MP efficiently.

National strategies that neglect domestic donor recruitment should therefore be critically re-assessed, especially if only few donors have been recruited so far.

As described by Schmidt et al [15], the probability *p(n)* for a random patient from a given population to find at least one matching donor in a registry including n individuals of a donor population is given with *p(n)* is the matching probability in " n" sample size, *fi* being the frequencies of the *i* -th genotype and *i* -th is any genotype from the rank of genotypes in the order of the highest to the lowest frequencies in a donor population. Genotype frequencies can be derived from the estimated HF under the assumption of Hardy-Weinberg equilibrium (HWE).

HF is calculated from DNA-typed registry donors with Markov Chain Monte Carlo (MCMC) algorithm *PHASE* [16]. Four-locus high-resolution HF (HLA-A, HLA-B, HLA-C, and HLA-DRB1) were used for adult donors. The HF and effective adult-donor registry size for each group were then put into a matching model that assumes genotypes are in HWE [17, 18]. The strategy involved modeling the likelihood that an 8/8 or 7/8 HLA-matched adult donor was available. For better analysis, information of adult-donor availability including donor refusal, discrepant donor typing and loss of contact would be desirable.

According to the calculations, the likelihood of finding an available 8/8 HLA matched donor is 75% for white patients of European descent but only 46%

for White patients of Middle Eastern or North African descent [19]. Similarly, the chance of finding an 8/8 HLA-matched donor for other groups is lower and varies with racial and ethnic background. For Black Americans of all ethnic backgrounds, the probabilities are 16 to 19%; for Asians, Pacific Islanders, and Native Americans, they range between 27% and 52%.

As it was reported that adult-donor availability differs according to racial and ethnic background [19], models including this variable have substantially lower match likelihoods than those which did not take into this account. Although the likelihood of HLA matching is the greatest with donors from the patient’s racial and ethnic group, donors from other racial and ethnic groups may increase this likelihood. Patients from groups with relatively low inter-racial or inter-ethnic marriage, such as Asian groups, are less likely to have donors identified from outside their group.

The overall available rate is only 29%.

We therefore estimated the donor pool and matching probability in this study based on our previous published gene and haplotype frequencies in Hong Kong population [20].

MATERIALS AND METHODS

Sample Collection and genotyping

As reported previously, 7, 595 voluntary unrelated bone marrow donors recruited by the HKBMDR between January 2013 and June 2014 were included in the analysis [20]. All donors are of Chinese origin, HLA-A, -B, -C and -DRB1 genotypes were obtained using polymerase chain-reaction

sequence-specific oligonucleotide probe methods using LifeCodes HLA-SSO Typing Kit (Gen-Probe, Stamford, CT) when analysed by Luminex 200™ system (Luminex Corp., Austin, TX). Typing ambiguity was resolved using sequence specific primer or sequence based typing methods utilising the specific primers of SBTexcellerator® HLA typing Kit (Genome Diagnostics, Utrecht, the Netherlands). Alleles were determined according to IMGT/HLA Database release 3. 18. 0.

Statistics Analysis

The frequencies of HLA-A, -B, -C and -DRB1 alleles were calculated from the number of observed genotype. Hardy-Weinberg equilibrium for each loci was assessed by PyPop using MCMC simulation from Guo and Thompson [21], and genotype frequency deviance within each loci was detected by PyPop invoking Arlequin [22]. P value of <0. 01 was considered as statistical significant.

By using the formulae described by Schmidt et al [15] with modification, the probability *p(n)* for a random patient from a given population to find at least one matching donor in a registry including n individuals of a donor population is given with *p(n)* is the matching probability in “ n” sample size, *fi* being the frequencies of the *i* -th genotype and *i* -th is any genotype from the rank of genotypes in the order of the highest to the lowest frequencies in a donor population.

RESULTS

The HLA genotypes and haplotypes frequency mentioned in the following section have been recently published [20]. HLA-A, -B, -C and -DRB1 genotypes deviated from the expected Hardy-Weinberg Equilibrium Proportions (HWEP) (p < 0. 001) shown in Table 1 [20]. Four-locus haplotype frequencies were estimated using the Markov Chain Monte Carlo algorithm *PHASE* [16]; adherence to HWEP was also assessed using PyPop 0. 7. 0 [23]. A few but significant deviations from HWEP were detected for all the four loci, HLA-A, -B, -C and -DRB1. Deviation from HWEP detected at the HLA-A locus is derived primary from an excess of A*02: 01 + A*02: 03 genotypes (247 observed, 218. 5 expected; p = 0. 0007) and an undercount of A*02: 06 + A*02: 03 genotypes (16 observed, 48. 2 expected; p = <0. 0001). List of the deviations from HWEP at HLA-A, -B, -C and -DRB1 are in Table 2. The allele and haplotype frequencies data are available in the Allele Frequencies Net Database under the population name 'Hong Kong Chinese BMDR' and the identifier (AFND 3357) [24].

Summary statistics for Hong Kong haplotypes is shown in Table 3. 2, 146 A-C-B-DRB1 haplotypes with frequencies > 0. 006% were estimated from these donors. The cumulative frequency distributions for HLA-A, -B, -C and -DRB1 loci in this Hong Kong Chinese cohort are shown in Table 4. Top twenty Haplotype A-C-B-DRB1 frequencies are shown in Table 5 [20]; nine of them have frequencies of greater than 1%. Our findings on HLA alleles and haplotypes frequencies were found to be very similar to those of Asian/Pacific Islander (A/PI) Race/Ethnicity of the NMDP Registry and other studies on Han Chinese population [25]. The most common haplotype *A*33: 03-C*03: 02-B*58: 01-DRB1*03: 01* ranked second in the A/PI of NMDP

registry (2. 3%) and top in Singapore Chinese (5. 1%) [26]. The second most common haplotype *A*02: 01-C*01: 02-B*46: 01-DRB1*09: 01* was one of most frequent haplotypes among Chinese populations, especially the southern area of China and Guangdong [27, 28]. However, the fifth common haplotype *A*02: 03-C*07: 02-B*38: 02-DRB1*16: 02* , was found to be less common in the A/PI of NMDP Registry (0. 4%) and the mainland China (0. 3%) [25, 28].

We compared the top 100 haplotypes of HKBMDR & HKCBB by haplotype frequencies with the two publications [25, 26]; we noted that 88 are in common, the rank correlation is 0. 909 for HLA-A-B-DRB1 haplotype. There appears to be no excessive immigration from other places to Hong Kong. We also compared the China population paper which had provided the detailed top haplotypes for 4 loci, we found that 43 are common in HLA-A-C-B-DRB1 haplotype and the correlation is low with only 0. 477 [28].

With the use of MCMC algorithm to estimate HLA haplotype frequencies [14], it was found that the number of haplotypes increases with number of donor samples studies as summarized in Table 6. Originally we tested the HLA haplotype frequencies in 2, 500 samples and noted a bigger number of haplotypes as compared with other papers. Then we increased the sample size to 5, 000 and 7, 500 and noted that the increase was quite significant in our population with many more haplotypes. However, we usually observed a plateau of number of haplotypes even with increase in sample size in the Caucasians and European populations.

As of December 2015, there were only around 100, 000 donors in the HKBMDR. Applying the similar methodology in calculating the likelihood of finding a “ matched” donor in US [19], likelihood of finding an 8/8 HLA match or > 7/8 HLA Match by different donor registry size in the HKBMDR was shown in Figure 1. The likelihood of finding an available 8/8 HLA matched donor is 45% while increases to 65% for finding 7/8 HLA matched donor. It is similar to the finding of other studies conducted among Asians, Pacific Islanders, and Native Americans which reported a likelihood ranging between 27% and 52% [19].

DISCUSSION

The chance of successful engraftment and disease free survival are associated with the HLA compatibility between the recipient and the prospective donor. The diversity of the HLA genes at the allelic level and the heterogeneity of HLA data of the registered donors have a significant bearing on the probability of finding a volunteer unrelated HSC donor for patients from a particular population. This can be seen in the existence of many populations including Hong Kong or Chinese with significant heterogeneity among recruitment centers. HLA frequencies estimated at the Hong Kong Bone Marrow Donor Registry or China Marrow Donor Program Registry are not in equilibrium and should not be relied on as characteristic of a ” Chinese population”.

The probabilities of finding a match would increase substantially when the registry size grows.

As reported in [19], the NMDP has added slightly more than 1 million adult donors to the registry in 2012 and plans recruitment growth of 9% cumulatively each year through 2017.

HLA typing of Chinese in Hong Kong were found to be more heterogeneous and this points to the need of a larger donor pool in bone marrow registry to optimize the chance of successful matching. The study findings provide vital information for defining donor recruitment target and planning for extra resources in order to support the cost in donor recruitment and HLA typing. Establishment of a more cost-effective bone marrow donor registry with a larger pool of donors could increase chance of matching and the success rate of haematopoietic stem cell transplantation.

Assuming 25, 000 per 10-year age range of even distribution, it is projected that the number of retired and non-contact to be around 2, 000. Based on the projection in Figure 1, if one would like to achieve MP for 50% 8/8 HLA Match or 70% > 7/8 HLA Match, HKBMDR should have about 150, 000 donors. Considering the HKBMDR registry size to grow to 150, 000 in five-year time, it will require 12, 000 new donors recruitment per year. To further increase MP to nearly 55% for 8/8 HLA Match or about 75% > 7/8 HLA Match, donor registry size should be expanded to 200, 000 (Figure 1). Similarly, an annual recruitment of 22, 000 new donors is required. Either of them is much higher than the current recruitment target of 5, 000 donors per year. As such, the associated resource implication in donor recruitment and HLA typing will need to be carefully addressed. In our previous study on the survey on Hong Kong donation [29], factors associated with HSC donation motivation in Hong Kong were identified. The results highly suggested

that recommendations on promoting BM donation to the younger and higher education may allow better recruit rate and longer maintenance for donation. The government should consider launching educational activities such as bone marrow donation campaign, educational series and school talks to students and parents.

However, it should be noted that the above estimation has not taken into account of the potential matches from around 2, 400, 000 Chinese donors registered in China and Taiwan registries. In addition, the use of cord blood units which are readily available and require less stringent HLA matching has not been added into the matching probability. Many transplant centers in particular those in the States and East Asia would switch to use cord blood when adult donor is not available. But the relatively low stem cell dose may be inadequate for adult size recipient. Recently, double cord blood or even haploidentical transplant has been applied clinically with success. Whether they will eventually replace the need of a large registry is currently under debate. But at the moment, these approaches are mainly indicated when conventional related or unrelated donors are not readily available or accessible. On the other hand, one should also be bear in mind the time required from matching, donor work up to donation of overseas donors and other cost implication factors when building up the model for estimation of registry size