

## Literature Review

# Next Generation Sequencing and Histocompatibility in Hematopoietic Stem Cell Transplantation

Rhonda Holdsworth<sup>1</sup>, Mary Diviney<sup>2</sup> and Catherine Hart<sup>2</sup>

<sup>1</sup>Transplantation and Immunogenetics Services; <sup>2</sup>Victorian Transplantation and Immunogenetics Service, Australian Red Cross Blood Service, Australia

The human leucocyte antigen (HLA) molecules are encoded by a series of closely linked genes on the short arm of chromosome 6 and are part of the major histocompatibility complex (MHC). The HLA genes encode antigen presenting proteins of the immune response molecules, HLA-A, -B, -C, -DRB1, -DOB1, -DPB1 which have a crucial role in the adaptive immune system. The HLA system is the most polymorphic gene system in the human genome, now termed hyper polymorphic, more than 18,000 HLA alleles have been described on The International Immunogenetics Information System (IMGT database). <http://www.ebi.ac.uk/ipd/imgt/hla/stats.html>

The importance of HLA allele matching in haematopoietic stem cell transplantation (HSCT) has been described in many publications over the past 3 decades.<sup>1-6</sup> Over this period of time, HLA typing methods and technology has evolved from serological and cellular typing methods to molecular methodologies achieving greater definition of the HLA antigens and alleles. HLA nomenclature has similarly changed from an antigen based nomenclature (serology) to a HLA allele based nomenclature evolving over time to accommodate the increasing variation identified at the molecular level. The development and application of next generation sequencing (NGS) has enabled HLA nucleotide sequence to be described across all exons and introns characterising an unambiguous HLA allele result at the 4 field level for up to 11 HLA genes.

Assessment of matching and outcome risk for stem cell transplantation varies depending on the source of stem cells e.g. sibling donors with inherited compatibility to unrelated or cord blood donors with matching based on individual HLA loci or estimated haplotypes.

In the past, HLA typing methods were aimed at detecting variation across the antigen recognition site (ARS) coded by exon 2 and 3 for HLA class I and exon 2 for HLA class II. HLA variation across the ARS influences the specificity of bound peptides and the interaction of ligands including the T cell receptor and KIR impacting on the immune response in the recognition of self and non self. Therefore, HLA variation across this region is clinically significant for matching in HSCT.

The WMDA guidelines for matching published in 1999<sup>7,8</sup> indicated a minimal assessment of loci to be considered in matching were the HLA -A,- B, -DRB1 loci only. However, few stem cell transplant units would limit HLA matching assessment on these 3 loci today. Risk assessment of a potential stem cell donor using HLA typing data by NGS has provided a broader concept of matching or mismatching but it is not yet clear the clinical impact of this level of resolution.

## HLA typing

The hyper polymorphism of the HLA genes has presented significant technical challenges for HLA typing. HLA typing by serology was dependent on allo-antisera or monoclonal antibodies and limited to known polymorphisms in the HLA-A,-B,- DRB1 and some -DOB1. Serology was quickly superseded by molecular methodologies in the post PCR<sup>9</sup> era and the number of HLA alleles described increased dramatically. Initial molecular methods such as sequence specific primer (SSP) or sequence specific oligonucleotide primers (SSO) were practical and robust but limited to identifying specific exons and technically difficult as the HLA alleles increased.

The development and application of sanger sequenced based typing (SBT)<sup>10</sup> methods for HLA typing enabled further characterisation of the HLA alleles and has been a reliable workhorse for HLA typing over many years. Sanger SBT was the catalyst for the update of the HLA nomenclature to HLA v3 as many new alleles were described using this methodology.

However, Sanger SBT had several drawbacks including sequencing for short range segments, heterozygote phase ambiguity and does not detect variation across exons or introns outside of the ARS. Consequently, some combinations of HLA alleles were ambiguous and it was not possible to detect HLA alleles defined outside the ARS making it difficult to accurately determine HLA type. Additional sequencing methods were required to finalise a result.

The development of NGS<sup>11,12</sup> or massively parallel sequencing technologies provides cost effective and accurate HLA allele assignment through elimination of phase ambiguities and multiplexing of samples enabling specific assignment of HLA alleles. As NGS detects variation across the whole HLA gene, many new HLA alleles have been identified making assessment of matching in HSCT more complex.

In addition, there are variations of workflows using NGS. For example amplicon based NGS is a high throughput cost effective protocol for HLA typing. However, amplicon based NGS protocols do not sequence the whole HLA gene. The result may be reported as an NGS derived high resolution result but the range of sequencing is limited highlighting the importance of defining the level of HLA typing in results.

### Defining levels of HLA typing

The definition of allele level resolution<sup>11</sup> has changed with the implementation of NGS for HLA typing. HLA typing by NGS questions the definition of a high resolution result, which until recently was the HLA allele characterised by the nucleotide sequence across the antigen recognition site but should now be the HLA

allele characterised by whole HLA gene sequencing. A generic result remains as a 2 digit or 1 field HLA result describing the antigen group e.g. HLA-B\*44. This is sometimes described as a serological equivalent, a legacy from serology HLA typing era. A high resolution typing can currently be performed by either Sanger SBT or NGS and the result for both technologies can be reported as 2, 3 or 4 field e.g. HLA-B\*44:02, 44:02:01 or B\*44:02:01:01. The definition of a high resolution result is evolving and moving towards a definition where full HLA gene typing by NGS is used to define a specific HLA allele.

### Assessing 'Matching' for HSCT using NGS

Assessing HLA mismatching for HSCT in the NGS era has created new challenges as patients and their potential donor(s) can be matched across the ARS but mismatched in their high resolution specific HLA alleles as determined by NGS. As the clinical implications of HLA variation outside of the ARS are not yet determined,<sup>13,14</sup> it is difficult to assess the impact of mismatching for potential donor and recipients in HSCT. An excellent review by Hurley et al<sup>14</sup> suggests that risk assessment of matching for HSCT should remain within the ARS for now. A decade of Sanger sequencing established the current understanding of the differences in HLA alleles and therefore assessment of mismatching for related and unrelated stem cell donors or cord bloods based on the ARD.

The International Histocompatibility Workshop HSCT project was established for the 13<sup>th</sup> IHIWS held in 2003 and lead by Prof Effie Petersdorf. The aim of the project was to accumulate a large, population diverse data set for HSCT to assess the effect of mismatching across HLA alleles or loci and aimed at identifying permissive mismatches. This international collaborative group of the IHIWS has continued to accrue data across 14<sup>th</sup>, 15<sup>th</sup>, 16<sup>th</sup> and 17<sup>th</sup> IHIWS and comprises an extremely large data set to assess HLA compatibility with transplant outcome.

NGS data was used for the first time in the 17<sup>th</sup> IHIWS HSCT project and may provide some information about whole HLA gene variation and the impact on HSCT outcomes. NGS challenges the traditional approach of reporting matches per loci. To date publications on HSCT reports on 8/8 or 10/10 when comparing matching across the HLA loci: HLA -A,-B,-C, -DRB1, -DOB1 and -DPB1. However, this approach has historically referred to HLA allele matching across the ARS. If the patient and unrelated donor are HLA typed using NGS, it is possible that the alleles will be different and matching reported as 4/8 but 8/8 across the ARS. As laboratories implement NGS as the routine method for a high resolution HLA typing for HSCT, the compatibility assessment between the patient and donor needs to state what part of the gene is included in the assessment.<sup>15</sup>

Recent discussions with our HSCT clinical colleagues as to how we should report recipient and donor compatibility has suggested that HLA typing reports using NGS data should report the specific HLA alleles for donor and recipient and qualifying matching or mismatching at the ARS.<sup>16</sup> A recipient and donor may therefore be HLA allele mismatched but matched across the ARS (this should be stated) and any assessment of number of loci matched or mismatched remain within the current assessment of sequence homology across the ARS. This would ensure consistency with the current understanding of risk assessment in the HSCT setting.

#### **Searching for unrelated donors and cords in the NGS era**

Stem cell registries contain very large databases of donor HLA types acquired over many years and characterised using different HLA typing methods. The detail of a HLA result will vary depending on the testing protocol used. Not only has the donor or cord blood been HLA typed with differing methods over time, how these are reported in registries can vary. A HLA result can appear to be a high resolution result but has only been detected using a molecular method such as SSO or SBT without qualifying the limitations of the

technology. Whilst most transplant centres request verification typing to ensure an up to date HLA result the cost of selecting donors for verification can limit the number of potential donors any one centre can request.

The HLA G group nomenclature provides a short hand format grouping together HLA alleles, which share sequence homology across the ARS but may be different outside of the ARS region. Some registries have utilised this HLA nomenclature format so that donors who are matched across the ARS can be found and further assessed for a recipient. This G code smart approach has provided a practical solution for registries as the patient may be HLA typed using NGS and therefore express a HLA allele expressing variation outside of the ARS and therefore not identify potentially matched donors.<sup>16</sup> HLA alleles which are designated new are not able to be searched in registries without an assigned allele nomenclature.

#### **NGS data analysis**

There are several excellent commercial kit options for HLA typing by NGS. All kits provide good coverage of the HLA genes and are able to assign a high resolution HLA allele result with their software.<sup>17</sup> There are also software providers who can assist with data analysis for in house protocols. Whilst there may be some variation on the laboratory workflows in testing, it is really up to the individual labs to assess their requirements based on sample numbers and throughput as to particular kit selection. In addition to laboratory workflows considerations, NGS generates a lot of data and analysis time can be critical for testing requirements in the routine tissue typing laboratory.<sup>18</sup> Laboratories test turnaround times (TAT) are measured as part of laboratory accreditation requirements and slow data analysis can negatively impact on test turnaround times.

It cannot be emphasised enough that when implementing NGS for HLA typing there should be substantial consideration on the server type, configuration and capacity for data analysis<sup>19</sup> and storage. NGS will

generate several orders of magnitude increase in data compared to Sanger SBT and as more and more HLA sequences are characterised using NGS, reference data bases are increasing in size. In HSCT testing, TAT can affect verification typing and donor selection as delays potentially impede the transplant going ahead.

### Future technologies

The complexity of the HLA system and its application in transplantation or immunogenetics and genomics will continue to drive development of newer technologies. How the HLA typing result is determined can characterise the level of resolution or impact the perceived structure and function of the HLA molecule as determined by nucleotide variation. Initially only the exons of the HLA gene structure was thought to be important but our knowledge on the interactive effect of the whole HLA gene on expression of the alleles or other HLA loci is rapidly growing.

HLA typing by NGS has provided as insight into the potential importance of the degree of polymorphism in transplantation and HSCT in particular.<sup>20</sup> Assessing transplant risk with HLA mismatching will most likely change as the technologies advance and more detailed results can be generated in short timeframes enabling donor selection to be more rapid and more specific.

### Conclusion

The HLA system is one of the most polymorphic gene systems in the human genome. HLA allele mismatching has a well described clinical impact on transplant outcome in HSCT. Assessing the risk of HLA mismatching in HSCT is dependent on an accurate and timely HLA typing method. Until recently HLA typing was mainly focused on defining variation across the nucleotide sequence which coded for the ARS for a number of key genes. In the era of NGS full gene sequencing is now possible. Consequently, the concept of HLA allele mismatching has changed. However, there is not yet sufficient evidence to assess the clinical impact of

nucleotide variation outside of the ARS. NGS provides a cost effective and accurate method for HLA typing but more work is needed to fully understand the clinical impact of variation across the whole HLA gene. Coupled with historic HLA typing data in bone marrow registries and cord blood banks, assessment of HLA matching or mismatching had become more complex.

### References

1. Petersdorf EW. Genetics of graft-versus-host disease: the major histocompatibility complex. *Blood Rev.* 2013;27:1-12.
2. Shaw BE, Arguello R, Garcia-Sepulveda CA, Madrigal JA. The impact of HLA genotyping on survival following unrelated donor haematopoietic stem cell transplantation. *Br J Haematol.* 2010;150:251-8.
3. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood.* 2007;110:4576-83.
4. Morishima Y, Kashiwase K, Matsuo K, Azuma F, Morishima S, Onizuka M, et al. Biological significance of HLA locus matching in unrelated donor bone marrow transplantation. *Blood.* 2015;125:1189-97.
5. Fernandez-Vina MA, Wang T, Lee SJ, Haagenson M, Aljurf M, Askar M, et al. Identification of a permissible HLA mismatch in hematopoietic stem cell transplantation. *Blood.* 2014;123:1270-8.
6. Qi J, Wang TJ, Chen LP, Wang XF, Wang MN, Wu JH. Utility of next-generation sequencing methods to identify the novel HLA alleles in potential stem cell donors from Chinese Marrow Donor Program. *Int J Immunogenet.* 2018;45:225-9.
7. Hurley CK, Wade JA, Oudshoorn M, Middleton D, Kukuruga D, Navarrete C, Christiansen F, et al. Histocompatibility testing guidelines for hematopoietic stem cell transplantation using volunteer donors: report from The World Marrow Donor Association. *Bone Marrow Transplant.* 1999;24:119-21.
8. Hurley CK, Wade JA, Oudshoorn M, Middleton D, Kukuruga D, Navarrete C, et al. A special report: histocompatibility testing guidelines for hematopoietic stem cell transplantation using volunteer donors. *Tissue Antigens.* 1999;53:394-406.
9. Erlich H. HLA DNA typing: past, present, and future. *Tissue Antigens.* 2012;80:1-11.
10. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA.* 1977;74:5463-7.
11. Gabriel C, Fürst D, Faé I, Wenda S, Zollikofer C, Mytilineos, et al. HLA typing by next-generation sequencing - getting closer to reality. *Tissue Antigens.* 2014;83:65-75.

12. Lind C, Ferriola D, Mackiewicz S, Heron S, Rogers M, Slavich L, et al. Next generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Hum Immunol.* 2010;71:1033-42.
13. Hou L, Vierra-Green C, Lazaro A, Brady C, Haagenson M, Spellman S, et al. Limited HLA sequence variation outside of antigen recognition domain exons of 360 10 of 10 matched unrelated hematopoietic stem cell transplant donor-recipient pairs. *HLA.* 2017;89:39-46.
14. Hurley CK, Ng J. Continue to focus clinical decision making on the antigen recognition domain for the present. *Hum Immunol.* 2018 Apr 21. pii: S0198-8859(18)30117-4.doi:10.1016/j.humimm.2018.04.010. [Epub ahead of print]
15. Nunes E, Heslop H, Fernandez-Vina M, Taves C, Wagenknecht DR, Eisenbrey AB, et al. Definitions of histocompatibility typing terms. *Blood.* 2011;118:e180-3. doi:10.1182/blood-2011-05-353490. Epub 2011 Oct 14.
16. Tiercy JM. How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica.* 2016;101:680-7.
17. 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.
18. Weimer ET. Clinical validation of NGS technology for HLA: An early adopter's perspective. *Hum Immunol.* 2016;77:820-3.
19. Larjo A, Eveleigh R, Kilpeläinen E, Kwan T, Pastinen T, Koskela S, et al. Accuracy of programs for the determination of human leukocyte antigen alleles from next generation sequencing data. *Front Immunol.* 2017;8:1815.
20. Hosomichi K, Shiina T, Tajima A, Inoue I. The impact of next-generation sequencing technologies on HLA research. *J Hum Genet.* 2015;60:665-73.

