The Major Histocompatibility Complex: A Model for Understanding Graft-versus-Host Disease

Effie W. Petersdorf, MD

Professor of Medicine, University of Washington

Member, Division of Clinical Research

Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

Address:

Fred Hutchinson Cancer Research Center

Division of Clinical Research, D4-115

1100 Fairview Avenue North

Seattle, WA 98109

Tel 206-667-5244

FAX 206-667-5255

epetersd@fhcrc.org
Abstract

Acute graft-versus-host disease (GVHD) afflicts up to 80% of all patients receiving an unrelated donor hematopoietic cell transplant (HCT) for the treatment of blood disorders, even with optimal donor HLA matching and use of prophylactic immunosuppressive agents. Of patients who develop acute GVHD, many are at risk for chronic GVHD and bear the burden of considerable morbidity and lowered quality of life years after transplantation. The immunogenetic basis of GVHD has been the subject of intensive investigation, with the classical HLA genetic loci the best-characterized determinants. Recent information on the major histocompatibility complex (MHC) region of chromosome 6 as an important source of untyped genetic variation has shed light on novel GVHD determinants. These data open new paradigms for understanding the genetic basis of GVHD.
Introduction

The success of hematopoietic cell transplantation is influenced by many factors including HLA mismatching, MHC region variation, minor histocompatibility targets of allorecognition, regulatory elements that affect gene expression, and genetic variation that affects immune responses. The individual contribution from each of these factors and pathways may be challenging to quantify in individual patients, however sophisticated tools and analysis methods are available for assessing the global impact of genetic variation on transplant outcome.

The best studied region of the genome is the major histocompatibility complex (MHC) on chromosome 6. The MHC is a 7 megabase (Mb) region with dense clustering of over 300 genes that have coordinated roles in immune function. With an average gene density of 1/16 kilobases (kb), it is no surprise that over 420 associations to more than 35 common complex autoimmune, inflammatory, infectious diseases and cancer have been described. The classical HLA loci are the best characterized MHC genes. HLA matching lowers the risks of non-engraftment, acute and chronic GVHD and mortality, however GVHD remains a significant complication in 20–80% of patients, and points to non-HLA variation as potential factors.

A comprehensive survey of MHC polymorphisms for potential GVHD determinants requires a complete and precise map of sequence variation at the level of the HLA haplotype, a resource that was made available in 1999 by The MHC Sequencing Consortium, and extended through the efforts of The MHC Haplotype Project, The International HapMap Consortium, and The 1000 Genomes Project Consortium. These resources have provided an unprecedented view of the MHC as a rich source of novel variation that has implications in disease and transplantation. Although the exploration of the MHC in the transplant model is still in the early stages, new information points to the contribution of MHC haplotype-linked variation from both the patient and the donor, and enlightens the possible mechanistic
pathways through which GVHD arises. In this review, the fundamentals of mapping the MHC for functional variants in unrelated donor HCT are described. Terminology commonly used in HLA immunogenetics and disease mapping are provided in Table 1.

The Genetic Landscape of the MHC: Coordinated Immune Function, Diversity, and Linkage Disequilibrium

Three hallmarks of the MHC are the very dense clustering of genes with related immunological function, extreme sequence diversity of expressed proteins, and strong long-range positive linkage disequilibrium (LD). MHC region gene families include co-receptors (HLA-F, HLA-G), molecules for antigen presentation (26 class I and 24 class II genes), the innate immune response (MICA, MICB, HCP5), inflammation (NKBBil, LTA, TNF, LTB, LST1, NCR3, AIF1), immune receptors (LY6), heat shock proteins (HSPA1L, HSPA1A, HSPA1B), complement cascade, regulatory receptors (NOTCH4), antigen processing (TAP, HLA-DM, HLA-DO) and peptide transport (RING1). In addition to these gene clusters, the MHC is home to several large families of “house-keeping” genes including over 66 histone, 36 zinc finger and 157 tRNA genes. The co-existence of genes having coordinated immunological function is thought to reflect some survival advantage to keeping these genes together on the same haplotype.

The extreme polymorphism of classical HLA genes is part and parcel of their role in peptide presentation for host defense against pathogens, and as ligands for natural killer receptors involved in immune surveillance of altered or loss of self MHC. Linkage disequilibrium (LD) describes the non-random association of two or more markers at frequencies that are higher than would be predicted by chance alone, and is the driving force behind the maintenance of alleles on HLA haplotypes. The physical distance over which LD prevails can create exceptionally long MHC haplotypes that measure up to 9
mega basepairs (Mb) in some populations.\textsuperscript{22,23} New haplotypes are generated by the shuffling of blocks of conserved sequences at points of recombination.\textsuperscript{19}

The first complete sequence of the MHC was derived from multiple different HLA haplotypes.\textsuperscript{12} Subsequently, sequencing of cell lines homozygous at HLA loci provided haplotype-specific information on common and rare variants, information that has been instrumental for fine mapping.\textsuperscript{13-16,24-28} The complete sequencing of HLA haplotypes shows that the MHC is a microcosm of the human genome, encoding single nucleotide polymorphisms (SNPs), substitutions and deletion-insertion polymorphisms, short tandem repeats (microsatellites), and copy number variations (CNVs). The simplest form is the SNP, which is a powerful tool for surveying regions containing disease-associated genes.\textsuperscript{27,29-34} Strong positive LD among SNPs gives rise to “blocks” of short SNP haplotypes that can be readily identified by genotyping one representative “tag SNP” for the block. TagSNPs that have portability across ethnically diverse populations are particularly powerful tools for disease mapping disease-susceptibility genes (Figure 1).\textsuperscript{35-37}

**The MHC and GVHD**

The MHC is a prime region of interest for the exploration of novel GVHD determinants because of the density of genes with immune function that currently are not tested in routine clinical practice.\textsuperscript{11,40,41} Untyped patient and donor variation responsible for GVHD can now be surveyed with great precision with the availability of SNP arrays. The identification of GVHD determinants is performed using association mapping techniques. Candidate gene studies focus on specific genes that have a plausible basis to the disease. Genome-wide association studies (GWAS) test associations of SNPs to clinical endpoints without \textit{a priori} assumptions as to the nature of the gene(s) involved in disease pathogenesis. Both approaches provide important information on the genes and allele variants that may be clinically relevant in transplantation. Candidate gene approaches may permit genes with known putative function
to be analyzed individually or together with other candidate genes to assess the effects of gene-gene interactions that may affect the same biologic process or clinical endpoint. GWAS approaches, on the other hand, permit novel, heretofore unknown genes and pathways to be discovered, which can subsequently be further studied in directed analyses of specific genes or pathways. The choice of method is largely driven by whether there are specific mechanistic hypotheses that can best be addressed through the study of specific candidate genes, or whether the research question is to broadly explore and identify genome-wide factors, variants that have previously been characterized in candidate gene studies, as well as novel polymorphisms. Both approaches require *in vitro* or *in vivo* functional studies to ultimately define how the sequence variation affects gene function and how the polymorphism may lead to disease.

The number of candidate gene studies in hematopoietic cell transplantation far exceeds that for GWAS. Among candidate gene studies, the heterogeneity of results has prompted recent analyses of independent cohorts. Many factors could influence the associations of a given variant to a clinical endpoint, including patient and donor characteristics such as age, sex, parity of female individuals, race and ethnicity, to name a few. Additionally, heterogeneity with respect to the intensity of the conditioning regimen, the use of T cell depletion for GVHD prevention, and post-transplant cellular therapy. Hence, interpretation of the results of retrospective genetic association studies ideally includes careful consideration of the characteristics of the study population. Equally important for candidate gene studies, is the oftentimes lack of uniformity of the specific polymorphisms within the candidate gene that are genotyped. For example, a regulatory region of interest may be polymorphic, with several different sites of variation; when different sites are typed and analyzed, the conclusions may or may not pertain to the untyped sites. These methodologic differences add to the complexity of interpreting results across studies.
Candidate MHC gene studies in HCT

The most extensively studied non-HLA MHC genes to date are involved in inflammation, the stress response, innate immunity, and NK-T cell allore cognition (Table 2): TNF-α,43-49 MICA,50-52 HLA-E,48;53-58 HLA-G59-61 and HSPA1L.62

**TNF-α.** The TNF block is a 7 kb region within class III that encodes TNF-α, TNF-β, lymphocytoxin (LTA)-α, leucocyte-specific transcript (LST)-1 and allograft inflammatory factor (AIF).28;63;64 These genes share coordinated function as inflammatory mediators.65 TNF-α has been the most comprehensively studied candidate gene,43-49 and its sequence diversity and haplotype organization have been defined.48;49;66;67

TNF-α is a pro-inflammatory Th1 cytokine involved in different phases of GVHD.68 In phase I, TNF-α is associated with tissue injury and the initiation of the cytokine storm; the activation of antigen-presenting cells enhance alloantigen presentation and induce inflammatory cytokines leading to the recruitment of the effector cells that ultimately mediate damage of the target organs. In phase III, TNF-α and IL1 participate in the cytokine storm.68 TNF-α sequence variation is associated with GVHD, transplant-related mortality (TRM) and survival (Table 2).43;45-49;69 Sequence variation at the -308 (A/G) and the -863 (A/C) positions in the 5′ promoter account for differential expression of TNF-α which correlate with risk of acute GVHD.70

**MICA.** MICA is polymorphic,71 has strong haplotypic associations to HLA-B,72;73 is inducible by stress,74 and is a ligand for activating NKG2D receptors.75 A potential role for MICA in GVHD might involve T or NK-mediated pathways, and/or inflammatory and stress mediators. One of the earliest investigations of the MICA region demonstrated the association of block matching with improved survival after HCT,76 paving the way for additional studies of this locus.50;52 Methionine (M) or valine (V) at residue 129 correlates with the strength of binding of MICA to its NKG2D receptor.77 Patients with 129V alleles have an increased risk of chronic GVHD after sibling donor HCT.50 Furthermore, patient-donor MICA
mismatching has correlated with an increased risk of II-IV acute GVHD, particularly of the gastrointestinal tract after HLA-matched and HLA-mismatched HCT. In HLA-matched unrelated donor HCT, MICA*008-positive patients had less acute GVHD. These data collectively suggest that the presence of specific MICA phenotypes may serve as antigenic targets that lead to GVHD, or that influence the interaction of MICA to its receptors.

**HLA-E.** HLA-E plays a role in innate and adaptive immunity by functioning as a ligand for inhibitory CD94/NKG2A receptors and by presenting peptides. The two major alleles, E*01:01 and 01:03, differ from one another at residue 107 (glycine, G, versus arginine, R) of the α2 heavy chain. HLA-E alleles have different levels of cell surface expression, with HLA-E*01:01 expressed at lower levels than E*01:03. Homozygosity for E*01:03 was associated with a lower risk of acute GVHD and TRM, and lower TRM and improved survival. In other studies, associations of HLA-E with acute GVHD have been heterogeneous. Patient-donor mismatching for the G107R substitution defined by rs126457 is associated with lower survival.

**HLA-G.** This non-classical class I gene has several unique features including a 14 bp insertion or deletion located in its 3’ untranslated region (UTR) and at least 15 isoforms that are generated through alternative splicing. Because of its multiple allotypes, HLA-G has been studied as a potential source of donor-recipient disparity in transplantation. The 14 basepair insertion/deletion was associated with acute GVHD in some but not other studies, although homozygosity for the 14 bp deletion did correlate with lower survival and disease-free survival.

**GWAS**

GWAS studies of GVHD in transplantation ideally require upwards of 4000 transplants for sufficient power across a range of odds ratios. GWAS of HLA-matched sibling or unrelated donor transplant populations permit the risks associated with non-chromosome 6 genes to be measured without the
effects of HLA mismatching. However, HLA-mismatched populations provide unique information on the additive or synergistic effects of HLA mismatching with variation both within and outside of the MHC.

The Japan Marrow Donor Program (JMDP) has conducted the largest GWAS analysis of transplant outcomes to date. The first stage assumed no HLA restriction of SNPs and identified rs6937034 near HLA-DP as a marker of grades II-IV acute GVHD. The second phase tested the hypothesis that SNPs function as HLA-restricted minor histocompatibility antigens; rs17473423 (chromosome 12) associated with risk of grades III-IV acute GVHD in HLA-A*24-B*52-C*12-DRB1*15-DQB1*06-positive transplants. Furthermore, rs9657655 (chromosome 9) was a marker for HLA-A*33:03-B*44:03-C*14:03-positive transplants, and 4 additional loci were identified in association with HLA-DQB1*05:01, C*01:02, B*52:01 and C*12:02. These data suggest a role for non-HLA polymorphisms as minor histocompatibility antigens restricted by HLA on extended Japanese HLA haplotypes. The investigators further identified 2 SNPs associated with patient genotype (chromosome 12 rs5998746 and chromosome 18 rs11873016); no GVHD-associated SNPs were identified in the donor genotype model.

**MHC region SNP discovery.** Whereas GWAS studies survey polymorphisms across the entire human genome, a focused examination of dense SNPs within the MHC provides a first-hand look at untyped variation carried on HLA haplotypes of the patient and donor. As described above, the rationale for such “MHC-was” studies is based on the concept that unrelated individuals with the same tissue type may have different HLA haplotypes that give rise to intergenic sequence variation (Figure 2) including very common HLA haplotypes (Figure 3). Focused SNP mapping of the MHC is applicable not only to HLA-mismatched but also to HLA-matched transplant pairs. Like GWAS, there are no *a priori* regions or genes that are more or less likely to cause GVHD.
SNPs are proxies for genes that cause disease. In transplantation, risks might be conferred by patient genotypes, donor genotypes and/or patient-donor mismatching. The patient and the donor genotype models test the hypothesis that the presence of specific variants affect GVHD, relapse or survival. This concept is akin to classical disease-association mapping. SNPs may be proxies for regulatory variants that affect the transcription and translation of the gene or reside a distance away from the target gene(s), including genes that reside on other chromosomes. Alternatively, SNPs may result in non-synonymous coding substitutions that affect the structure and function of proteins, ligand-receptor interactions, or possibly interactions with other molecules in shared pathways. The most conceptually straightforward model is that of patient-donor mismatching, in which the patient’s SNP allele differs from the donor’s. This model indirectly tests whether there are donor polymorphisms recognized by the patient (donor AG versus patient AA or GG), patient variation recognized by the donor (patient AG versus donor AA or GG), or both (patient AA versus donor GG). The identification of any SNP in any model sheds light on what candidate gene(s) are involved in GVHD, and the potential functional consequences of DNA variation.

Two recent studies demonstrate that the MHC harbors clinically relevant variation responsible for GVHD and support the concept that GVHD is a polygenic disease. A discovery-validation study of over 4200 HLA-A,C,B,DRB1,DQB1-matched (“10/10”) unrelated donor transplants identified two SNPs associated with clinical outcome. Donor mismatching at rs887464 (donor AG versus patient AA) was associated with lower survival and disease-free survival. Interestingly, rs887464 is a risk marker for type 1 diabetes and a putative expression quantitative locus (eQTL). These data suggest that GVHD might share common pathways with other autoimmune diseases, and offer new approaches for fine mapping the gene(s) that are involved in GVHD.
In HLA-matched unrelated donor HCT, patient-donor mismatching at rs2281389 (patient AG versus donor GG) increased the risks of grades II-IV acute GVHD.\textsuperscript{40} SNP rs2281389 resides centromeric to the HLA-DPB1 3’ UTR and shows complex haplotype linkage to HLA-DP; its biological implications are yet to be elucidated. Both rs887464 and rs2281389 were identified in the mismatch model, suggesting that HLA-matched patients and donors have untyped disparity outside of the classical HLA loci that can increase risks of GVHD and mortality.

The evaluation of HLA-mismatched unrelated donor HCT introduces different challenges than the HLA-matched setting. The hypothesis that deleterious HLA haplotype-linked polymorphisms can synergize with HLA mismatches \textit{per se}, has recently been tested in a retrospective study of 2628 transplants with a single HLA-A,C,B,DRB1, or DQB1 (“9/10”) mismatch.\textsuperscript{41} Untyped SNPs “hitch-hike” with the HLA alleles and the rate of SNP mismatching may be particularly high near the HLA-mismatched locus. A survey of over 1100 SNPs identified 12 candidate markers for future validation. Risks of acute and chronic GVHD, relapse, and mortality increased with increasing numbers of unfavorable SNPs, and parallel observations from GWAS mapping of complex autoimmune diseases.\textsuperscript{87} Finally, SNPs define HLA-SNP haplotypes. The most common Caucasian haplotype worldwide, HLA- A*01:01-B*08:01-DRB1*03:01 (7.4% frequency\textsuperscript{88}) displays variation at rs429916, a marker for survival.\textsuperscript{41} These data suggest that knowledge of the patient’s and donor’s haplotypes not only for the specific HLA alleles, but also haplotype-linked SNPs, are needed to fully understand the risks to individual patients after HCT. Future studies of SNPs and HLA haplotypes will be needed to define the extent to which the SNP-HLA haplotypes are shared among ethnically diverse populations, and those that are unique.

\textit{Impact of HLA mismatching}. If risks can stem from non-HLA variants that are carried on HLA haplotypes, then cumulative effects of both HLA and non-HLA variation may be significant. To test this hypothesis, the risks associated with a single HLA-A, B, DRB1, DQB1 allele or antigen, or single HLA-C allele
mismatch relative to a single HLA-C antigen mismatch were defined in models that adjusted for the contribution of the 12 SNP genotypes or mismatching. The risks associated with allele or antigen mismatching at each locus were similar to HLA-C antigen mismatches with the exception of HLA-DQB1 mismatches which, as a group, were less risky. Although these data suggest that, in general, any HLA mismatch outside of HLA-DQ is deleterious, some HLA mismatch combinations are more likely to encode favorable SNPs than others. It is intriguing to speculate that the more favorable outcomes of HLA-C allele mismatches might be in part contributed by the hitchhiking of favorable SNPs on the patient’s and donor’s haplotypes. That the permisivity of HLA mismatches might be defined by synergistic effects between HLA coding region variation and haplotype-linked variation, introduces new hypothesis that will require large cohorts of SNP-typed transplants for testing in the future.

**Comparison of HLA 10/10 with 9/10 unrelated donor HCT.** The identification of clinically relevant MHC region SNPs permits a formal comparison of survival after 10/10-matched and 9/10-matched unrelated donor HCT (Figure 4). Survival after 9/10 HCT with 0–1 unfavorable SNP is similar to that after 10/10 HCT with favorable SNPs, and both are superior to transplantation from 9/10 donors with 2 or more unfavorable SNPs or from 10/10 donors with an unfavorable SNP. Intermediate risks are observed after 10/10 transplantation with a “neutral” SNP. These data shed new light on clinical settings where the use of a 9/10 donor might yield comparable outcomes to that of a 10/10 donor and suggest that prospective patient and donor evaluation for SNPs may aid the selection of donors with optimal MHC SNPs.

**After SNP Discovery: Replication and Fine Mapping**

Candidate SNPs that are replicated for the same model (genotype or mismatch) and for the same clinical endpoint (GVHD, relapse, survival) represent *bona fide* tags for the causative genes. Fine mapping involves focused studies to narrow down the regions harboring the causative gene(s) for comprehensive
analysis of sequence variation. Often, the physical proximity of the SNP to a gene is used to initiate direct examination of gene polymorphism and function. When the SNP itself defines a non-synonymous substitution, the protein-coding gene is a prime suspect. In gene-dense regions, sequencing several hundred kb 5’ and 3’ of the target gene may be necessary, depending on local block structure.

Often the SNP associations have biological plausibility because they define known genes with known immunological function. In some instances, the SNP may be shared among different diseases; for example, Crohn’s colitis, multiple sclerosis, psoriasis, rheumatoid arthritis, systemic lupus erythematosus and type 1 diabetes share up to 44% of SNP associations. Shared SNPs provide an approach for fine mapping across different diseases and phenotypes.

Dense SNPs panels and “deep sequencing” with next-generation sequencing (NGS) are two of the most common methods used for fine mapping. New genome-wide approaches that facilitate the mapping of SNPs to structural alleles offer promising methods for analyzing many genomes concurrently. When haplotype block structure varies among populations, SNPs that are shared among different haplotypes may be helpful for ruling in or out candidate genes (Figure 1). In some instances, it may not be feasible to identify a single causative variant, but rather haplotypes of tightly-linked markers.

**Future Applications to Clinical Practice**

Mapping GVHD-associated SNPs in a retrospective population of paired patients and donors provides information on the frequency of SNPs among donors who were selected for transplantation using criteria other than SNPs. The likelihood that a given patient has HLA-matched donors who also have favorable SNPs can be estimated by examining SNP frequencies in a pool of individuals who share the same tissue type. In a retrospective study of 230 patients referred to a single center for initiation of an unrelated donor search, patients for whom at least 2 HLA-matched donors were identified were each
genotyped for the GVHD-associated rs2281389. Among 131 patients with two donors each, 82% had at least one rs2281389-matched donor and 48% had more than one SNP-matched donor. Among 17 patients with 6 HLA-matched donors each, 94% had at least one rs2281389-matched donor and 88% had more than one SNP-matched donor. These data suggest that rs2281389-matched donors can be identified among HLA-matched donors, and that the likelihood of finding such donors increases with more donors tested per patient. Similar frequency data are needed in the future for validated SNPs, as the frequency of donors with favorable SNPs may depend on their haplotypes.  

The translation of SNP data to clinical practice in the future will depend on what kind of polymorphism should be tested, whether quantitative protein expression are robust indicators of GVHD risk, and whether tissue-specific patterns of expression are clinically relevant. The most straightforward translation of validated SNPs to the clinical setting requires efficient and highly robust read-out assays. One available method for SNP genotyping is the TaqMan® 5’ nuclease allelic discrimination assay, an efficient, low cost method where data interpretation is unambiguous because only 2 alleles are possible (Figure 5A).  

Direct sequencing is also feasible and useful for selected applications (Figure 5B). When clear associations between GVHD and expression variants are defined, then functional assays may ultimately require assessment of individual alleles, genes, haplotypes, cellular compartments and tissues.  

Future needs to facilitate the translation of association-mapping data for clinical purposes will require careful assessment of the relative importance of genetic and non-genetic factors to outcome. In allogeneic transplantation where disease diagnosis and stage, patient/donor age, sex, ABO blood type, CMV serostatus may all be important risk factors, the relative importance of each of these variables to genetic factors remains an important research objective. This information will facilitate the prospective selection of potential transplant donors especially when patients have a choice among several donors.
Conclusions

The MHC remains a model system for understanding the immunogenetic basis of GVHD and transplant outcomes. GVHD is a polygenic disease where risk is contributed by the MHC haplotypes of the patient and the transplant donor. A new paradigm is emerging that includes consideration for both HLA coding and HLA haplotype-linked variation as important factors in unrelated donor HCT. The identification of specific novel variants within the MHC that play a role in GVHD underscores the need for more complete information on MHC haplotype diversity and organization of sequence variation in ethnically diverse populations.

Acknowledgments

The author thanks Dr. Mari Malkki and Dr. Tao Wang for assistance with figures, and Ms. Courtney Preusse for preparation of the manuscript. Dr. Petersdorf is supported by National Institutes of Health grants AI069197 from the National Institute of Allergy and Infectious Diseases, and CA100019 and CA18029 from the National Cancer Institute.

Authorship

EWP wrote the manuscript.

Conflict of Interest Disclosure

The author declares no competing financial interests.
References


Ref Type: Online Source


Ref Type: Online Source


65. Mulcahy H, O'Rourke KP, Adams C, Molloy MG, O'Gara F. LST1 and NCR3 expression in autoimmune inflammation and in response to IFN-gamma, LPS and microbial infection. Immunogenetics 2006;57(12):893-903.


84. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661-678.


Ref Type: Online Source


Ref Type: Online Source


Table 1. Terms commonly used in HLA immunogenetics and disease mapping.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association mapping</td>
<td>A method for linking clinical characteristics to genetic traits.</td>
</tr>
<tr>
<td>Block</td>
<td>The physical region of DNA encoding two or more SNPs or genes in strong positive linkage disequilibrium with one another, such that any one polymorphism is a marker for all the others in that region.</td>
</tr>
<tr>
<td>Candidate gene</td>
<td>A gene that has been identified through discovery methods, but not yet validated.</td>
</tr>
<tr>
<td>Deep sequencing</td>
<td>Sequencing approaches that incorporate multi-fold coverage of nucleotide positions.</td>
</tr>
<tr>
<td>Direct sequencing</td>
<td>PCR-based sequencing approaches that that use the target DNA strand for amplification and sequencing.</td>
</tr>
<tr>
<td>Discovery-Validation</td>
<td>The use of two independent cohorts, where one cohort is used to “discover” new trait-associated markers (“candidate” SNPs or genes), and the second cohort is used to “validate” the candidates.</td>
</tr>
<tr>
<td>study design</td>
<td></td>
</tr>
<tr>
<td>Expression quantitative locus</td>
<td>A genetic locus that regulates the level of expression of mRNAs and proteins.</td>
</tr>
<tr>
<td>Fine mapping</td>
<td>Approaches that use dense SNPs or sequencing to narrow down the region(s) that most likely harbor the causative gene(s).</td>
</tr>
<tr>
<td>Genome-wide association study</td>
<td>The examination of variants across the human genome, for association to specific traits or diseases.</td>
</tr>
<tr>
<td>(GWAS)</td>
<td></td>
</tr>
<tr>
<td>Genotype</td>
<td>An individual’s DNA sequence (gene, allele, SNP), to be distinguished from phenotype, the trait of the expressed protein, or a clinical characteristic.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Haplotype</td>
<td>Markers that travel together on the same chromosome 6 strand. Example: HLA-A1-B8-DR3</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>Two different HLA alleles/antigens at the HLA locus. Example: HLA-A1,2</td>
</tr>
<tr>
<td>HLA</td>
<td>HU-1 (human-1 antigen) and LA (leukocyte antigen), antigens that govern tissue type and histocompatibility.</td>
</tr>
<tr>
<td>HLA Class I</td>
<td>The telomeric (further from the centromere) region of the MHC encoding HLA genes. The Class I region contains HLA-A, HLA-B, HLA-C “classical class I genes”) and HLA-E, F, G; MICA, MICB (“non-classical class I”).</td>
</tr>
<tr>
<td>HLA Class II</td>
<td>The centromeric (closer to the centromere) region of the MHC encoding HLA genes. The Class II region contains HLA-DR, HLA-DQ, HLA-DP.</td>
</tr>
<tr>
<td>HLA Class III</td>
<td>The region of the MHC mapping between class I HLA-B and class II HLA-DR. The Class III region contains Tumor Necrosis Factor (TNF) and the complement pathway genes.</td>
</tr>
<tr>
<td>Homozygous</td>
<td>Two copies of the same polymorphism or allele. Example: SNP genotype AA, TT, CC or GG; HLA genotype HLA-A*02:01,02:01</td>
</tr>
<tr>
<td>Imputation (of SNPs)</td>
<td>Methods that employ linkage disequilibrium to infer SNPs that have not been directly genotyped.</td>
</tr>
<tr>
<td>Linkage Disequilibrium (LD)</td>
<td>The non-random association of alleles or antigens. Example: HLA-B8 with HLA-DR3.</td>
</tr>
<tr>
<td>MHC</td>
<td>Major Histocompatibility Complex, a 7.6 megabase region of chromosome 6p21.3 that encodes the histocompatibility genes important in transplantation.</td>
</tr>
<tr>
<td>Phenotype</td>
<td>A trait. Phenotype can refer to the expressed protein (HLA-A2), or a clinical</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>characteristic</td>
<td></td>
</tr>
<tr>
<td>Proxy</td>
<td>A genetic marker that can be used in lieu of another markers, usually in high linkage disequilibrium.</td>
</tr>
<tr>
<td>Single Nucleotide Polymorphism (SNP)</td>
<td>The simplest form of human genetic variation. A SNP is a biallelic nucleotide change. Example: A vs C.</td>
</tr>
<tr>
<td>TaqMan</td>
<td>A real-time 5’ nuclease method for genotyping biallelic SNPs.</td>
</tr>
<tr>
<td>TagSNP</td>
<td>A SNP that is in 100% LD with another SNP(s) so that they serve as proxies for one another.</td>
</tr>
</tbody>
</table>
Table 2. Summary of MHC region polymorphisms associated with clinical outcome after allogeneic HCT.  

(A) Candidate gene studies. In these studies, the specific MHC region gene or allele or SNP was tested individually. Genes are presented in chromosomal order.  

(B) MHC region SNPs identified by GWAS of the entire genome. In this study, GWAS was performed using a commercial platform.  

(C) SNPs identified by “MHC-was” SNPs survey. In these studies, SNPs genotyping was performed using commercial platforms. SNPs are presented in chromosomal order.

<table>
<thead>
<tr>
<th>A. Candidate Gene/Polymorphism</th>
<th>Model</th>
<th>Study Population</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-G 14 bp insertion/deletion in 3' UTR</td>
<td>Homozygosity for del/del</td>
<td>53 HLA matched URD</td>
<td>Increased acute GVHD</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Homozygosity for ins/ins</td>
<td>47 HLA matched URD, siblings, cord blood</td>
<td>Lower survival and DFS</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>157 HLA identical siblings</td>
<td></td>
<td>Increased severe acute GVHD</td>
<td>50</td>
</tr>
<tr>
<td>HLA-E E<em>01:01 and E</em>01:03 alleles</td>
<td>Patient and donor homozygosity for E*01:03</td>
<td>187 HLA identical siblings (ref 54); 33 HLA identical siblings; 12 HLA matched URD; 11 haploidentical mismatched relateds (ref 57)</td>
<td>Lower grades III-IV acute GVHD, lower TRM, improved survival</td>
<td>54, 57</td>
</tr>
<tr>
<td></td>
<td>Patient E*01:03 homozygosity</td>
<td>67 HLA matched siblings; 16 HLA matched URD</td>
<td>Higher survival; higher DFS; lower TRM</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Donor E*01:03 homozygosity</td>
<td>124 HLA matched URD</td>
<td>Lower grades II-IV acute GVHD; lower chronic GVHD; higher relapse</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116 HLA matched URD</td>
<td>No correlation with clinical outcome</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>E*01:01/01:03 mismatch</td>
<td>116 HLA matched URD</td>
<td>No correlation with clinical outcome</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Rs1264457 residue 128 arg (R) versus gly (G) mismatch</td>
<td>922 HLA matched and mismatched URD</td>
<td>Lower survival</td>
<td>48</td>
</tr>
<tr>
<td>MICA</td>
<td>Residue 129 valine (V) versus methionine (M) patients homozygous for 129V</td>
<td>211 HLA matched siblings</td>
<td>Higher chronic GVHD</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Mismatch</td>
<td>236 HLA matched and mismatched URD</td>
<td>Higher grades II-IV acute GVHD, notably gastrointestinal GVHD</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38 HLA matched URD</td>
<td>No association</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Binding affinity for M129V</td>
<td>38 HLA matched URD</td>
<td>No association</td>
<td>52</td>
</tr>
</tbody>
</table>
### B. GWAS Chromosome 6 Signals/Closest Locus

<table>
<thead>
<tr>
<th>Model</th>
<th>Study Population</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DP</td>
<td>Mismatch</td>
<td>1598 HLA matched URD</td>
<td>Higher grades II-IV acute GVHD</td>
</tr>
</tbody>
</table>

### C. "MHC-was"/Closest Locus

<table>
<thead>
<tr>
<th>Model</th>
<th>Study Population</th>
<th>Effect</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIM27</td>
<td>Mismatch</td>
<td>2628 HLA mismatched URD</td>
<td>Higher grades III-IV acute GVHD</td>
</tr>
<tr>
<td>HLA-A</td>
<td>Mismatch (donor-versus-patient)</td>
<td>2628 HLA mismatched URD</td>
<td>Higher chronic GVHD</td>
</tr>
<tr>
<td>PSOR1C3/POU5F1</td>
<td>Mismatch (patient-versus-donor)</td>
<td>4205 HLA matched URD</td>
<td>Lower survival</td>
</tr>
<tr>
<td>HCP5</td>
<td>Donor genotype GG</td>
<td>2628 HLA mismatched URD</td>
<td>Higher relapse</td>
</tr>
<tr>
<td>LTA</td>
<td>Patient genotype AT</td>
<td>2628 HLA mismatched URD</td>
<td>Lower TRM</td>
</tr>
<tr>
<td>SNP</td>
<td>Gene</td>
<td>Mismatch (donor-versus-patient)</td>
<td>Genotype/Allele(s)</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>rs2242656</td>
<td>BAG6</td>
<td>Mismatch</td>
<td></td>
</tr>
<tr>
<td>rs2075800</td>
<td>HSPA1L</td>
<td>Patient genotype GG</td>
<td></td>
</tr>
<tr>
<td>rs3830076</td>
<td>FKBP1</td>
<td>Mismatch</td>
<td></td>
</tr>
<tr>
<td>rs394657</td>
<td>NOTCH4</td>
<td>Donor genotype AA</td>
<td></td>
</tr>
<tr>
<td>rs2071479</td>
<td>HLA-DOB</td>
<td>Mismatch</td>
<td></td>
</tr>
<tr>
<td>rs429916</td>
<td>HLA-DOA</td>
<td>Patient genotype AA</td>
<td></td>
</tr>
<tr>
<td>Rs2281389</td>
<td>HLA-DPB1 3’UTR</td>
<td>Mismatch</td>
<td></td>
</tr>
<tr>
<td>rs986522</td>
<td>COL11A2</td>
<td>Donor genotype CG</td>
<td></td>
</tr>
<tr>
<td>rs107822</td>
<td>RING1A</td>
<td>Mismatch</td>
<td></td>
</tr>
</tbody>
</table>

UTR, untranslated region; del, deletion; ins, insertion; DFS, disease-free survival; TRM, transplant-related mortality

*SNP was imputed.

†SNP was validated in a discovery-validation study design.

‡SNP was identified in a discovery set; validation is pending.
Figure Legends

Figure 1. SNP haplotypes in Caucasian, African, Asian and Admixed populations from the 1000Genomes datafile. Continental populations may differ for the minor allele frequencies of SNPs, the strength of LD, and recombination, factors that define haplotype block structure within the MHC. Illustrated is the 20.8 kb TNF block consisting of the LTA-TNF-LST-NCR loci for samples from the 1000Genomes Project Consortium for individuals of diverse backgrounds: (A) European Caucasian, n = 379; (B) African, n = 246; (C) Asian, n = 286 and (D) Admixed American (includes self-described Hispanic), n = 181. The intensity of the red color in each pairwise SNP comparison is a measure of the strength of LD as measured by D'. Bright red boxes indicate a D’ of 1 (100% linkage) and the gradation of color to white indicates progressively smaller D’ values indicative of weaker LD. The different block structure and SNP haplotypes among these ethnically diverse populations may facilitate the fine mapping of SNP association signals within this region of the MHC.

Figure 2. HLA-mismatched patient-unrelated donors may encode variation in regions between classical HLA loci. HLA-matched patients and donors may encode different sequences in between HLA-A, C, B, DRB1 and DQB1 and these variants might be a source of disparity that gives rise to GVHD after HCT. In HLA-mismatched HCT, the frequency of patient-donor sequence differences and the extent of the MHC that is different may be pronounced and the effects of such untyped variation may be additive to that from the patient-donor HLA mismatch.

Figure 3. Comparison of single nucleotide polymorphisms (SNPs) encoded on HLA-A1-B8-DR3, HLA-A3-B7-DR15 and HLA-A2-B7-DR15 haplotypes. SNP variation encoded on the three most common Caucasian haplotypes, HLA-A1-B8-DR3, A3-B7-DR15 and A2-B7-DR15, are a source of patient-donor variation. Data are shown for 28 HLA-A1-B8-DR3 chromosomes, 4 HLA-A3-B7-DR15 chromosomes, and 4 HLA-A2-B7-DR15 chromosomes. HLA-A1-B8-DR3 chromosomes share conserved sequences indicated by
the green bar; SNP substitutions are indicated by vertical black bars and black rectangles. HLA-matched and HLA-mismatched patients and donors may differ at SNP positions across the 7 Mb of the MHC. For example, a patient with the HLA-A1,3-B7,8-DR15,3 tissue type and donor with the HLA-A1,2-B7,8-DR15,3 tissue type (HLA-A mismatched) will have disparities at multiple nucleotide positions. These untyped SNPs may be a source of variation associated with increased risk of GVHD and transplant-related mortality.41

**Figure 4.** Kaplan-Meier probability of survival after HLA 10/10-matched and HLA 9/10-matched unrelated donor HCT according to the number of favorable SNPs. A total of 4205 HLA 10/10 allele-matched40 and 2628 HLA 9/10 single locus mismatched41 unrelated donor-patient pairs were examined for SNP associations with survival: 10/10-matched and favorable rs887464 (SNP defined in reference 40) (solid red curve; n = 126); 9/10-matched and 0 – 1 unfavorable SNP (SNPs defined in reference 41) (solid blue curve; n = 73); 10/10-matched and neutral rs887464 (solid black curve; n = 3521); 9/10-matched and more than 1 unfavorable SNP (hatched blue curve; n = 1849), and 10/10-matched and unfavorable rs887464 (hatched red curve; n = 92). Overall survival after 10/10-matched unrelated donor HCT with favorable SNP is similar to that after 9/10-matched unrelated HCT with 0 - 1 unfavorable SNP. The negative effect of cumulative unfavorable SNPs is additive to HLA mismatch-associated risks.40,41

**Figure 5.** (A) TaqMan® assay for typing biallelic single nucleotide polymorphisms (SNPs). Actual TaqMan® typing result for the G → A rs2075800 SNP (E → K at residue 602) of HSPA1L, a marker for GVHD in HLA-mismatched unrelated donor HCT.41 The assay is a real-time PCR method in which the wild type allele is PCR-amplified separately from the alternative allele using fluorescent allele-specific probes. Shown is the allelic discrimination plot in reverse orientation: CC homozygous (blue circles; reverse orientation for GG); TT homozygous (red circles; reverse orientation for AA); CT heterozygous (green circles). Fluorescent signals are interpreted with the use of software that detects the presence of one
dye (either homozygous wild type represented by one color or homozygous alternative allele represented by a different color) or two dyes (heterozygous wild type and alternative allele). (B)

**Sequence-based typing of SNPs.** Actual sequence chromatogram (reverse orientation) of a MICA*008,018-heterozygous sample illustrating the C (blue) \( \rightarrow \) T (red) substitution corresponding to valine \( \rightarrow \) methionine at residue 129 (asterisk). In this sample, MICA*008 encodes valine and MICA*018 encodes methionine. Methods previously described in reference 96.
Figure 2

Patient Haplotypes

Potential Donor Haplotypes

Haplotype-matched

Haplotype-mismatched
Figure 3

A1-B8-DR3

A3-B7-DR15

A2-B7-DR15

A

C B

DRB1 DQB1 DPB1
The major histocompatibility complex: a model for understanding graft-versus-host disease

Effie W. Petersdorf