Population Frequencies Determined by Next-generation Sequencing Provide Strategies for Prospective HLA Epitope Matching for Transplantation

Jenny Tran
University of British Columbia https://orcid.org/0000-0002-4777-165X
Oliver Gunther
Gunther Analytics
Karen Sherwood
Vancouver General Hospital
Franz Fenninger
University of British Columbia
Lenka Allan
University of British Columbia
James Lan
University of British Columbia
Ruth Sapir-Pichhadze
McGill University and MU-HRI
Rene Duquesnoy
University of Pittsburgh Medical Center
Frans Claas
Leids Universitair Medisch Centrum
Steven Marsh
Anthony Nolan Research Institute https://orcid.org/0000-0003-2855-4120
W. Robert McMaster
University of British Columbia
Paul Keown (paul.keown@ubc.ca)
University of British Columbia

Article

Keywords: HLA epitopes, antibody-verified, HLA alleles, next-generation sequencing, biomarkers, kidney transplantation

Posted Date: September 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-50893/v1

POPULATION FREQUENCIES DETERMINED BY NEXT-GENERATION SEQUENCING PROVIDE STRATEGIES FOR PROSPECTIVE HLA EPITOPE MATCHING FOR TRANSPLANTATION

Tran JN[1], Günther OP [5], Sherwood KR[1], Fenninger F[1], Allan LL[1], Lan J[1,2], Sapir-Pichhadze R[6], Duquesnoy R[7], Claas F[8], Marsh SGE[9], McMaster R[3,4], *Keown PA[1,2,4], for the Genome Canada Transplant Consortium

Departments of: [1]Pathology and Laboratory Medicine, [2]Medicine, and [3]Medical Genetics, and the [4]Infection & Immunity Research Centre, University of British Columbia, Vancouver, Canada; [5]Günther Analytics, Vancouver, Canada; [6]Department of Medicine, McGill University and MU-HRI, Montreal, Canada; [7]Department of Pathology, University of Pittsburgh, Pennsylvania, USA; [8]Department of Immunohematology and Blood Transfusion, University of Leiden, Netherlands; [9]Anthony Nolan Research Institute & UCL Cancer Institute, Royal Free Campus, London.

Genome Canada Transplant Consortium: Members of the research team are listed in the appendix. Research is supported by Genome Canada, Genome BC, Genome Quebec, Genome Alberta and the Canadian Institutes for Health Research, and funded by awards LSARP 273AMR and GP1-155871.

Keywords: HLA epitopes, antibody-verified, HLA alleles, next-generation sequencing, biomarkers, kidney transplantation

Word Count: Abstract: 147 Text: 4,974 Methods: 829 Tables: 1 Figures: 9

Corresponding Author*: Paul A. Keown
Director, Immunology
University of British Columbia
855 W 12th Ave, Vancouver, BC V5Z 1M9
Tel: 604-875-4111, Ext. 54393
paul.keown@ ubc.ca

ABSTRACT

Compatibility for human leukocyte antigen (HLA) genes between transplant donors and recipients improves graft survival but prospective matching is rarely performed due to the vast heterogeneity of this gene complex. To reduce complexity, we have combined next-generation sequencing and *in silico* mapping to determine population frequencies and matching probabilities of 150 antibody-binding eplets across all 11 classical HLA genes in 2000 ethnically heterogeneous renal patients and donors. We show that eplets are more common and more uniformly distributed between donors and recipients than the respective HLA isoforms. Simulation of targeted eplet matching shows that a high degree of overall compatibility, and perfect identity at the clinically important HLA class II loci, can be obtained within a patient waiting list of approximately 250 subjects. Internal epitope-based allocation is thus feasible for most major renal transplant programs, while regional or national sharing may be required for other solid organs.

INTRODUCTION

Transplantation is the treatment of choice for irreversible renal failure, offering superior survival, quality of life and economic costs compared to alternative options [1,2]. But despite superb initial success (1 year kidney graft survival often exceeds 95%), many grafts fail within the first decade [3]. While several factors may jeopardize the transplanted organ, graft rejection remains the overwhelming cause of failure [4,5] and antibody-mediated rejection (AMR) is the most serious and destructive form of this injury. It may occur early or late in the transplant course, presenting a spectrum that ranges from common acute and fulminant graft injury to chronic and progressive graft destruction. There is currently no effective therapy to reverse AMR, so that measures to prevent this complication by reducing immunogenicity and modulating immunity are vital.

Human Leucocyte Antigen (HLA) genes located within the Major Histocompatibility Complex on the human 6th chromosome are the most polymorphic in the human genome with over 25,000 alleles now identified [6]. The classical polymorphic members of this family are divided into three class I genes (HLA-A/B/C) and eight class II genes (HLA-DRB1/3/4/5, DQA1/B1, DPA1/B1). These genes code for an array of HLA protein isoforms expressed on nucleated cells which are highly immunogenic and are known to be the principal transplantation antigens and primary targets of graft rejection [7]. Compatibility for Human Leukocyte Antigen (HLA) genes between donor organs and graft recipients ensures excellent outcome in live donor transplantation, and improves transplant survival in deceased donor transplantation [8,9], but is difficult to achieve due to the heterogeneity of this gene complex.

Discrete motifs on these proteins are central to both antigen recognition and response. Structural epitopes, clusters of amino acids on the surface of the HLA protein isoforms that are accessible to

and are bound by antibody, encompass smaller eplets lying in a 3 Angstrom radius containing at least one polymorphic amino acid, that interact directly with the antibody paratope [10]. The quantitative mismatch of these eplets between donor and recipient provides an index of the risk of rejection, though not all epitope mismatches may be of equal importance and those occurring at certain HLA class II gene loci may be particularly critical. For example, Wiebe et al. have shown that HLA-DRB1 and -DQB1 mismatches are independent predictors of *de novo* class II donor-specific antibody development [11], and a nested case-control study found that the odds ratio for transplant glomerulopathy increases incrementally with increasing HLA-DR + DQ eplet mismatches [12].

Quantitative epitope mismatch analysis is normally proposed as a post-hoc measure to predict rejection risk and to guide immunosuppressive treatment. However, prospective use of eplet matching to guide recipient selection offers a novel method to actively reduce donor immunogenicity, and the limited number of eplets may enable efficient matching. To determine whether this prospective strategy is clinically feasible for patients awaiting transplantation requires precise population data on donor and recipient frequency distributions. Here we present the first large-scale data comparing human HLA allele and eplet frequencies defined by high-resolution next-generation sequencing (NGS) in a heterogeneous population, focusing on antibody-verified HLA eplets in light of their proven target role. We describe combinatorial epitypes comprising the array of eplets expressed by the HLA isoforms on each individual donor and recipient and their population clustering, and model the probabilities of achieving eplet identity at all or individual HLA gene loci to confirm the feasibility of prospective eplet matching within a national transplant program.

RESULTS

Transplant patients and donors

A total of 2,000 subjects from the BC renal transplant program had full NGS sequence data at all 11 allelic HLA loci for the study. Of these, 154 subjects expressed alleles that were not yet present in the HLAMatchmaker database (47 alleles, average carrier rate ≤0.18%, Supplemental Figure 1) and were excluded from this analysis. Carrier rates of other HLA alleles in these subjects were otherwise comparable to the overall population (Supplemental Figure 1). The remaining 1846 subjects included 1049 patients with kidney failure and 797 kidney donors. Patient and donor groups were 62% and 52% males, with a mean age of 56 and 48 years, respectively. Four subgroups were included to control for bias, comprising patients prior to (n=611) or post-transplant (n=438), and deceased (n=243) or living donors (n=554).

Visualization of eplets on an HLA protein

A freely accessible protein visualization software, Chimera, was used to visualize eplets on HLA proteins [13]. In Figure 1, a DR protein expressed by DRA1 and DRB1*04:01 is depicted. Eplets are highlighted on the DRB1 chain and are found mainly on the beta helix of the peptide-binding region. Three eplets are found in the region encoded by exon 3, indicating that antibody binding can occur outside the HLA region involved in T-cell receptor binding. Of note, no eplets were located within the groove, which is occupied by processed peptide.

Converting HLA alleles to eplets

The 564 class I and 290 class II common and well-documented alleles in HLAMatchmaker were used to define a string of eplets for each allele. The network diagrams in Figure 2 show the extensive sharing of eplets by alleles within and between HLA class I A, B and C gene loci, while class II eplets were shared only by alleles within the same gene locus. DQA1 and DPA1 contained two mutually exclusive allele groups of eplet expression (Fig. 2e and 2f). For example, DPA1 alleles expressed either 50QA or 50RA.

A total of 361 unique HLA alleles were identified among the 1846 subjects, 206 of which were at HLA class I loci (59 at A, 107 at B, and 40 at C loci) and 155 at HLA class II loci (56 at DRB1, 7 at DRB3, 4 at DRB4, 6 at DRB5, 20 at DQA1, 18 at DQB1, 7 at DPA1 and 37 at DPB1 loci) (Table 1). The class I alleles individually encoded 0 to 11 eplets and class II alleles encoded 0 to 17 eplets. In total, the 361 alleles in the study population encompassed 150 eplets of which 59 were at class I loci (31 at A, 26 at B and 16 at C) and 91 were at class II loci (38 at DRB1/3/4/5, 11 at DQA1, 32 at DQB1, 2 at DPA1 and 8 at DPB1) (Table 1). The reduction in complexity in conversion of HLA alleles to eplets is depicted in Figure 3.

Numerous intra-locus eplets were identified, encoded by multiple alleles within the same gene. The class I eplet 131S, for example, was encoded by 90 class I alleles whereas the 163RG eplet was encoded by only 2 alleles (A*01:01 and A*01:02), and the class II eplet 25R by 67 alleles while the 25Q eplet was encoded by only 1 allele (DRB1*07:01). Multiple inter-locus eplets were also present, encoded by more than 1 gene within each gene region (Supplemental Table 2). Thirteen of these (22%) were encoded by more than one class I gene, with one eplet (163EW) occurring by all three class I genes and the remainder by two genes and fourteen (15%) were

encoded by more than one class II gene, restricted to the DRB1/3/4/5 alleles. No eplets were shared between class I and class II alleles.

Relative frequencies of HLA alleles and eplets

Most of the 361 alleles observed occurred with low population frequencies (Figure 4a and 4b, Supplemental Table 3). Less than 2% (n = 7) were carried by more than 30% of subjects (class I: HLA-A*02:01; and class II: HLA-DPA1*01:03; DPB1*04:01; DRB4*01:03; DQB1*03:01; DRB3*02:02; and DQA1*01:02) while over half (n=188) were present in less than 1% of subjects producing a highly skewed frequency distribution (Figure 4a and 4b).

In contrast, eplets generally occurred with much higher frequencies (Figure 4c and 4d, Supplemental Table 4). Of the 150 eplets identified, over three quarters (n = 113) were carried by more than 30% of subjects and the most frequent class I (79GT and 69TNT) and class II eplets (85VG(DRB), 25R, and 77T) were carried by 90% and 99% of subjects respectively. Even the least frequent eplets (class II: 164VQ, 40D2 and class I: 17RS, 62LQ) were carried by 3-5% of subjects.

Eplet frequencies are comparable in patients and donors.

Despite close linear correlation in allele frequencies between patients and donors (r = 0.975, Figure 5A), some important differences were observed. For example, the most common class I alleles, HLA-A*02:01, occurred respectively in 31% of patients and 40% of donors while the most common class II allele DPA1*01:03 occurred in 83% of patients and 91% of donors. Less common alleles were infrequent in both groups, for example, B*42:01, was present in 0.2% of patients and 0.1% of donors. Comparison of patient and donor sub-groups showed close correlation between

patients prior to or post-transplantation (r = 0.992, SFig 2A) though greater disparity of both groups between deceased donors (r = 0.949, SFig 2C; r = 0.961, SFig 2E). Comparison of patients and living donors (r = 0.973, SFig 2B; r = 0.980, SFig 2D), and living and deceased donors (r = 0.985, SFig 2F) produced similar results. See S2 Table for allele frequencies in table format.

Eplet frequency profiles for donors and recipients were comparable across the frequency spectrum (r = 0.991) (Figure 5B). The most common class I eplet, 79GT, encoded by all but 4 identified HLA-A alleles (n=48), was present in 96% of patients and 97% of donors while the least common class I eplet, 17RS, encoded by 6 HLA-A alleles was present in 4% of patients and 3% of donors. Similar results were observed for class II eplets where 85VG(DRB) occurred in 99% of both recipients and donors. Comparison of patient and donor sub-groups showed similar close correlation between patients prior to and post-transplantation with little disparity between both groups and deceased and living donors (SFig 2G- 2L). See S3 Table for eplet frequencies in table format.

Genotype frequencies

A total of 1800 discrete genotypes, comprising the 16 - 18 alleles encoded at each of the 11 HLA gene loci on both chromosomes (note that DRB3/4/5 may be absent or hemizygous in an individual genotype) were identified in the 1846 study subjects and combinations of these are shown in Table 1. The 206 class I alleles identified were combined in 1572 discrete genotypes and the 155 class II alleles in 1509 discrete genotypes. Diversity at a single gene locus ranged from a maximum of 107 alleles and 602 genotypes at HLA-B to 7 alleles and 14 genotypes at DPA1 (Table 1).

Genotype distribution differed between patients and donors: 1017 complete genotypes (comprising all loci) were observed uniquely in patients and 756 uniquely in donors, with only 27 genotypes

(1.5%) occurring in both groups (Figure 6a). The number of shared genotypes increased as fewer gene loci were considered; for example, 6% of class I genotypes and 7.4% of class II genotypes were shared between patients and donors. The specific HLA gene locus was of primary importance: 30% of genotypes were shared at DRB1/3/4/5+DQB1, 37% at DRB1/3/4/5, 51% at DPB1, 64% at DPA1, and 78% at DQA1. Genotype sharing was most common at DQB1, with 90 genotypes (79%) occurring in both groups. No class I or class II genes had 100% of genotypes present in both patients and donors.

Epitype frequencies

A total of 1793 discrete epitypes, comprising the eplets expressed on the HLA proteins encoded by each of the 11 gene pairs (for example, the A*30:01/A*33:01 genotype encodes the epitype: 17RS, 56R, 62RR, 62QE, 79GT, 138MI, 253Q) were identified in the 1846 subjects, with frequencies ranging from a maximum of 4 (0.22%) to a minimum of 1 (0.05%) (Table 1). The number of identified epitypes was marginally lower than that of genotypes at all individual loci and locus combinations. For example, the 59 HLA class I eplets identified were combined in 1487 discrete epitypes and the 91 class II eplets in 1086 discrete epitypes. Diversity at a single gene locus ranged from a maximum of 26 eplets and 288 epitypes at HLA-B, to 2 eplets and 3 epitypes at DPA1.

Epitype distribution also differed between patients and donors: 1010 epitypes (comprising all loci) were observed uniquely in patients and 751 uniquely in donors, with only 32 epitypes (1.8%) being observed in both groups (Figure 6b). The number of shared epitypes increased as fewer gene loci were considered; for example, 8% of class I epitypes and 19% of class II epitypes were shared between patients and donors. The specific HLA gene locus was also important: 45% of epitypes

were shared at DRB1/3/4/5+DQB1, 57% at DRB1/3/4/5, and 79% at DPB1. For DPA1, DQA1, and DQB1, 100% of epitypes occurred in patients and donors, with a significant decrease in the number of genotypes compared to epitypes.

Cluster analysis

Cluster analysis showed patterns of eplet expression in patients and donors (heat maps that visualize these patterns can be found in Supplemental Figure 3). Certain class I eplets (e.g. 79GT, 9GT, 69TNT, 193PV, 131S, 150AAH, 76VRN, 76 VS, 253Q, 76ESN, 90D, 18MI, 144KR, and 62QE) occur commonly in patients while others are infrequent (56R, 17RS, 144QL, and 62LQ). Certain eplets such as 82LR, 80I, 65GK, 62EE, 166DG, 144TKH, and 127K, are prevalent in clusters, related to the presence of specific alleles such as 44KM3, 163RG, and 76ANT which are encoded by HLA-A*01:01 and/or A*01:02. These clustering patterns were also present in donors, including the commonly occurring class I eplets stated previously.

Class II cluster analysis showed even less variability than class I. Clusters of common and rare eplets were more distinct for class II. Common eplets across patients and donors include 28T, 46VY, 52P, 77T(DRB),73A, 25R, 4R, and 85VG(DRB). Some clusters were related to specific alleles, such as 26WN, 18L, 48Q, 40YNL, 81Y, and 48Q encoded by DRB1*04 alleles, 180LT and 96Y encoded by DRB1*04 alleles, and 87F encoded by DQB1*06 alleles. These results are consistent with linkage disequilibrium between these alleles.

Eplet mismatch without prospective matching

Base-case modeling estimated the probability of patient-donor eplet mismatch by comparing all deceased donors and patients, generating a total of 92,756 potential matches. The numbers of eplet

mismatches across all genes and various gene combinations are shown in Figure 7A. Median mismatch (and ranges) were 27 (0-65) for all 11 genes, 10 (0-27) for class I genes, 17 (0-46) for class II genes, 6 (0-21) for DQB1 only, 6 (0-20) for DRB1/3/4/5 genes, and 12 (0-39) for DRB1/3/4/5 + DQB1 genes. The probability of a 0 eplet mismatch by chance alone was 12% at DQB1, 6% at DRB1345, and <5% for all other combinations.

Eplet mismatch frequencies were inversely correlated with total population frequencies ($r = -0.998$) (Figure 7B) in that common eplets were infrequently mismatched and vice versa. For example, the most common eplet 85VG(DRB) (present in 99% of subjects) occurred in 1.3% of possible mismatches. In comparison, 40D2 occurred in 2.7% of subjects but occurred in 97% of possible mismatches.

Furthermore, the mismatch results showed a high probability of identifying a recipient for each successive donor with a mismatch score of 0 at DRB1/3/4/5, at DQB1 and at these combined gene loci. For DRB1/3/4/5, 93% of donors matched at least one patient with a mismatch score of 0, while a further 5% had a minimum mismatch score of 1, and 0.8% a score of 2. For DQB1, all donors matched at least one patient with a mismatch score of 0 except for one donor with a minimum score of 1. For the combined DRB1/3/4/5 + DQB1 loci, 84% of donors could be matched with a patient having a mismatch score of 0 at all 5 loci, with a 95% having a mismatch score of score 1 or less, and 97.5% a score 3 or less.

Eplet mismatch with prospective matching

Simulation was performed using the Canadian national data comprising 9 provincial programs with a combined waiting list of 2,032 patients and 762 deceased donors (approximately 1500 kidneys) per year (Figure 8a). Prospective matching within this national pool enabled a high

degree of eplet match, with full compatibility at the critical HLA-DR and DQ loci. Compared with no matching, the median mismatch score (and range) declined from 27.35 (0 – 62) to 9.3 (0 – 22) for the full epitype; for class I from 10.2 (0 – 27) to 3 (0 – 11) and for class II from 16.8 (0 – 45) to 1 (0 – 13); and for DRB1/3/4/5 from 6 (0 – 20) to 0 (0 – 4), for DQB1 from 6 (0 - 21) to 0 (0 - 5), and for DRB1/3/4/5 + DQB1 combined from 12.3 (0 - 39) to 0 (0 – 10).

Modeling the average provincial program (waiting list 290 patients; 109 deceased donors / year) also predicted a high degree of eplet match (Supplemental Figure 4) with full compatibility at the critical HLA-DR and DQ loci. Median mismatch score (and range) declined from 27.15 (0 – 60) to 11.8 (0 – 31) for the full epitype; for class I from 10.25 (0 – 24) to 3.9 (0 – 13) and for class II from 16.8 (0 – 42) to 3.3 (0 – 27); and for DRB1/3/4/5 from 5.95 (0 – 19) to 0.1 (0 – 11), for DQB1 from 5.9 (0 - 21) to 0 (0 - 11), and for DRB1/3/4/5 + DQB1 combined from 12.2 (0 - 36) to 0 (0 – 21).

Scenario analyses performed across the range of provincial programs confirmed the importance of waiting list size in determining matching success, with an inflexion at approximately 250 recipients indicating the minimum number to optimize eplet matching (Figure 9A, 9B, and Supplemental Table 5). The probability of perfect eplet identity (mismatch score = 0) at HLA-DQB1 was 92% with a waitlist of 790 and compared with 65% with a waitlist of 88. Requirement for good, as opposed to perfect identity (a cumulative mismatch score of 10 or lower), improved the probability of successful matching across all programs and gene loci except for the full epitype (Figure 9B), with even the smallest program achieving 87% successfully matched pairs.

Extension of these models to other organs showed that regional or national sharing may be required to enable epitope compatibility for heart, lung and liver transplants whose national waiting lists

253 (n= 72, 150, and 285) and donor totals (n= 141, 306 and 430) are smaller than those of the kidney

254 [3].

255

256

257

258

259

260

261

DISCUSSION

Complementary genomic and proteomic methods have clarified the structural biology of HLA antigens, enabling more precise understanding of the complex and sequential mechanisms of allorecognition by T and B lymphocytes [14–16]. Two cardinal groups of epitopes are now recognized, those involved in indirect recognition of the donor HLA antigen array by recipient T cell which are predicted through the PIRCHE algorithm [17], and those which are antibody-accessible and are involved in the humoral response, predicted through the HLAMatchmaker algorithm [10]. Studies confirm that the mismatch between donor and recipient for each of these two sets of molecular targets is directly related to the risk of rejection and graft loss [11,12,18–20]. In this study we focus on antibody-accessible eplets, restricting our attention to those for which biological relevance has been verified by the detection of specific antibodies to the target (http://www.epitopes.net/publications.html). The majority of these are encoded within exons 2 and 3 for class I and exon 2 for class II antigens, though putative eplets may exist outside these regions on the expressed protein. We document the probability distributions of these eplets and the broader epitypes determined unambiguously by NGS among patients and donors in a large and ethnically diverse transplant program, and estimate the quantitative mismatches achievable at each HLA gene locus with or without prospective matching. These data provide the basis to inform strategic decisions for incorporating quantitative epitope mismatch data into clinical practice, either through retrospective use of a mismatch score to estimate risk and adjust immune suppression, or by prospectively matching donors and recipients to minimize incompatibility and improve overall outcomes.

The results reported reinforce the small proportion of documented HLA alleles commonly observed in routine practice [21]. Only 361 of the more than 25,000 class I and class II HLA alleles

(1-2%) were observed in the patient and donors from this highly ethnically diverse population, most of which were present in fewer than 5% of subjects; only 7 individual alleles were observed in more than 30% of cases. Matching for identity at the allele level is therefore challenging. Large donor registries have been established to achieve this in hematopoietic stem cell transplantation, a strategy not feasible in organ transplantation [22]. Eplets, in contrast, are fewer in number and often distributed across multiple alleles within or between gene loci, resulting in higher population frequencies and more linear distribution. Of the 150 eplets identified, three quarters were present in more than 30% of subjects and several occurred in over 90% of subjects in both patient and donor groups, increasing the potential for prospective matching to enhance compatibility.

Transplant patients and deceased donors often differ in ethnicity especially in an ethnically diverse population such as British Columbia [23], raising concern that allelic diversity may create an inequality barrier by increasing waitlist times for ethnic minorities with rare genotypes, diminishing the enthusiasm for stringent HLA matching for deceased donor transplantation in Canada and other locations of high population diversity [24–26]. We observed differences in allele carrier rates between deceased donors and patients in this study, while both patient sub-groups, prior to or post-transplantation serving as internal controls, showed tight correlation with each other ($r = 0.992$). Conversion from alleles to eplets not only reduced the HLA complexity but minimized the consequences of ethnic diversity. Every eplet was present in both patients and donors, compared with 29% of alleles that were absent in one or other of these groups. In consequence, when examined at the epitope level, correlation was robust between donors and patients prior to ($r=0.987$) or post-transplant ($r=0.980$), comparable to the correlation between these two individual patient groups ($r = 0.996$).

Although eplets are widely expressed in donors and patients, often overlapping in clusters related to the presence of common class I or class II alleles, the number of discrete epitypes (eplets present at all 12 gene loci) identified (n=1793) was very similar to the number of HLA genotypes (n=1800). No more than 1-2% of epitypes were shared between donors and patients, with only 6% sharing at the class I region and 7.4% at the class II region, indicating that identical matching at the epitype or at each gene region is improbable in a diverse population. However, increasing data suggests that eplet compatibility at certain class II gene loci, particularly HLA-DRB1/3/4/5 and DQB1 is of primary importance in minimizing graft injury [11,12,20], reflecting the frequent occurrence of antibodies to these gene products in AMR [27–29]. Our studies (R.S-P) utilizing the US Scientific Registry of Transplant Recipients (SRTR) confirm this understanding, documenting an increased risk of transplant glomerulopathy and death-censored graft failure in donor/recipient pairs mismatched at these loci [30]. The data reported here demonstrate that the probability of identity at these class II loci is substantially higher than for the full epitype, ranging from 30% at HLA-DRB1/3/4/5 + DQB1 to 79% at DQB1 alone, so providing a logistical basis for deliberate matching at these loci.

In focusing on AMR, we have primarily examined antibody-verified eplets and employed mismatch counts/scores as a measure of incompatibility. However, since our understanding of relative immunogenicity remains limited, it is likely that not all polymorphisms have the same biological effect in the rejection process and that structural properties of the eplet could be equally or more important than the simple quantity of mismatches. This raises critical questions related both to the specific eplet differences and the quantitative sum of these differences between recipient and donor. For example, a higher mismatch score may confer a summative biological effect, or simply increase the probability of a highly immunogenic eplet being present in the donor.

The immunogenicity of an eplet is relevant only in context of the three-dimensional structure of the HLA proteins expressed by the transplant recipient. Kosmoliapsis et al. also consider the disparity of physiochemical properties (i.e. electrostatic score, EMS) of mismatched amino acids, and report a greater EMS to be an independent predictor of the formation of antibodies to donor DR and DQ targets [31]. The location of the mismatched eplet may also play a role in its immunogenicity as described by Tambur et al [32]. Polymorphisms occur at different locations in the HLA protein, including both the peptide binding cleft and the outer aspects of the molecule, which influence how peptide is presented to the T cell. More precise understanding of eplets and their role in transplant rejection is therefore critical to refine matching algorithms.

The use of retrospective mismatch scores to select patients at risk of rejection and to modify management is valuable, but does not maximize the benefit of eplet compatibility. As shown here, the allocation of deceased donors to transplant patients without deliberate HLA matching showed a median epitope mismatch of 27 across the full epitype, 10 across class I genes and 17 across class II genes. Even at individual gene loci, the estimated median mismatch score ranged from 6 (range: 0-21) at DQB1 alone to 12 (range: 0-39) at DRB1/3/4/5 + DQB1, suggesting a wide range of rejection risk in non-matched transplant populations. But the implementation of prospective matching requires thoughtful consideration of logistics. Here we show that the opportunity to maximize compatibility by prospective eplet matching is influenced by the size of the transplant wait-list. Simulation modeling based on a broad range of donor and patient frequencies indicated that a wait list of 250 or more active recipients offers the greatest opportunity to achieve optimal matching at key class II HLA gene loci. This number, which is consistent with the waiting list in major transplant centres in America, Europe and many other countries, is important in informing organ sharing policies within regions, particularly in geographic areas with low population

densities, since the logistics, costs and storage time involved in organ sharing must be balanced against the benefits achieved in prolonging outcomes [33–36].

These results provide compelling evidence that prospective donor/recipient eplet matching is feasible in the Canadian population and, while not enabling full epitype identity, matching may successfully achieve a very low – or zero – antigen mismatch at the critical HLA class II loci in the majority of patients awaiting transplant. Further, the results indicate that a high degree of successful matching at these loci may be achieved within program or region, assuming a minimum waiting list of 250 patients. This is of vital importance since, while national organ sharing is routine for highly sensitized patients, the costs and logistical complexities of transporting all organs nationally would be substantial. Our data suggest that, in certain cases, a small number of contiguous programs may need to be combined to ensure transplant regions with adequate wait-list numbers for kidney, and that regional or national sharing will be required for non-renal organs. But graft success or economic costs are not the sole arbiters of policy and the utility donated organs must be balanced by equality of access to them [24]. Matching at the eplet level may more closely approximate this latter goal than the simple use of allele compatibility, though accommodation must still be considered for recipients with uncommon eplets of high biological importance.

Limited approaches to eplet-matching have been incorporated by other programs. Eurotransplant has successfully used class I eplet-matching to expand the donor pool for highly-sensitized patients [37–39] and eplet matching of class I (<10 eplets) and class II (<30 eplets) has been performed in Australian pediatric patients[40]. Although these examples are small the results show promise.

This study has certain limitations which we are working to address. We restricted this analysis to antibody-verified eplets because of their demonstrated clinical importance in AMR. It is possible

that putative B-cell eplets for which antibodies have not yet been identified, or other eplets which are recognized primarily by the T-cell receptor (PIRCHE), also play an important role in graft rejection and we are therefore evaluating these in subsequent analyses. A small number of subjects (8%) could not be included since they expressed one or more alleles which are not present in HLAMatchmaker. We are working to update HLAMatchmaker to include these alleles and will re-evaluate once this process is complete. Subject selection and sequencing were performed in a single provincial program, raising potential concern for both precision and representativeness. B.C. has served as the lead program for evaluation, validation and implementation NGS for HLA genes in Canada, and the accuracy and reproducibility of these assays have been fully validated according to ASHI standards. But the ethnic diversity of the program and the small number of subjects excluded, we believe the data is highly representative of the epitope frequencies observed across Canada. We are currently engaged in a larger study to confirm these national data. In modeling matching probabilities, we extrapolated the allelic and eplet frequencies observed in this program to the broader pool of patients and donors across Canada. Precise national population frequencies will enable us to refine these model probabilities. And we acknowledge the limitations of model parameters, which we have deliberately restricted to allocation by blood group identity and optimal eplet match. Recognizing these considerations, we present the first data describing HLA eplet frequencies in patients and donors in a highly diverse ethnic population. We show that the conversion of alleles to eplets reduces the HLA complexity and enables matching at selected clinically important gene loci. And in a simple allocation model we demonstrate that a high degree of eplet compatibility can be achieved at these loci with a relatively small waiting list, reducing the requirement for national transport of all donor organs and so minimizing costs and ex-vivo storage time. We are now proceeding with studies to define more precisely the immunogenicity of

dominant eplets and to incorporate eplet frequencies from other regions in a more comprehensive model to guide allocation within a national program.

REFERENCES

1. Tonelli, M. *et al.* Systematic review: Kidney transplantation compared with dialysis in clinically relevant outcomes. *Am. J. Transplant.* **11**, 2093–2109 (2011).

2. Fu, R., Sekercioglu, N., Berta, W. & Coyte, P. C. Cost-effectiveness of Deceased-donor Renal Transplant Versus Dialysis to Treat End-stage Renal Disease. *Transplant. Direct* **6**, e522 (2020).

3. Organ replacement in Canada: CORR annual statistics, 2019 | CIHI. https://www.cihi.ca/en/organ-replacement-in-canada-corr-annual-statistics-2019.

4. Sellarés, J. *et al.* Understanding the causes of kidney transplant failure: The dominant role of antibody-mediated rejection and nonadherence. *Am. J. Transplant.* **12**, 388–399 (2012).

5. Halloran, P. F. *et al.* Disappearance of T cell-mediated rejection despite continued antibody-mediated rejection in late kidney transplant recipients. *J. Am. Soc. Nephrol.* **26**, 1711–1720 (2015).

6. Robinson, J. *et al.* IPD-IMGT/HLA Database. *Nucleic Acids Res.* **48**, D948–D955 (2020).

7. Valenzuela, N. M. & Reed, E. F. Antibody Binding and Mechanisms of Injury. 1–26 (2014) doi:10.1007/978-1-62703-493-7.

8. Opelz, G. & Döhler, B. Effect of human leukocyte antigen compatibility on kidney graft survival: Comparative analysis of two decades. *Transplantation* **84**, 137–143 (2007).

9. Opelz, G. Commentary: Is HLA matching useful? *Int. J. Immunogenet.* **37**, 327–328 (2010).

10. Duquesnoy, R. J. HLA epitope based matching for transplantation. *Transpl. Immunol.* **31**, 1–6 (2014).

11. Wiebe, C. *et al.* Class II HLA epitope matching - A strategy to minimize de novo donor-specific antibody development and improve outcomes. *Am. J. Transplant.* **13**, 3114–3122 (2013).

12. Sapir-Pichhadze, R. *et al.* HLA-DR and -DQ eplet mismatches and transplant glomerulopathy: A nested case-control study. *Am. J. Transplant.* (2015) doi:10.1111/ajt.12968.

13. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

14. Duquesnoy, R. J. A Structurally Based Approach to Determine HLA Compatibility at the Humoral Immune Level. *Hum. Immunol.* **67**, 847–862 (2006).

15. El-Awar, N., Jucaud, V. & Nguyen, A. HLA Epitopes: The Targets of Monoclonal and Alloantibodies Defined. *J. Immunol. Res.* **2017**, (2017).

16. Marino, J., Paster, J. & Benichou, G. Allorecognition by T lymphocytes and allograft rejection. *Front. Immunol.* **7**, 1–9 (2016).

17. Geneugelijk, K. *Indirect recognition of HLA epitopes in solid organ transplantation*. (2017).

18. Geneugelijk, K. *et al.* Exploratory Study of Predicted Indirectly ReCognizable HLA Epitopes in Mismatched Hematopoietic Cell Transplantations. *Front. Immunol.* **10**, 1–8

(2019).

19. Daniëls, L. *et al.* The clinical significance of epitope mismatch load in kidney transplantation: A multicentre study. *Transpl. Immunol.* **50**, 55–59 (2018).

20. Wiebe, C. *et al.* Class II eplet mismatch modulates tacrolimus trough levels required to prevent donor-specific antibody development. *J. Am. Soc. Nephrol.* **28**, 3353–3362 (2017).

21. Mack, S. J. *et al.* Common and well-documented HLA alleles: 2012 update to the CWD catalogue. *Tissue Antigens* **81**, 194–203 (2013).

22. Aljurf, M. *et al.* "Worldwide Network for Blood & Marrow Transplantation (WBMT) special article, challenges facing emerging alternate donor registries". *Bone Marrow Transplantation* vol. 54 1179–1188 (2019).

23. Census Profile, 2016 Census - British Columbia [Province] and Canada [Country]. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&SearchText=Canada&SearchType=Begins&SearchPR=01&TABID=1&G=1&Geo1=PR&Code1=59&Geo2=PR&Code2=01&type=0&B1=Ethnic origin.

24. Lee, D., Kanellis, J. & Mulley, W. R. Allocation of deceased donor kidneys: A review of international practices. *Nephrology* **24**, 591–598 (2019).

25. Tiercy, J. M. & Claas, F. Impact of HLA diversity on donor selection in organ and stem cell transplantation. *Hum. Hered.* **76**, 178–186 (2014).

26. Gill, J. S. *et al.* Allocation and Reporting of Deceased Donor Kidney Transplantation in

Canada. *Curr. Transplant. Reports* **6**, 344–347 (2019).

27. Wiebe, C. *et al.* Rates and determinants of progression to graft failure in kidney allograft recipients with de novo donor-specific antibody. *Am. J. Transplant.* **15**, 2921–2930 (2015).

28. Ntokou, I. S. A. *et al.* Long-term follow up for anti-HLA donor specific antibodies postrenal transplantation: High immunogenicity of HLA class II graft molecules. *Transpl. Int.* **24**, 1084–1093 (2011).

29. Guidicelli, G. *et al.* Non-complement-binDing de novo donor-specific anti-HLA antibodies and kidney allograft survival. *J. Am. Soc. Nephrol.* **27**, 615–625 (2016).

30. Sapir-Pichhadze, R. *et al.* Epitopes as characterized by antibody-verified eplet mismatches determine risk of kidney transplant loss. *Kidney Int.* **97**, 778–785 (2020).

31. Kosmoliaptsis, V. *et al.* Alloantibody Responses After Renal Transplant Failure Can Be Better Predicted by Donor–Recipient HLA Amino Acid Sequence and Physicochemical Disparities Than Conventional HLA Matching. *Am. J. Transplant.* **16**, 2139–2147 (2016).

32. Tambur, A. R. HLA-Epitope Matching or Eplet Risk Stratification: The Devil Is in the Details. *Front. Immunol.* **9**, 2010 (2018).

33. Hart, A. *et al.* OPTN/SRTR 2018 Annual Data Report: Kidney. *Am. J. Transplant.* **20**, 20–130 (2020).

34. About Eurotransplant - Eurotransplant. https://www.eurotransplant.org/.

35. Home - NHS Blood and Transplant. https://www.nhsbt.nhs.uk/.

36. ANZDATA - Australia and New Zealand Dialysis and Transplant Registry - ANZDATA. https://www.anzdata.org.au/anzdata/.

37. Heidt, S., Witvliet, M. D., Haasnoot, G. W. & Claas, F. H. J. The 25th anniversary of the Eurotransplant Acceptable Mismatch program for highly sensitized patients ☆ , ☆☆. (2015) doi:10.1016/j.trim.2015.08.006.

38. Claas, F. H. J., Rahmel, A. & Doxiadis, I. I. N. Enhanced kidney allocation to highly sensitized patients by the acceptable mismatch program. *Transplantation* **88**, 447–452 (2009).

39. Duquesnoy, R. J., Witvliet, M., Doxiadis, I. I. N., De Fijter, H. & Claas, F. H. J. HLAMatchmaker-based strategy to identify acceptable HLA class I mismatches for highly sensitized kidney transplant candidates. *Transpl. Int.* **17**, 22–30 (2004).

40. Kausman, J. Y. *et al.* Application of an epitope-based allocation system in pediatric kidney transplantation. *Pediatr. Transplant.* **20**, 931–938 (2016).

41. Pieper, J. *et al.* Memory T cells specific to citrullinated α-enolase are enriched in the rheumatic joint. *J. Autoimmun.* **92**, 47–56 (2018).

42. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**, 1129–1164 (1991).

43. Connors, J. *et al.* Circos : An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

44. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances

circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

TABLES

	Alleles	Genotypes	Eplets	Epitypes
All Loci	361	1800	150	1793
HLA Class I	206	1572	59	1487
HLA Class II	155	1509	91	1086
DRB1/3/4/5+DQB1	91	710	70	365
A	59	253	31	128
B	107	602	26	288
C	40	227	16	74
DPA1	7	14	2	3
DPB1	37	146	8	24
DQA1	20	109	11	10
DQB1	18	115	32	34
DRB1	56	404	28	161
DRB3	7	14	10	7
DRB4	4	8	9	3
DRB5	6	13	11	4
DRB1/3/4/5	73	564	38	199

Table 1. Summary of HLA alleles, genotypes, eplets, and epitypes in the total study population of 1846 subjects at various loci combinations and individual loci.

FIGURES

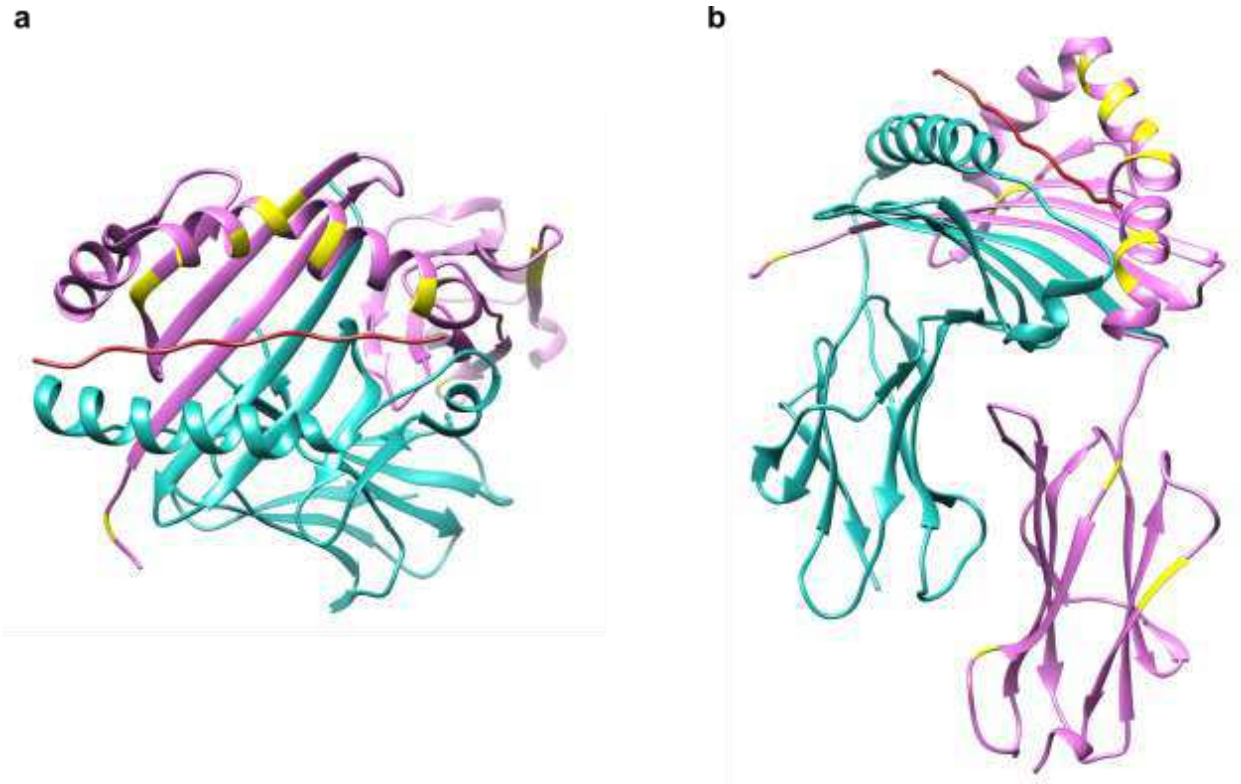


Figure 1. A 3D visualization of an HLA-DR protein with eplets highlighted. The DRA chain is depicted in aqua, DRB1*04:01 chain is depicted in magenta, and processed peptide in dark red. All 11 eplets are highlighted in yellow. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 [13] using the Protein Data Bank ID: 5NI9 [41]. (a) Top-down view of the protein, depicting the peptide-binding groove. (b) Side-view of the protein, depicting its full extracellular portion, showing eplets present outside of the peptide-binding groove.

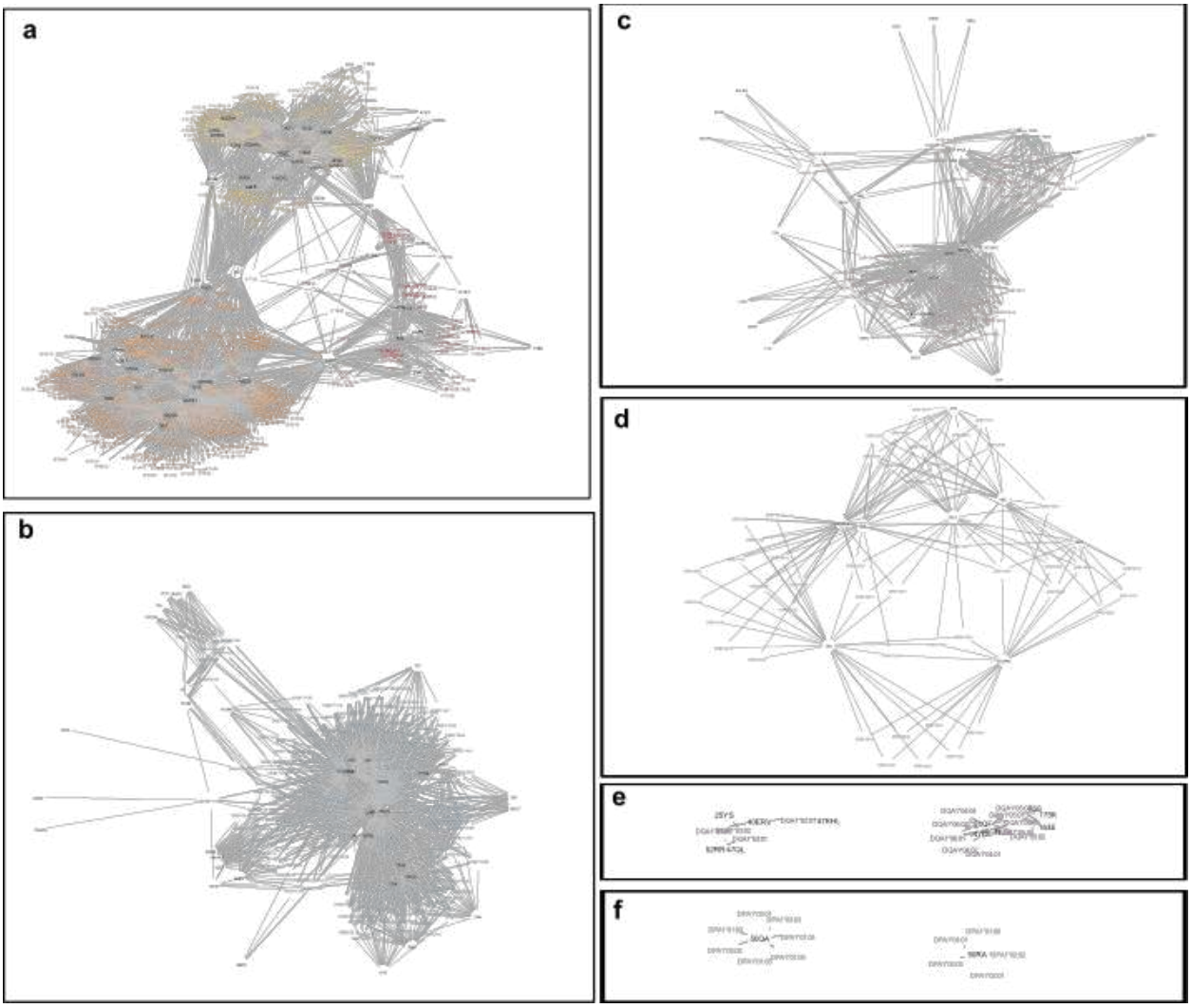
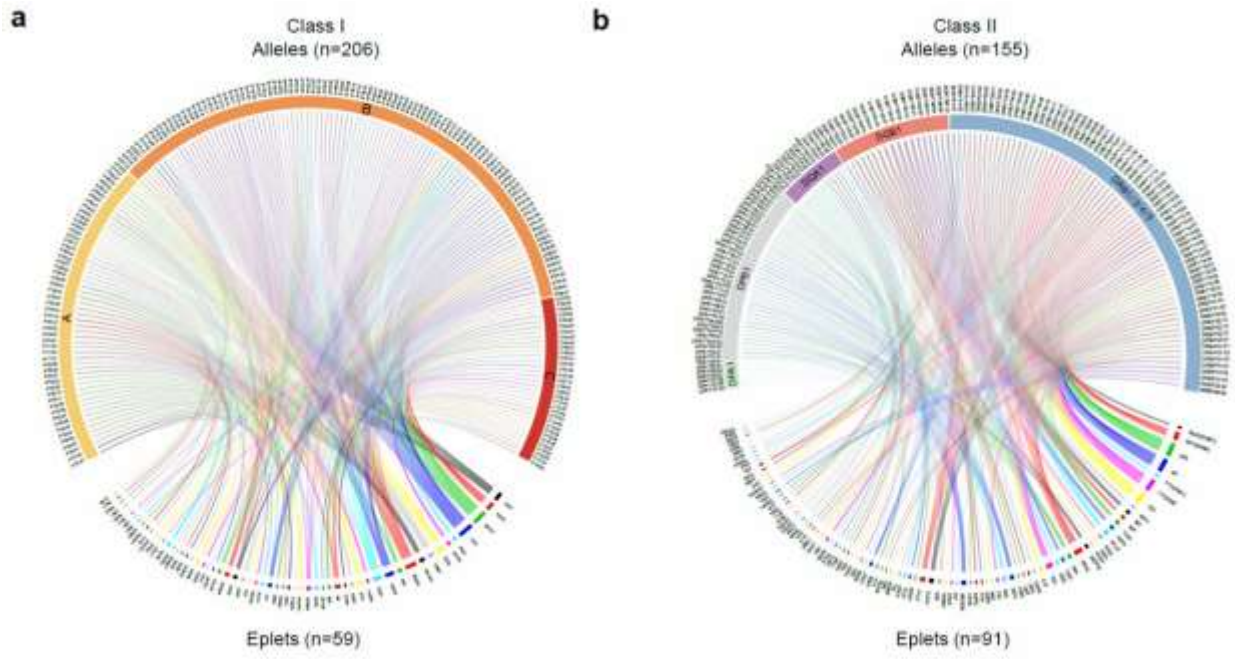


Figure 2. Bipartite networks [42] depicting the associations between the complete library of HLA alleles and eplets present in HLAMatchmaker. Each node represents an allele or eplet and lines represent either that the eplet is encoded by an allele or vice versa. (a) HLA Class I, (b) DRB1/3/4/5, (c) DQB1, (d) DPB1, (e) DQA1, and (f) DPA1.



547

548 **Figure 3.** Circos [43,44] diagrams depicting the reduction of HLA complexity of identified HLA
549 alleles (top portion) in the study population converting into eplets (bottom portion). Conversion of
550 alleles to eplets was determined by HLAMatchmaker. HLA alleles identified in the study
551 population are depicted at the top portion and eplets shown in the bottom portion. Interconnections
552 represent an allele encoding an eplet. (a) Class I alleles to eplets and (b) Class II alleles to eplets.

553

554

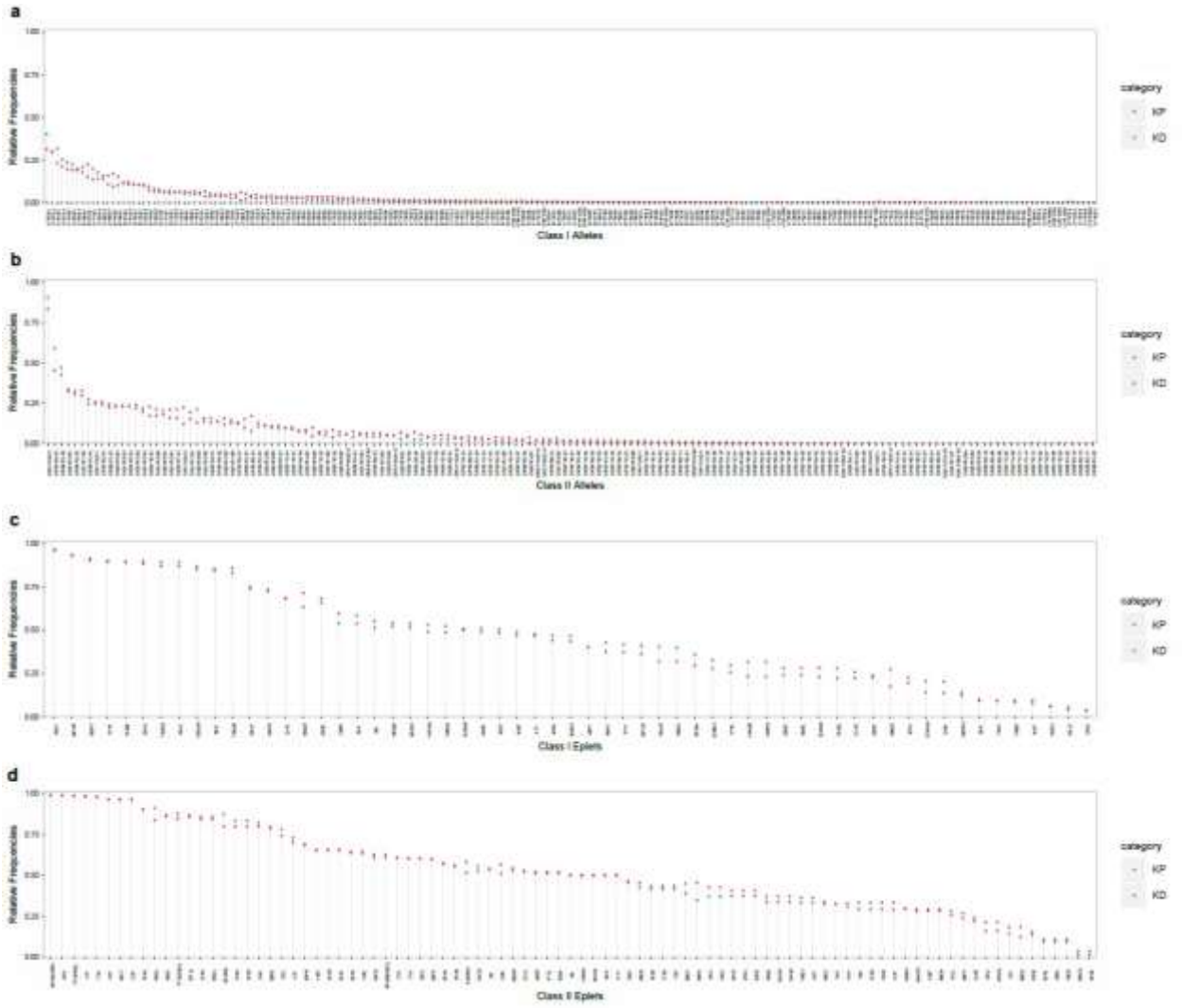


Figure 4. The relative frequencies of HLA alleles and eplets in the study population, calculated as the proportion of subjects expressing a particular allele or eplet. Images (a) and (b) depict the allele frequencies by class I and II, respectively. (c) and (d) depict eplet frequencies by class I and II, respectively. *KP = Kidney Patients, KD = Kidney Donors.*

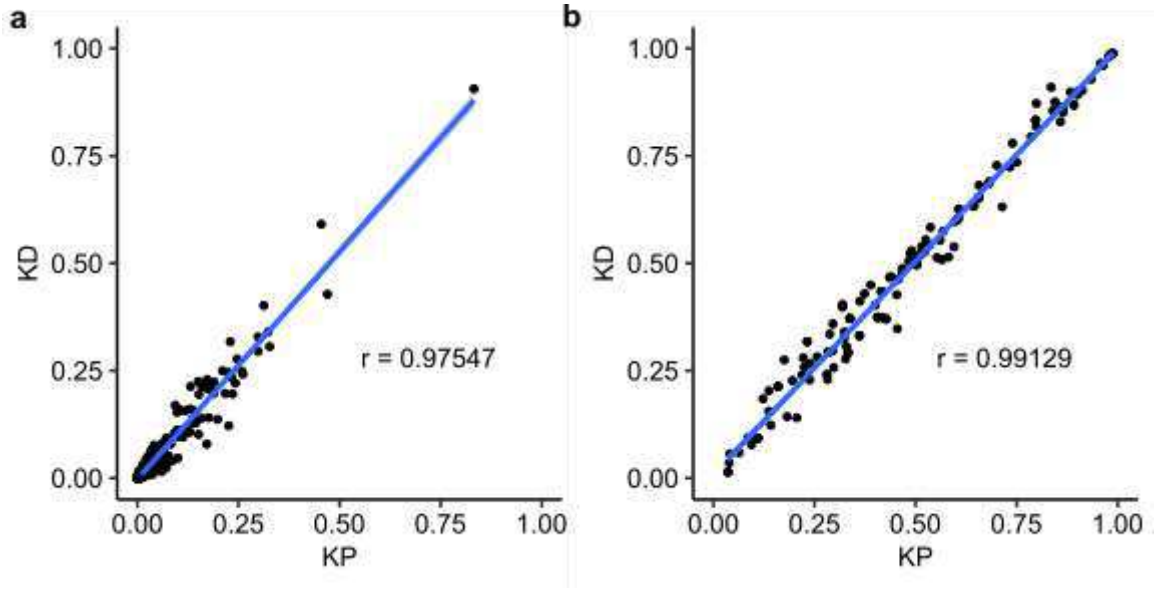


Figure 5. Pairwise analysis between total patient and total donor groups of HLA allele and eplet frequencies calculated as the proportion of subjects expressing a particular allele or eplet. The line and correlation coefficient were calculated using Pearson Correlation. (a) Each dot represents an allele and its frequency for a particular patient or donor group is plotted against another patient or donor group. (b) As above but for eplets. *KP = kidney patients (n = 1049), KD = kidney donors (n = 797).*

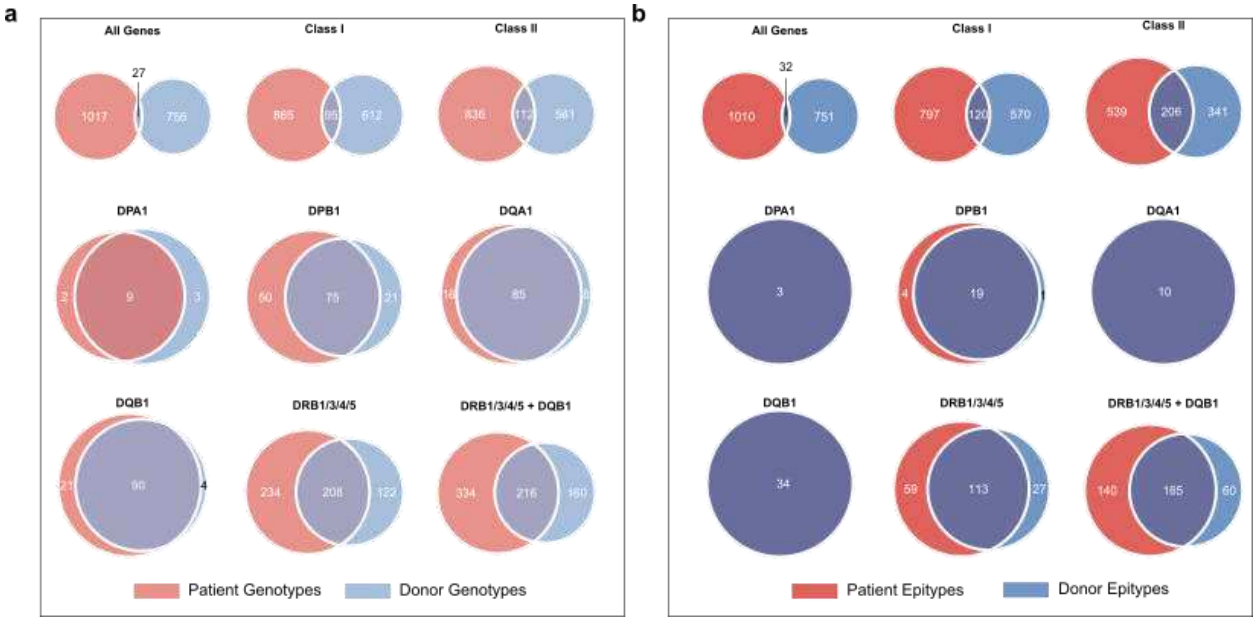
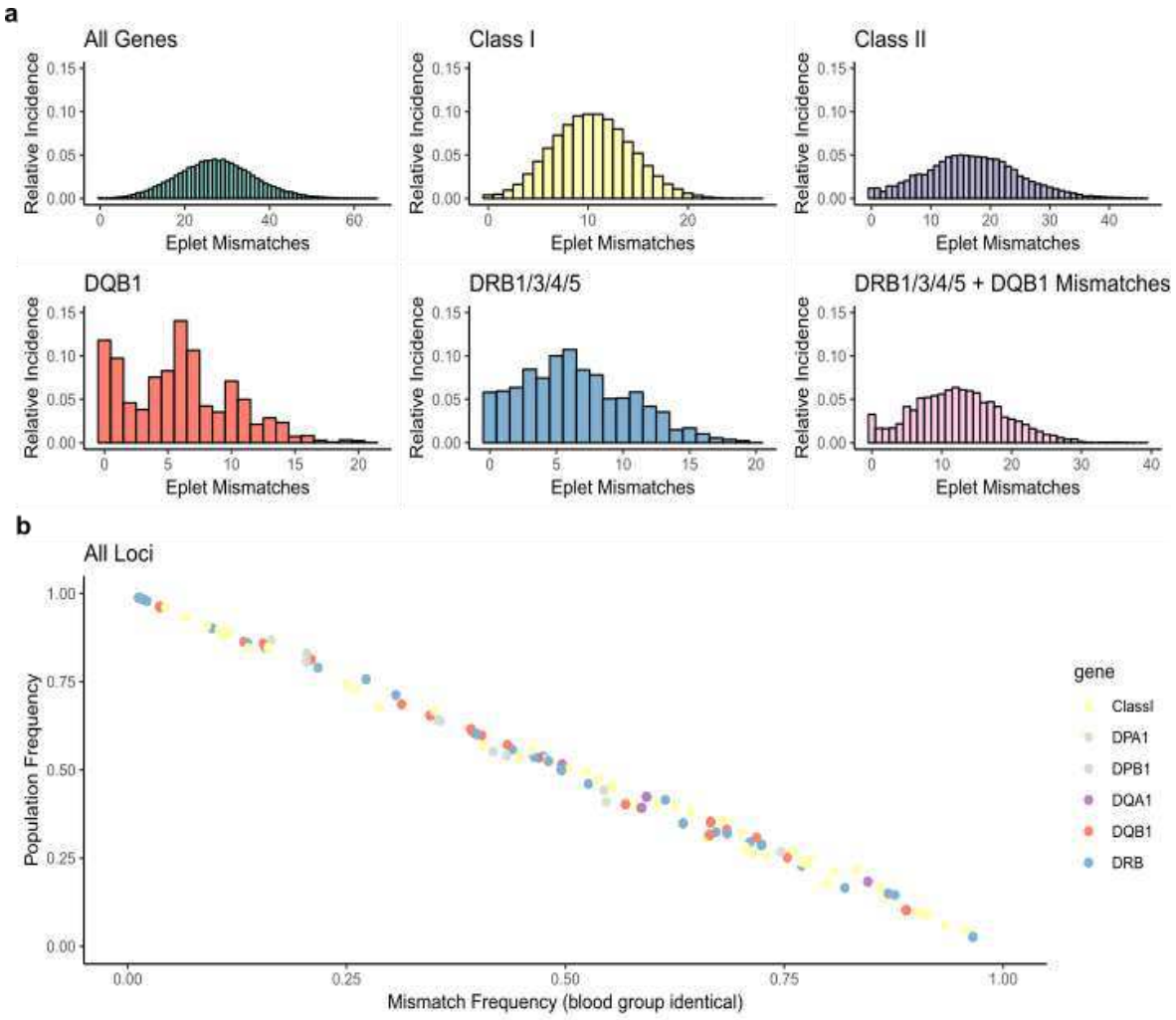


Figure 6. Genotype and epitype frequencies between patients and donors. (a) Venn diagrams of unique genotypes and how some or all occur between patient and donor groups. Genotypes were determined at the individual level, where alleles of a particular loci were sorted and combined. Analysis was performed on various loci combinations: all genes, HLA class I, HLA class II, DPA1, DPB1, DQA1, DQB1, DRB1/3/4/5, and DQB1 and DRB1/3/4/5 combined. (b) As above but for epitypes.



580

581 **Figure 7**. Characterization of eplet mismatches in all blood group identical matches between
582 kidney patients and deceased donors. 1049 kidney patients were matched against 243 deceased
583 donors by blood-group identity. Next, eplet mismatches across all genes (A, B, C, DRB1, DRB3,
584 DRB4, DRB5, DPA1, DPB1, DQA1, and DQB1), HLA class I, HLA class II, DQB1 only,
585 DRB1/3/4/5, and DRB1/3/4/5 and DQB1 were calculated. A mismatch was determined as an eplet
586 present in a donor that is not present in the patient. (a) Distribution of the relative incidences of
587 eplet mismatch scores at all gene combinations analyzed. Relative incidence was calculated by the
588 proportion of a particular quantitative eplet mismatch out of the total number of 92,756 blood-
589 group identical matches. (b) The linear correlation (Pearson's) of population frequency versus
590 mismatch frequency for individual eplets. A dot represents an individual eplet colour-coded by
591 gene combination, and all identified eplets are plotted (n = 150). Population frequency is the
592 proportion of individuals (kidney patients and donors) whom express the eplet, out of the total
593 study population (n = 1846). Mismatch frequency was the incidence of a particular eplet being
594 mismatched in blood-group identical matches, divided by the sum of the number of donors with
595 the specific eplet multiplied by the number of patients, restricted by blood group. The correlation
596 coefficient (r) is -0.998.

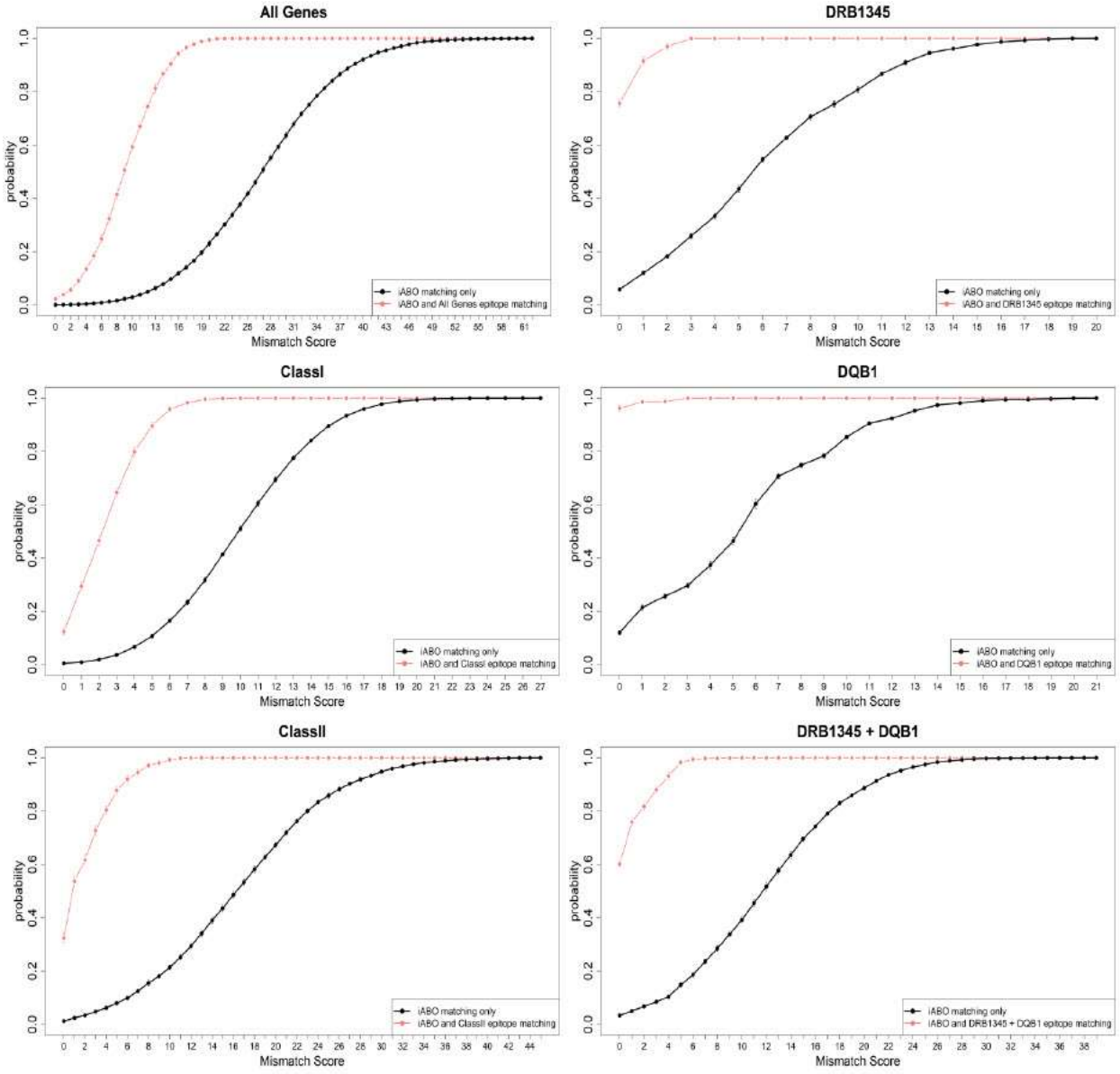


Figure 8. Matching simulations incorporating deliberate eplet and blood group identical matching (red curve) and baseline blood group identical matching only (black curve) in kidney patients and deceased donors. Plots represent simulations using Canadian waitlist and deceased donor data across all genes, HLA class I, HLA class II, DQB1, DRB1/3/4/5, and DQB1+DRB1/3/4/5 combined. Eplet mismatch scores for their respective genes on the x axis plotted against the cumulative probability of these scores in the matched population. Error bars are calculated as the standard deviation of 10 repeated simulation runs.

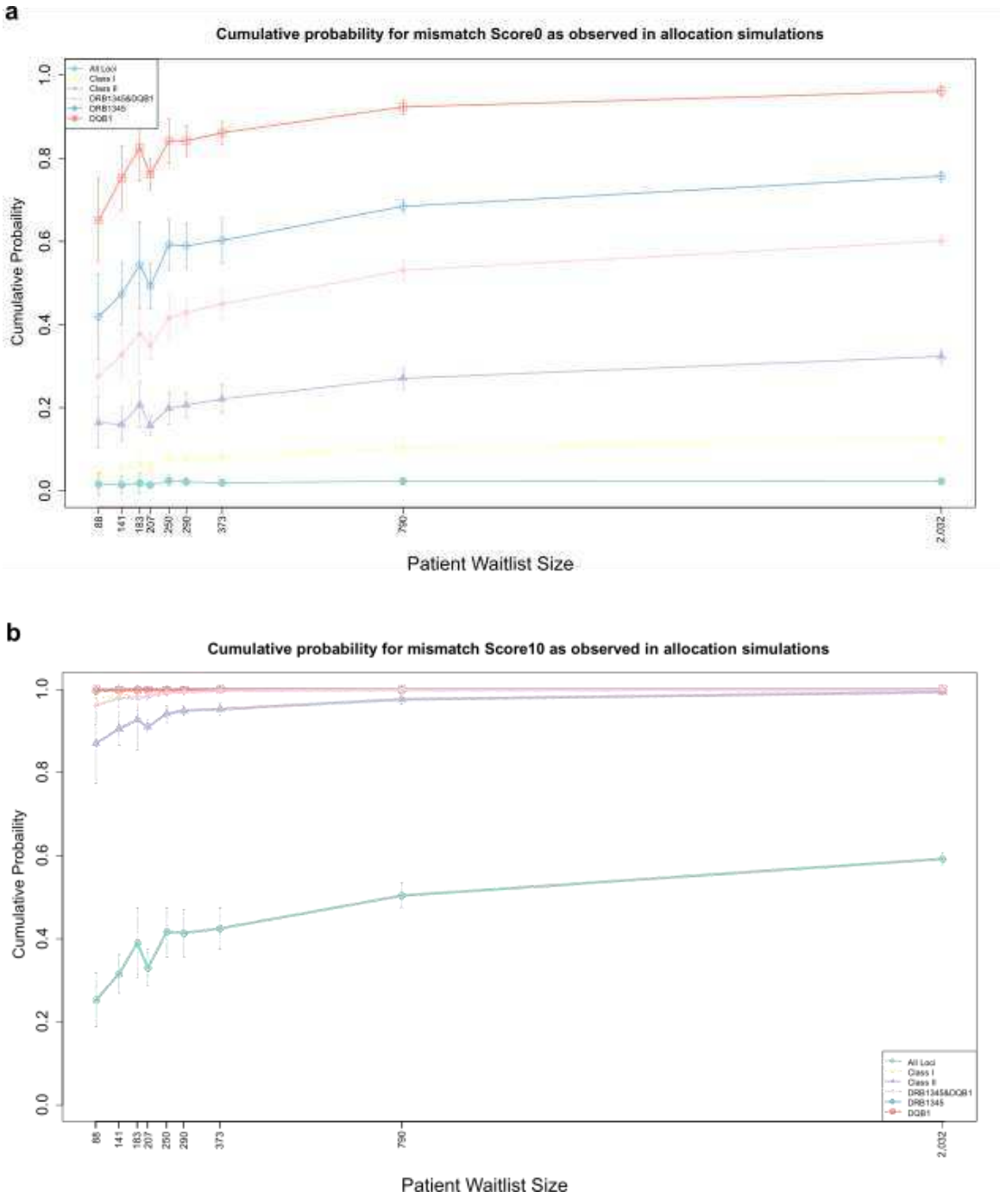


Figure 9. Results of an averaged set of 10 deliberative eplet-matching simulations at various gene loci. Provincial active waitlist and deceased donor numbers were used according to [3]. X axis shows the number of patients on provincial or national patient waiting lists (Saskatchewan 88, Atlantic provinces 141, Manitoba 183, British Columbia 207, Alberta 50, Average Province 290, Quebec 373, Ontario 790, Canada 2,032). Y axis shows the averaged cumulative probability of achieving a total mismatch score of 0 (a) or 10 or lower (b) at the end of the simulations.

MATERIALS AND METHODS

This population-based study included renal transplant patients and donors genotyped by next-generation sequencing (NGS) from October 2016 to January 2019 at the Provincial Reference Immunology Laboratory (Vancouver General Hospital, Vancouver, BC). This research was approved by the University of British Columbia Clinical Research Ethics Board (#H18-00090).

HLA gene sequencing, eplet analysis and carrier frequencies

DNA was extracted from whole blood using the EZ1 DNA Blood 350 µl Kit or QIAsymphony DSP DNA Mini Kit (192) (Qiagen, Hilden, Germany). Genotyping for HLA-A, B, C; DRB1, DRB3/4/5, DQA1, DQB1, DPA1, and DPB1 was performed using the Holotype HLA kit version 2 (Omixon, Budapest, Hungary). Libraries were sequenced using MiSeq Sequencer (Illumina, California, USA) and sequence data analyzed using the HLA Twin version 2.5 (Omixon, Budapest, Hungary).

HLA eplets were obtained from allelic data using the computer algorithm HLAMatchmaker v02 for HLA class I and v02.2 for HLA class II genes (www.epitopes.net) which uses a database of common and well-documented alleles to define a string of eplets for each allele. Only eplets registered to be experimentally confirmed by antibody testing (antibody-verified eplets) were used in this analysis.

Allele or eplet carrier rates (relative frequencies) were calculated respectively as the proportion of subjects expressing a specific allele or eplet within the total study sample. Composite genotypes (HLA types) and their corresponding epitypes were calculated respectively as the numbers of

subjects within the sample expressing each unique combination of alleles or eplets at all the 11 allelic HLA gene loci.

Correlations between allele and eplet frequencies in different sub-groups of renal patients (prior to or following transplantation) and donors (potential living or deceased donors) were determined using Pearson's correlation coefficient. Locus-specific epitype patterns for patient and donor populations were explored using hierarchical clustering approaches and heat map visualizations. For each gene of focus, a binary epitype vector was defined based on the presence or absence of eplets of interest. Epitype vectors were then clustered using the complete linkage method in the R hclust-function and the Manhattan distance metric which counts all eplet absence-presence differences between two epitype vectors with equal weight.

Eplet matching and simulation

An allocation simulation framework was implemented in R (MRAN 3.5.2 and 3.5.3) to model kidney matching between deceased donors and patients. Simulation was initialized using wait lists of specified size and bootstrapped recipients were added one at a time until the determined number was reached. This process produced an initial rank ordering where the first recipient added was at the top, and the last recipient added was at the bottom of the wait list. Each donor was considered to provide two donor kidneys for matching with recipients on the wait list using defined matching rules (see below). Before the next donor was entered into the model, two new recipients were randomly selected from the recipient distribution and added to the bottom of the wait list, keeping the wait list at a constant size over the course of the simulation. Simulation continued until all donor kidneys were allocated. Eplet mismatch was defined as the presence of an eplet in a donor that was not present in the recipient. Numerical eplet mismatches were calculated across all genes

or selected loci combined for each donor-recipient pair. Match-information, including eplet and epitype mismatch scores, were recorded and stored in a Match-List table for post-simulation analysis.

The baseline scenario was structured to approximate the current Canadian allocation model in which deceased donor allocation is performed primarily within the province of kidney origin, with national sharing for a small proportion of primarily highly-sensitized patients. Time on the wait list is the principal determinant of ranking order, adjusted to allow for clinical priority in a small proportion of subjects (e.g. children, loss of dialysis access, other cases of exceptional clinical urgency), and constrained by ABO identity to avoid over-allocation of group O donors to non-O recipients. Within this system, HLA compatibility is used only as a lower-level decision factor to select between individuals of otherwise equal ranking, and organs are therefore normally allocated independent of eplet mismatch.

Exploratory simulation models were then developed to examine the impact of deliberate eplet matching across a range of wait-list and donor pool sizes representing Canadian provinces obtained from the 2018 Canadian Institute of Health Information data (https://www.cihi.ca/en/organ-replacement-in-canada-corr-annual-statistics-2019). Allocation was modeled with constraints for ABO identity, and organs were assumed to be freely shared across Canada in the full national model. Eplet mismatching was calculated for all HLA 11 genes, for class I and class II regions, and for DRB1/3/4/5/+DQB1, DRB1/3/4/5 and DQB1 loci. Each donor was considered to be matched with the recipient having the lowest eplet mismatch score at the relevant HLA locus or loci, with the rank on the waitlist determining priority in cases of identical scores. Fifty-six sets of

simulations were performed with 10 replicates each (i.e. running the same simulation for different random orderings of recipients and donors) and the cumulative probability of increasing mismatch scores was derived for each scenario over the 10 replicates.

ACKNOWLEDGEMENTS

We are indebted to the members of the Genome Canada Transplant Consortium for their contribution to this research. Research was supported by Genome Canada, Genome British Columbia, Genome Quebec, and Genome Alberta, Canadian Institutes of Health Research, and partnered grants from Omixon.

AUTHOR CONTRIBUTIONS

J.N.T, O.P.G, K.R.S, and P.A.K contributed to the conception and the design of the study and wrote the manuscript. J.N.T performed the sequencing, collection of data, and analysis of alleles and eplets and their frequencies. O.P.G performed the network analysis and simulations. O.P.G and F.F performed the cluster analysis. P.A.K guided the research and provided oversight of the work. J.N.T, O.P.G, P.A.K, K.R.S, F.F, L.L.A, J.L, R.S-P, F.C, S.G.E.M, R.M, and R.D contributed to the review and interpretation of the data or results. All authors contributed to the drafting and critical review of the article and provided final approval of the manuscript.

Competing Interest Declaration

None of the authors have conflicts of interest in the research reported here.

GENOME CANADA TRANSPLANT CONSORTIUM MEMBERS

Keown, P. Sapir-Pichhadze, R. Bryan, S. Caulfield, T. Ragoussis, J. Oualkacha, K. Tinckam, K. Liwski, R. Campbell, P. Cardinal, H. DeSerres, S. Allan, L. Saw, CL. Mengel, M. Sis, B. Sherwood, K. Wagner, E. Berka, N. McManus, B. McMaster, R. Hebert, MJ. Foster, L. Rossi, F. Borchers, C. Piccirillo, C. Polychronakos, C. Ng, R. Jevnikar, A. Cullis, P. Filler, G. Wong, H. Foster, B. Gill, J. Kim, J. Tibbles, L. Humar, A. Lan, J. Shechter, S. Chaudhury, P. Fernandez, N. Fowler, E. Kiberd, B. Tennankore, K. Gill, J. Fortin, MC. Klarenbach, S. Balshaw, R. Mital, S. Mucsi, I. Ostrow, D. Stiller, C. Parekh, R. Richard, L. Senecal, L. Blydt-Hansen, T. (Canada); Duquesnoy, R. Erlich, H. Gebel, H. Weimer, ET. Kaplan, B. Burckart, G. (United States); Middleton, D. Marsh, SGE. (United Kingdom); Tilanus, M. Claas, F. van Gelder, T. (Netherlands); Opelz, G. Oellerich, M. (Germany); Marquet, P. (France); Marra, C. (New Zealand); Kalo, Z. (Budapest)

Figures

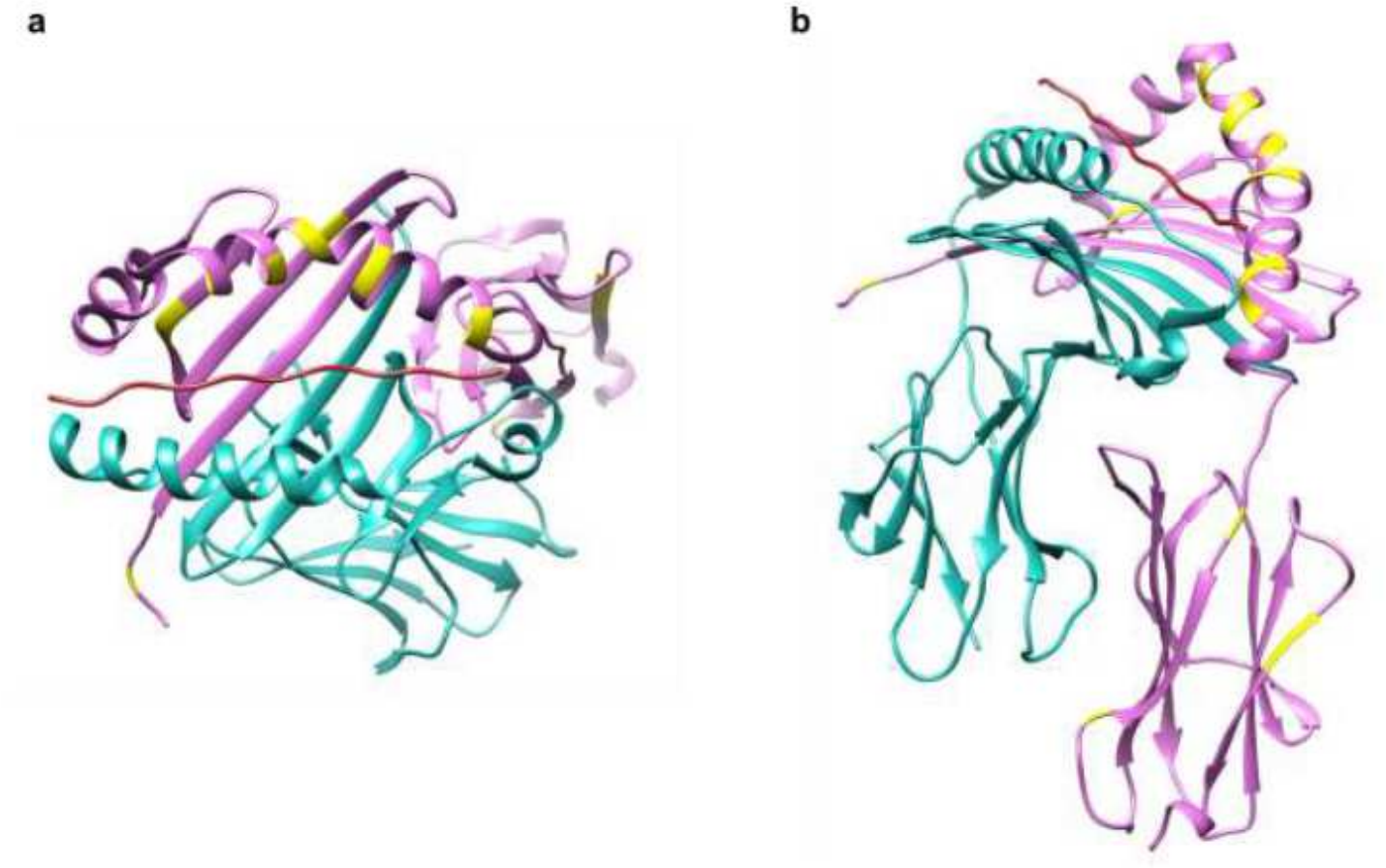


Figure 1

A 3D visualization of an HLA-DR protein with eplets highlighted. The DRA chain is depicted in aqua, DRB1*04:01 chain is depicted in magenta, and processed peptide in dark red. All 11 eplets are highlighted in yellow. Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 13 using the Protein Data Bank ID: 5NI9 41. (a) Top-down view of the protein, depicting the peptide-binding groove. (b) Side-view of the protein, depicting its full extracellular portion, showing eplets present outside of the peptide-binding groove.

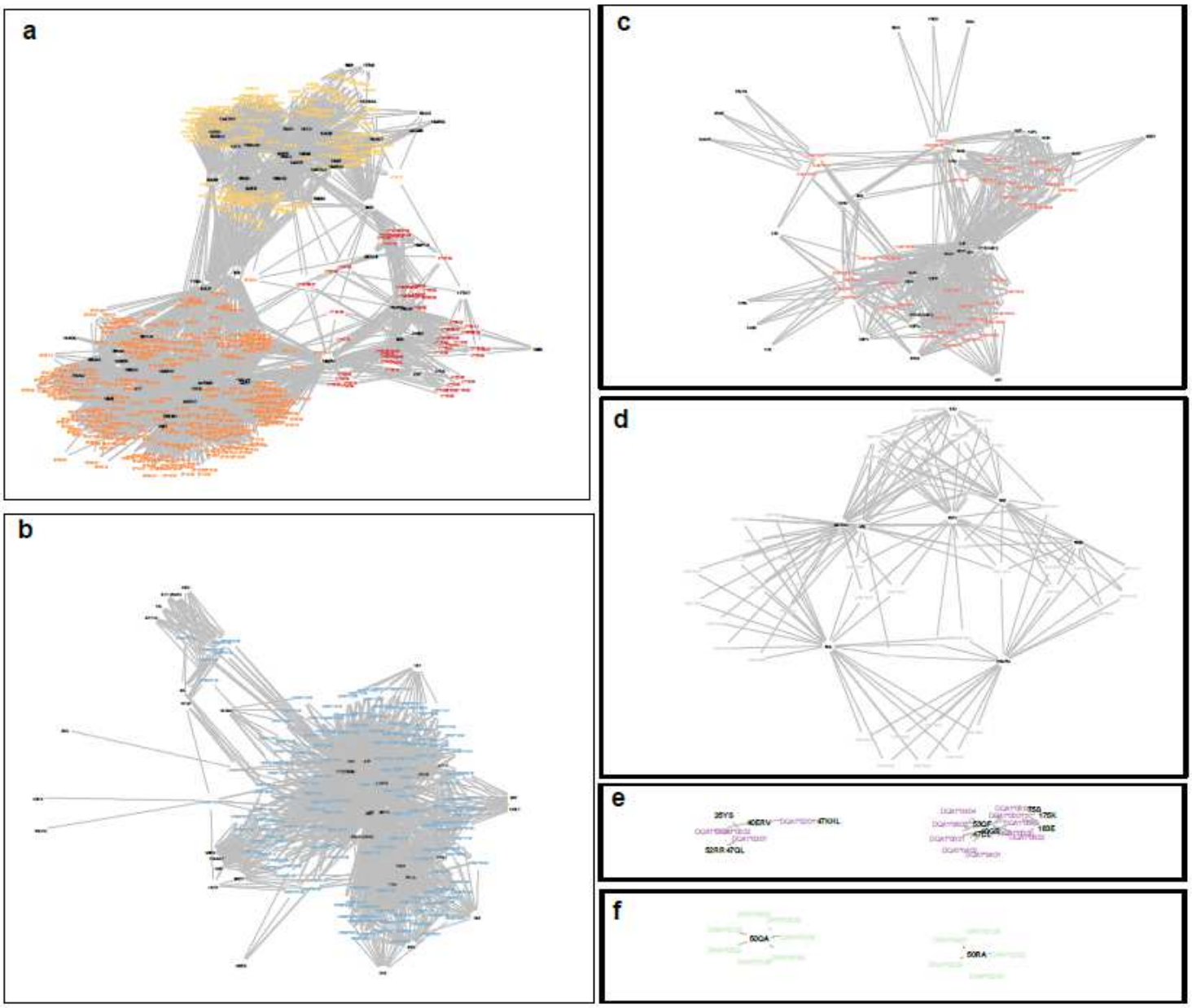


Figure 2

Bipartite networks 42 depicting the associations between the complete library of HLA alleles and eplets present in HLAMatchmaker. Each node represents an allele or eplet and lines represent either that the eplet is encoded by an allele or vice versa. (a) HLA Class I, (b) DRB1/3/4/5, (c) DQB1, (d) DPB1, (e) DQA1, and (f) DPA1.

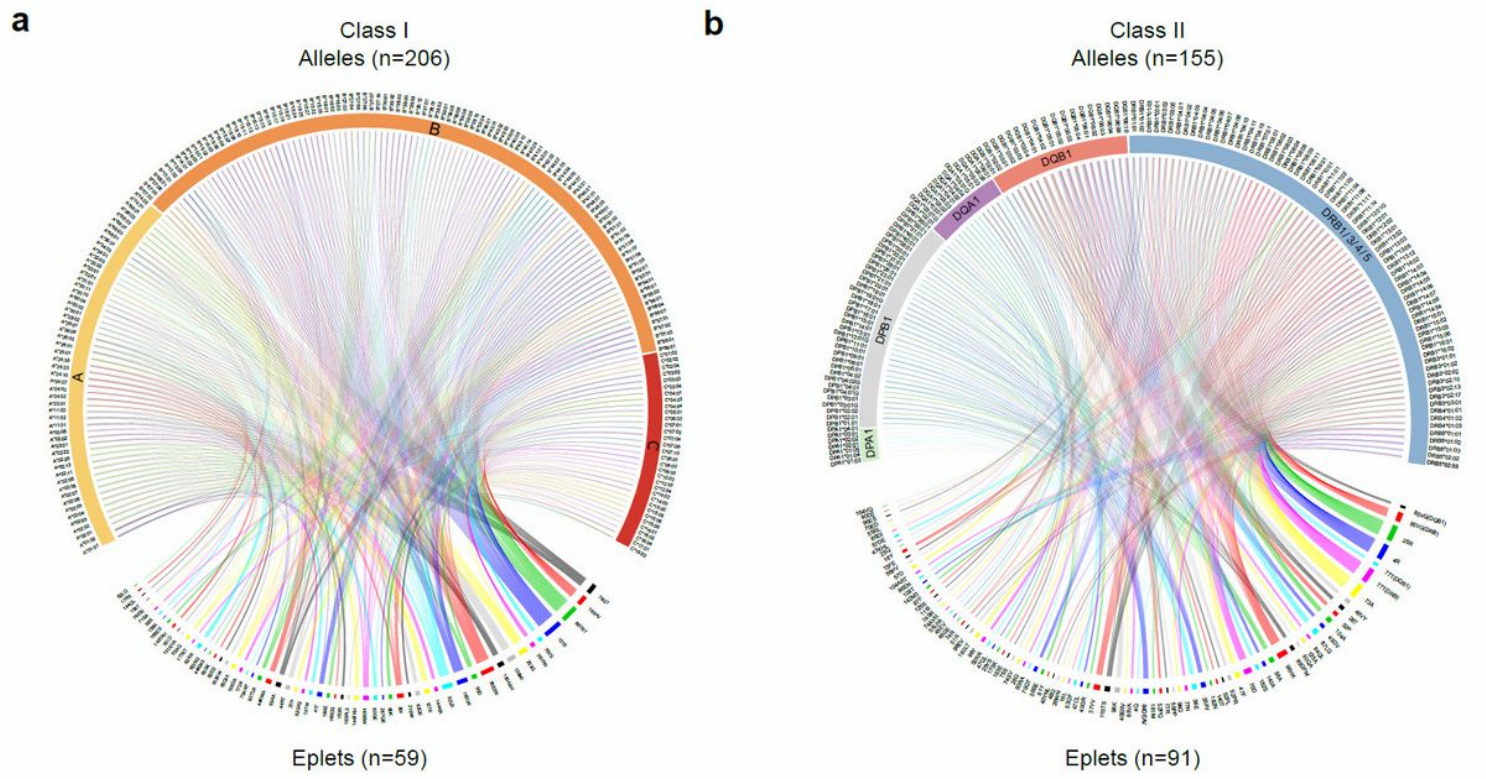


Figure 3

Circos 43,44 diagrams depicting the reduction of HLA complexity of identified HLA alleles (top portion) in the study population converting into eplets (bottom portion). Conversion of alleles to eplets was determined by HLAMatchmaker. HLA alleles identified in the study population are depicted at the top portion and eplets shown in the bottom portion. Interconnections represent an allele encoding an eplet. (a) Class I alleles to eplets and (b) Class II alleles to eplets.

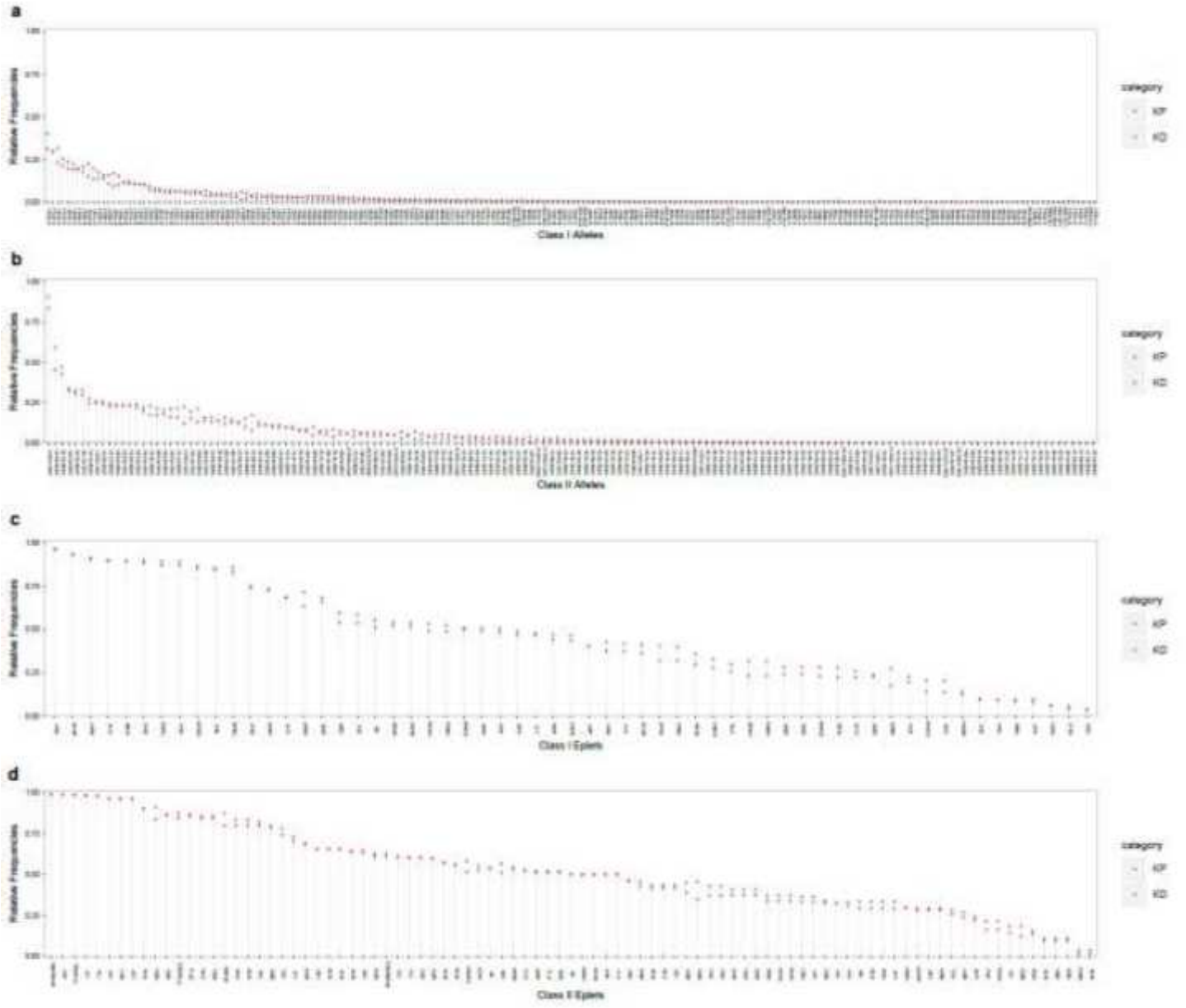


Figure 4

The relative frequencies of HLA alleles and eplets in the study population, calculated as the proportion of subjects expressing a particular allele or eplet. Images (a) and (b) depict the allele frequencies by class I and II, respectively. (c) and (d) depict eplet frequencies by class I and II, respectively. KP = Kidney Patients, KD = Kidney Donors.

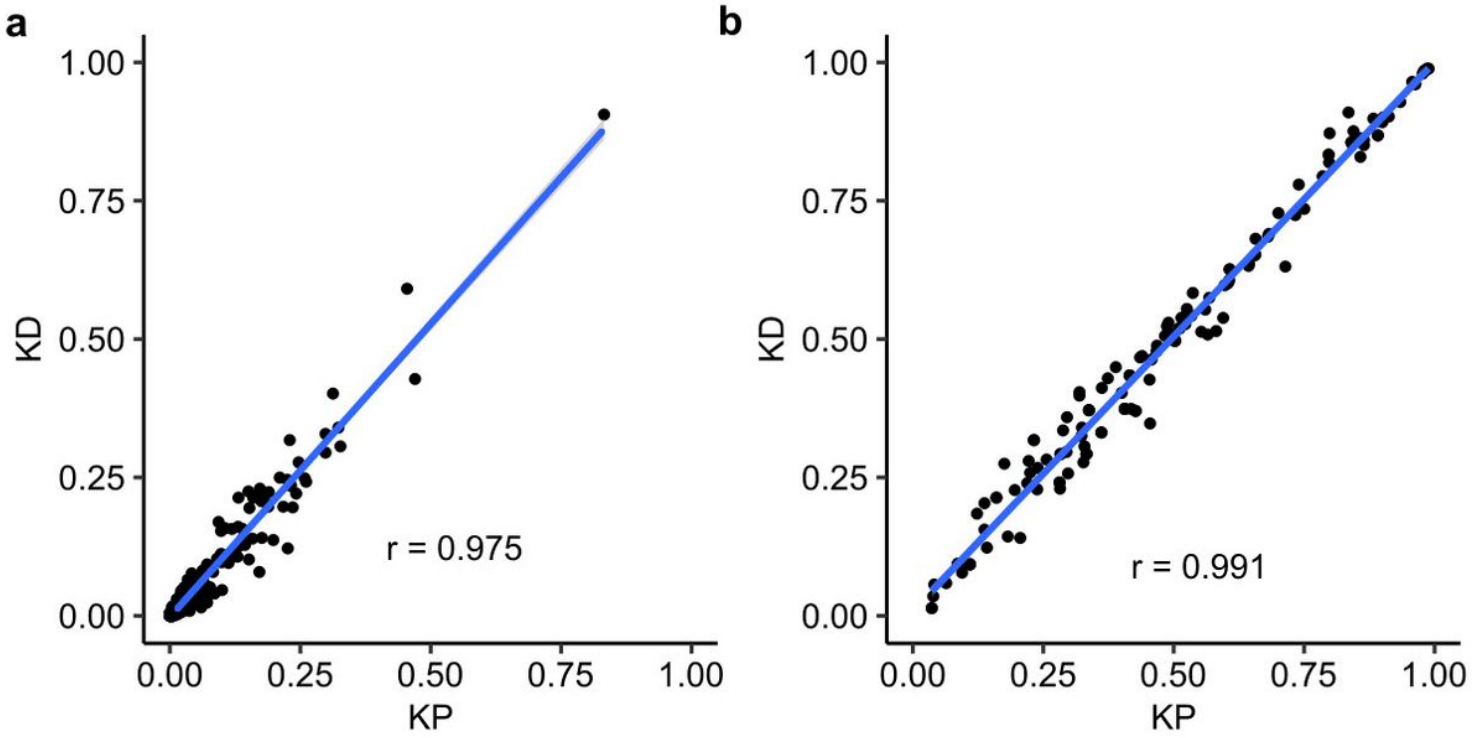


Figure 5

Pairwise analysis between total patient and total donor groups of HLA allele and eplet frequencies calculated as the proportion of subjects expressing a particular allele or eplet. The line and correlation coefficient were calculated using Pearson Correlation. (a) Each dot represents an allele and its frequency for a particular patient or donor group is plotted against another patient or donor group. (b) As above but for eplets. KP = kidney patients (n = 1049), KD = kidney donors (n = 797).

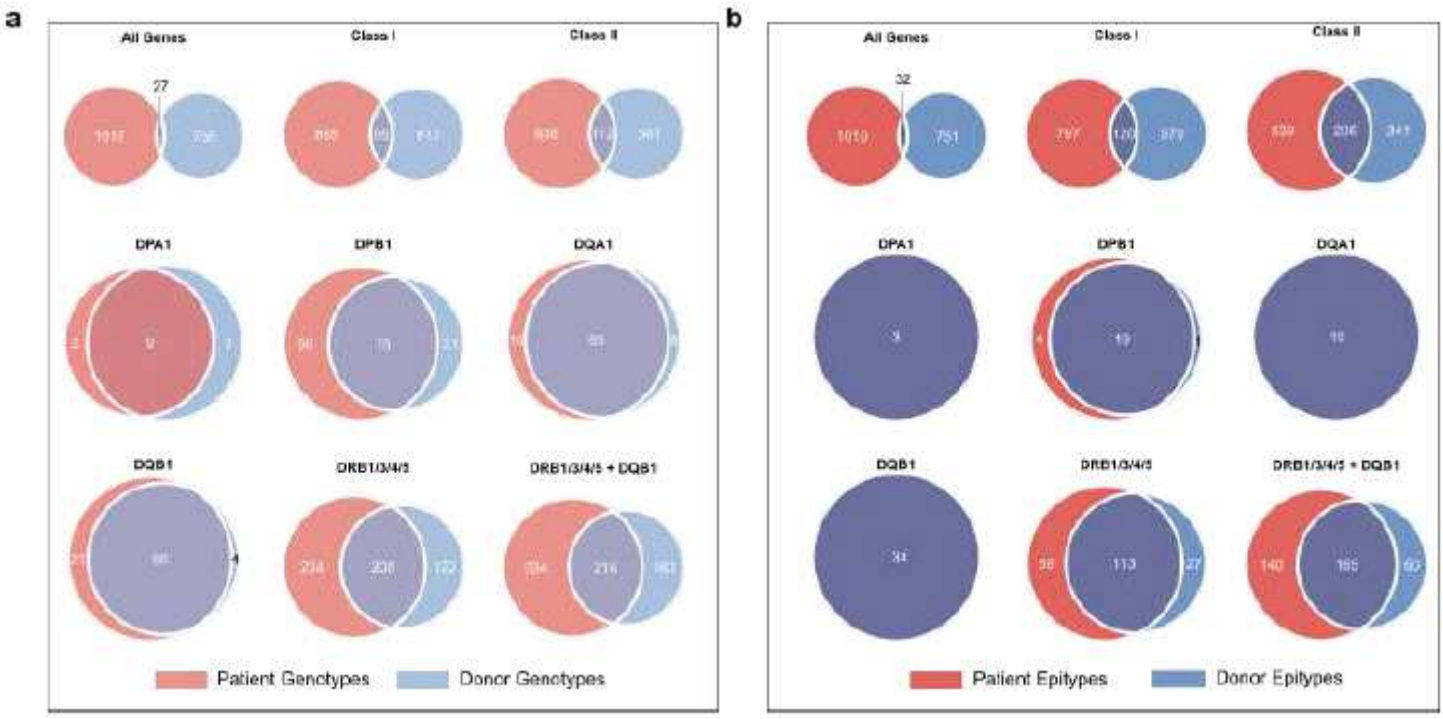


Figure 6

Genotype and epitype frequencies between patients and donors. (a) Venn diagrams of unique genotypes and how some or all occur between patient and donor groups. Genotypes were determined at the individual level, where alleles of a particular loci were sorted and combined. Analysis was performed on various loci combinations: all genes, HLA class I, HLA class II, DPA1, DPB1, DQA1, DQB1, DRB1/3/4/5, and DQB1 and DRB1/3/4/5 combined. (b) As above but for epitypes.

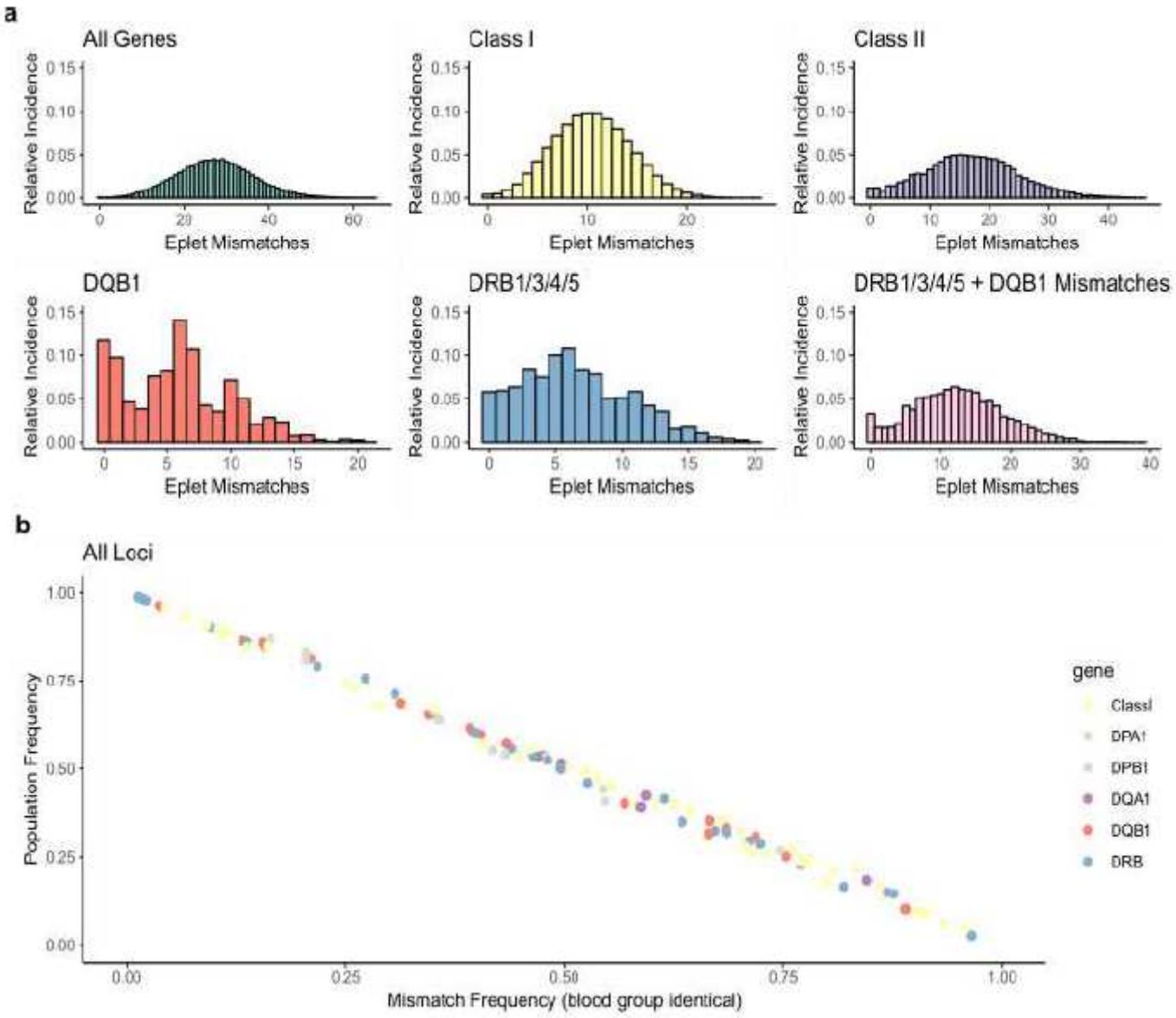


Figure 7

Characterization of eplet mismatches in all blood group identical matches between kidney patients and deceased donors. 1049 kidney patients were matched against 243 deceased donors by blood-group identity. Next, eplet mismatches across all genes (A, B, C, DRB1, DRB3, DRB4, DRB5, DPA1, DPB1, DQA1, and DQB1), HLA class I, HLA class II, DQB1 only, DRB1/3/4/5, and DRB1/3/4/5 and DQB1 were calculated. A mismatch was determined as an eplet present in a donor that is not present in the patient. (a) Distribution of the relative incidences of eplet mismatch scores at all gene combinations analyzed.

Relative incidence was calculated by the proportion of a particular quantitative eplet mismatch out of the total number of 92,756 blood-group identical matches. (b) The linear correlation (Pearson's) of population frequency versus mismatch frequency for individual eplets. A dot represents an individual eplet colour-coded by gene combination, and all identified eplets are plotted (n = 150). Population frequency is the proportion of individuals (kidney patients and donors) whom express the eplet, out of the total study population (n = 1846). Mismatch frequency was the incidence of a particular eplet being mismatched in blood-group identical matches, divided by the sum of the number of donors with the specific eplet multiplied by the number of patients, restricted by blood group. The correlation coefficient (r) is -0.998.

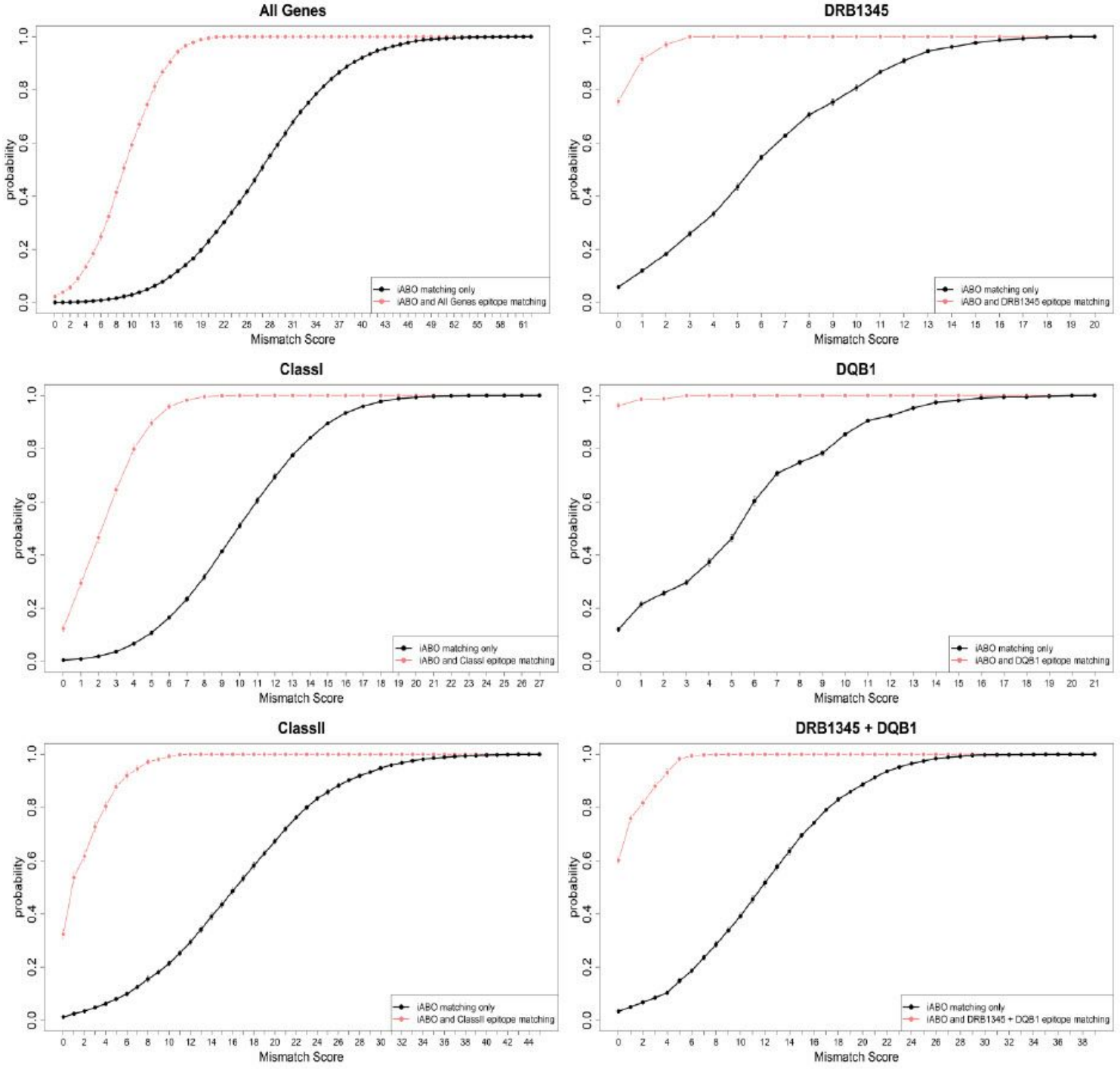


Figure 8

Matching simulations incorporating deliberate eplet and blood group identical matching (red curve) and baseline blood group identical matching only (black curve) in kidney patients and deceased donors. Plots represent simulations using Canadian waitlist and deceased donor data across all genes, HLA class I, HLA class II, DQB1, DRB1/3/4/5, and DQB1+DRB1/3/4/5 combined. Eplet mismatch scores for their respective genes on the x axis plotted against the cumulative probability of these scores in the matched population. Error bars are calculated as the standard deviation of 10 repeated simulation runs.

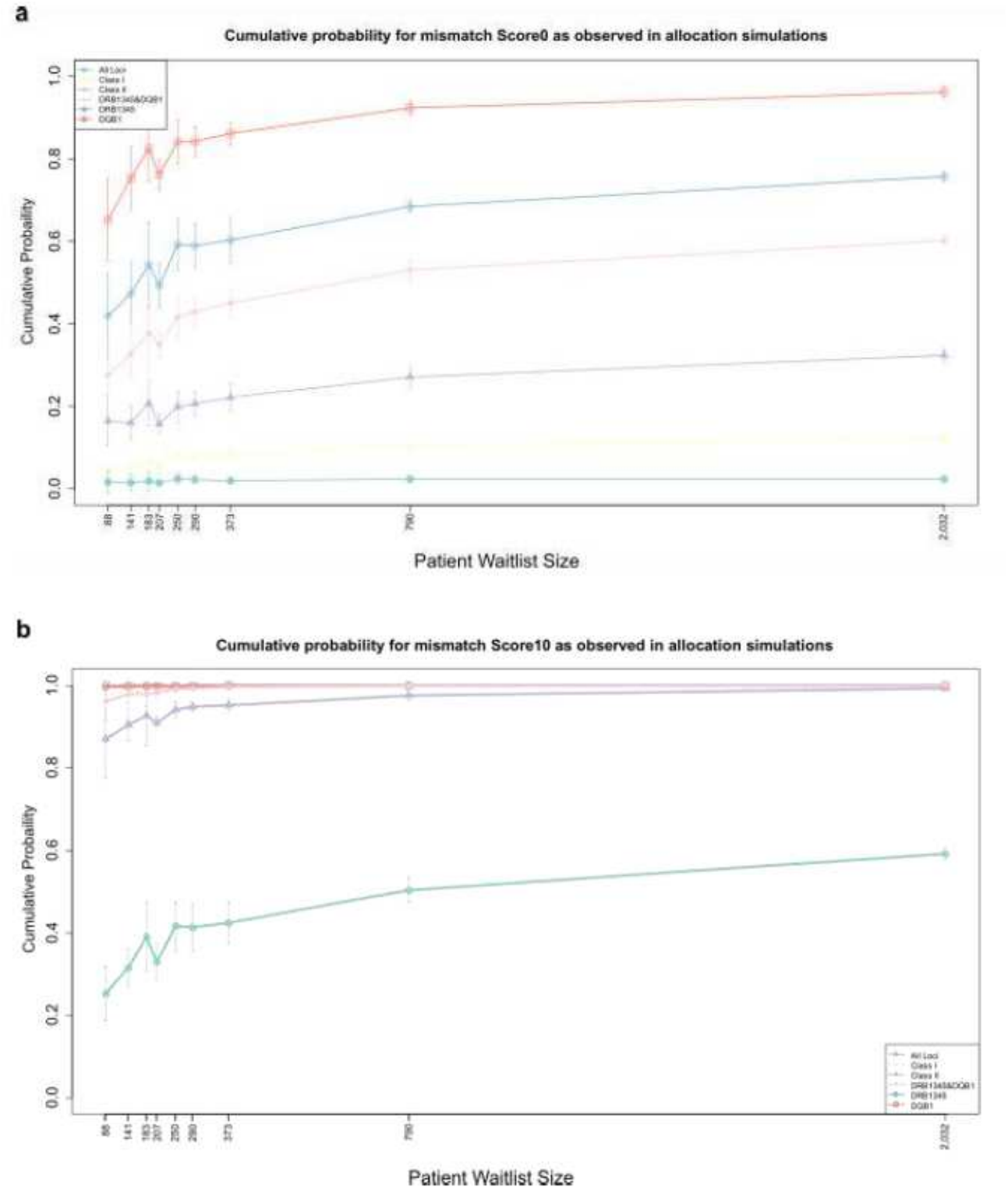


Figure 9

Results of an averaged set of 10 deliberative eplet-matching simulations at various gene loci. Provincial active waitlist and deceased donor numbers were used according to 3. X axis shows the number of patients on provincial or national patient waiting lists (Saskatchewan 88, Atlantic provinces 141, Manitoba 183, British Columbia 207, Alberta 50, Average Province 290, Quebec 373, Ontario 790, Canada 2,032). Y axis shows the averaged cumulative probability of achieving a total mismatch score of 0 (a) or 10 or lower (b) at the end of the simulations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalFigure1omittedsubjects.jpg
- SupplementalFigure2Linearregreesionallgroups.jpg
- SupplementalFigure3HeatMaps.jpg
- SupplementalFigure4AverageProvinceSimulation.jpg
- SupplementalTable1.xlsx
- SupplementalTable2.pdf
- SupplementalTable3.xlsx
- SupplementalTable4.xls
- SupplementalTable5.xlsx