Human Immunology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Human Immunology

journal homepage: www.elsevier.com/locate/humimm

Predicting an HLA-DPB1 expression marker based on standard DPB1 genotyping: Linkage analysis of over 32,000 samples

Bianca Schöne[a], Sabine Bergmann[a], Kathrin Lang[a], Ines Wagner[a], Alexander H. Schmidt[a,b], Effie W. Petersdorf[c], Vinzenz Lange[a,*]

[a] DKMS Life Science Lab, Blasewitzer Str. 43, 01307 Dresden, Germany
[b] DKMS, Kressbach 1, 72072 Tübingen, Germany
[c] University of Washington, Fred Hutchinson Cancer Research Centre, Seattle, WA, United States

ARTICLE INFO

Keywords:
HLA-DPB1
Gene expression
SNP
Linkage
Genotyping
GvHD
rs9277534
Permissive mismatch

ABSTRACT

The risk of acute graft-versus-host disease (GvHD) after hematopoietic stem cell transplantation is increased with donor-recipient HLA-DPB1 allele mismatching. The single-nucleotide polymorphism (SNP) rs9277534 within the 3′ untranslated region (UTR) correlates with HLA-DPB1 allotype expression and serves as a marker for permissive HLA-DPB1 mismatches. Since rs9277534 is not routinely typed, we analyzed 32,681 samples of mostly European ancestry to investigate if the rs9277534 allele can be reliably imputed from standard DPB1 genotyping. We confirmed the previously-defined linkages between rs9277534 and 18 DPB1 alleles and established additional linkages for 46 DPB1 alleles. Based on these linkages, the rs9277534 allele could be predicted for 99.6% of the samples based on DPB1 genotypes (99.99% concordance). We demonstrate that 100% prediction accuracy could be achieved if the prediction utilized exon 3 sequence information. DPB1 genotyping based on exon 2 data alone allows no unambiguous rs9277534 allele prediction but was estimated to maintain 99% accuracy for samples of European descent. We conclude that DPB1 genotyping is sufficient to infer the DPB1 expression marker rs9277534 with high accuracy. This information could be used to select donors with permissive HLA-DPB1 mismatches without directly screening for rs9277534.

1. Introduction

Hematopoietic stem cell transplantation (HSCT) from unrelated donors can cure various blood disorders [1]. A high level of donor-recipient HLA compatibility is crucial for the success of HSCT. Currently, matching for HLA-A, -B, -C, -DRB1 and -DQB1 alleles (10/10 match) is the gold standard for the selection of unrelated donors, while HLA-DPB1 is often not considered [2]. This leads to DPB1 mismatching in 80–86% of otherwise HLA-matched donor-recipient pairs [3,4]. Mismatching of HLA-DPB1 between donor and recipient is associated with a significantly higher risk of acute GvHD [5], which is a major impediment to successful transplantation.

Recent studies indicate that not all DPB1 mismatches confer equal risks after transplantation [6–9]. The presence of specific amino acid residues within hypervariable regions defined by DPB1 exon 2 has been shown to play a role in alloreactivity and risk of GvHD after transplantation [7,8]. In addition to a role for the recognition of T-cell epitopes, the level of HLA-DP expression in the patient influences the incidence of GvHD after HSCT, where mismatching against a highly-expressed patient allele leads to significantly higher GvHD risk compared to other combinations [9].

The expression level of an HLA-DPB1 allele was shown to be correlated with variant rs9277534 [10], an A → G SNP located within the 3′ UTR of DPB1 (Fig. 1A). This SNP therefore may serve as marker for HLA-DPB1 expression level even though the SNP itself is most likely not directly involved in DPB1 expression control. Conserved haplotypes of DPB1 exon 2 and rs9277534 were defined for common DPB1 alleles by direct phasing [9]. The rs9277534-A allele is associated with low DPB1 expression, whereas the rs9277534-G allele is associated with high DPB1 expression. When DPB1-matched donors are not available, this expression marker can be used to prospectively identify DPB1-mismatched donors who generate a permissive DPB1 mismatch against low-expression patient DPB1 alleles [9].

The 3′ UTR containing rs9277534 is currently not covered by

Abbreviations: GvHD, graft-versus-host disease; SNP, single nucleotide polymorphism; UTR, untranslated region; HSCT, hematopoietic stem cell transplantation; SMRT, single molecule real-time [sequencing

* Corresponding author.
E-mail addresses: schoene@dkms-lab.de (B. Schöne), bergmann@dkms-lab.de (S. Bergmann), lang@dkms-lab.de (K. Lang), schmidt@dkms.de (A.H. Schmidt), epetersd@fredhutch.org (V. Lange).

https://doi.org/10.1016/j.humimm.2017.11.001
Received 14 June 2017; Received in revised form 28 September 2017; Accepted 3 November 2017

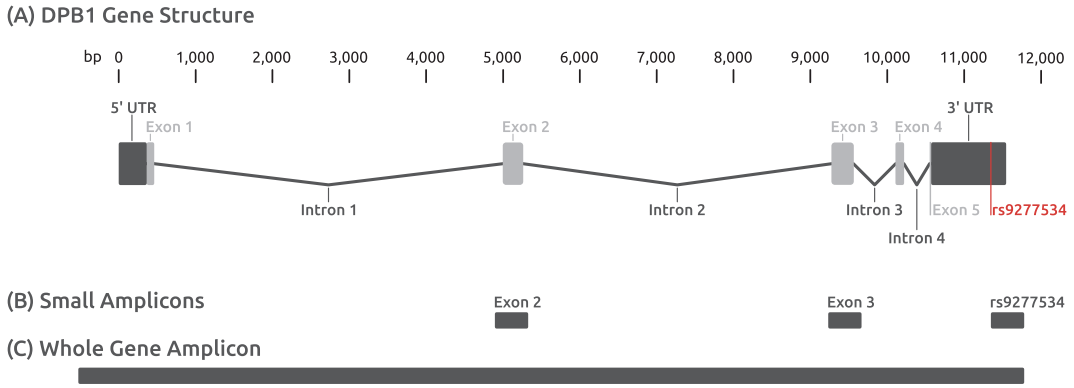

Fig. 1. HLA-DPB1 gene structure and regions analyzed in this study (schematic, to scale). (A) Gene structure of DPB1, including both UTRs (black boxes), exons (grey boxes), and introns (black lines). SNP rs9277534 (red) is located towards the end of the 3′ UTR. Positions are given relative to the start of the 5′ UTR of HLA-DPB1*02:01:02.(B) Selected gene areas were targeted with short amplicons (black boxes) for the primary analysis by Illumina sequencing.(C) The whole gene was targeted with a long amplicon (black box) used for secondary analysis by SMRT sequencing.

routine genotyping assays for HLA-DPB1. An additional screening would increase the cost of HLA genotyping in the clinical setting for donor selection. Therefore, the aim of our study was to verify if the rs9277534 genotype can be accurately predicted from high-resolution DPB1 genotyping. For this purpose, we examined DPB1 exon 2, exon 3, and rs9277534 as separate amplicons in 32,681 individuals using our lab's high throughput NGS workflow based on sequencing with Illumina MiSeq and HiSeq instruments (Fig. 1B) [11,12]. In addition, we applied full-length Single Molecule Real-Time (SMRT) sequencing for a small subset of 22 samples with unexpected linkage patterns after primary analysis. Our full-length sequencing approach covers the whole DPB1 gene including rs9277534 in the 3′ UTR in one amplicon (Fig. 1C) and therefore directly delivers the requisite information for phasing coding and non-coding polymorphisms.

2. Materials and methods

2.1. Samples

Samples were provided by DKMS donor centers in Germany (69.3%), USA (13.2%), UK (11.9%), and Poland (5.2%) and by the BMST (Bangalore Medical Services Trust) donor center in India (0.4%) between March 2016 and May 2016. The samples approximately reflect the ethnic diversity encountered in these countries. (Detailed information about the diversity of provenance for each donor center is listed in Supplementary Material 1.) The majority of samples (98%) were collected with nylon FLOQSwabs™ hDNA free (Copan Italia Spa, Brescia, Italy) and 2% as 150 µl whole blood in Venosafe 4 ml EDTA tubes (Terumo, Tokyo, Japan). During registration, donors signed an informed consent approving HLA genotyping and other analyses to facilitate or improve donor search for stem cell transplantation. No ethics committee approval was obtained as the described genotyping is within the scope of this consent form and performed as a genotyping service.

2.2. Isolation and quantification of DNA

DNA was isolated from buccal swabs or whole blood using the magnetic-bead-based "Chemagic DNA Buccal Swab kit Special" or "Chemagic DNA Blood Kit Special" (Perkin Elmer, Baesweiler, Germany), respectively. The isolated DNA was eluted in 100 µl elution buffer (10 mM Tris-HCl pH8.0), and its concentration was measured by fluorescence (SYBR Green, Biozym, Hessisch Oldendorf, Germany) using the TECAN infinite 200Pro plate reader (Tecan, Männedorf, Switzerland). Samples with DNA concentrations of < 2 ng/µl were excluded from genotyping.

2.3. Amplicon-based sequencing of HLA-DPB1 exons 2 and 3 and rs9277534

2.3.1. PCR

Exons 2 and 3 of the HLA-DPB1 gene were amplified in a multiplexed PCR as previously described [11], resulting in amplicons about 340 bp in length. The rs9277534 region was targeted by an additional primer group (forward: GAATTGACTGTATTTCAGTGAGCTGCC, reverse: ACATGTATTGCTTTGCTCTTTCCCCAG and ACGTATTGCTTTGCTCTTTCCCCAG), resulting in PCR products about 425 bp in length. (The length and position of all three amplicons is depicted in Fig. 1B.)

The two DPB1 PCR reactions were performed alongside PCRs for several other loci in 10 µl volume each, using 384-well plates with FastStart™ Taq DNA Polymerase (Roche, Basel, Switzerland) and the associated buffer system [12]. All amplification products belonging to one sample were pooled using volumes appropriate to obtain balanced read coverage for each amplicon. A secondary PCR was performed on the pool to elongate the amplicons with indexes and sequencing adapters for Illumina sequencing. Target-specific primers and indexing primers were obtained from Metabion (Metabion International AG, Planegg, Germany).

2.3.2. Sequencing

After indexing PCR, 384 barcoded samples were pooled together. Pooled PCR products were purified with SPRIselect beads (BeckmanCoulter, Brea, USA) with a ratio of 0.6:1 beads to DNA, and subsequently quantified by qPCR. Denaturation and dilution of the sequencing library were executed as recommended by Illumina (MiSeq Reagent Kit v2-Reagent Preparation Guide). Libraries with commonly 384 or 3840 samples were loaded at 12.5 pM onto MiSeq or HiSeq flow cells, respectively, with 10% PhiX spiked in. Paired-end sequencing was performed for 2 times 249 cycles with MiSeq Reagent Kit v3 or HiSeq Rapid SBS Kit v2 (Illumina, San Diego, USA) on MiSeq or HiSeq, respectively.

2.3.3. Genotyping

DPB1 alleles were assigned from exons 2 and 3 sequences using the genotyping software neXtype as previously described [11]. rs9277534 alleles were interpreted with FasType, developed in Python. FasType uses the GEM mapper [13] for mapping reads against reference sequences. According to the allele sequences in the IPD-IMGT/HLA database, all currently described full-length DPB1 sequences converge to two distinct sequences at the rs9277534 amplicon region. These two sequences differ by seven individual SNPs and were used as reference for rs9277534. Samples with less than 50 reads coverage on any of the three amplicons (exon 2, exon 3, rs9277534) were not included in the