PHASED FULL-LENGTH SMRT SEQUENCING OF HLA-DPB1

Kathrin Lang1, Gerhard Schöfl1, Carolin Zweiniger1, Maarten Penning2, Erik Rozemuller2, Sylvia Clausing2, Yannick Duport3,4, Nicola Gscheidl4, Sylke Winkler4, Vinzenz Lange1, Inna Böhme1, Alexander Schmidt1,5
1DKMS Life Science Lab, Dresden, Germany; 2GenDx, Utrecht, The Netherlands; 3CRTD - Center for Regenerative Therapies Dresden, Deep Sequencing Group, Dresden, Germany; 4Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany; 5DKMS German Bone Marrow Donor Center, Tübingen, Germany

Aim
In contrast to exon-based HLA-typing approaches, whole gene genotyping crucially depends on full-length sequences submitted to the IMGT/HLA database. Currently, full-length sequences are known for only 12 out of 550 HLA-DPB1 alleles (as of July 2015). Here, we present a whole-gene sequencing approach for DPB1 that allows full phase resolution to facilitate further exploration of the allelic structure at this locus.

Methods

Primer Design & PCR

![Diagram of Primer Design & PCR]

N G S L i b r a r y P r e p a r a t i o n f o r S M R T S e q u e n c i n g

![Diagram of NGS Library Preparation for SMRT Sequencing]

D a t a A n a l y s i s & H L A-T y p i n g

![Diagram of Data Analysis & HLA-Typing]

Results

We analyzed DPB1 for a set of 48 randomly picked donor samples. With 3 exceptions due to PCR failure, all genotype assignments conformed to previous typing results based on exon 2 and 3 short read sequencing. Allelic proportions for SMRT-sequencing-derived heterozygous positions were evenly distributed for all samples (range 0.4 - 0.6), suggesting unbiased long-range amplifications.

To verify PacBio read data, we also conducted standard 2x250 paired-end shotgun sequencing (Illumina MiSeq). Despite the high per-read raw error rates typical for SMRT sequencing (~15%), this comparison indicates an overall high level of agreement between the two sequencing technologies (Figure 1). Nevertheless, discrepancies arise at known problematic genomic positions and within specific sequence motifs (e.g. microsatellites and homopolymer stretches).

We describe novel intronic sequence variation for 5 previously described whole-length DPB1 alleles (Table 1). Additionally, we gathered whole-length sequences for 9 DPB1 alleles with so far unknown introns (Table 2). One of these alleles (HLA-DPB1*131:01) is classified as rare (Figure 2).

Conclusion
Here we present a whole gene amplification and sequencing workflow for DPB1 alleles utilizing single molecule real-time (SMRT) sequencing from Pacific Biosciences. Validation of consensus sequences against known exonic sequences highlights the reliability of this technology. This workflow will facilitate amending the IMGT/HLA Database for DPB1.