# Coding and Non-Coding Variation in 74 Novel Full-Length Sequences of KIR2DL1

**Steffen Klasberg**[1], Kathrin Lang[1], Marie Günther[1], Alexander H. Schmidt[2], Vinzenz Lange[1], Gerhard Schöfl[1]

[1]**DKMS** Life Science Lab • Blasewitzestr. 43 • 01307 Dresden
[2]**DKMS** gGmbH • Kressbach 1 • 72072 Tübingen

## Introduction

KIR2DL1 is a member of the killer-cell immunoglobulin-like receptor (KIR) family, which is an important factor of the human immune system. KIR genes are key regulators of natural killer cell activity and partly bind to proteins of the human leukocyte antigen (HLA) family, e.g. HLA-B and HLA-C in the case of KIR2DL1. Likely due to the complexity of the KIR locus with its extensive genetic variation, only little is known about the impact of allelic variation. Here, we identified novel KIR2DL1 alleles by routine high-throughput exon-based KIR genotyping and subsequently created full-length reference sequences. We analysed coding and non-coding variation to identify patterns of variation between alleles.

## Methods

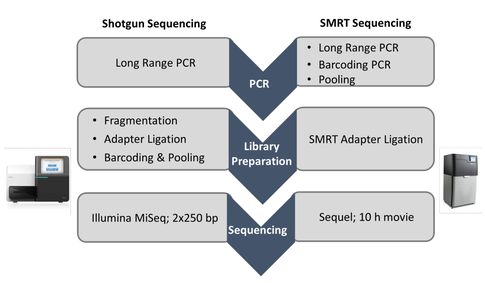


**Figure 1.** *Sequencing workflow*

We sequenced whole 16 kb amplicons of the complete KIR2DL1 gene, including the **5′**- and **3′**-UTRs, using shotgun sequencing (Illumina MiSeq) and single molecule real time (SMRT) sequencing (PacBio Sequel). Using the R package DR2S[1], we combined phase information from the long SMRT sequencing reads with the accuracy of shotgun sequencing to generate high-quality phased full-length sequences (**Figure 1**).

[1]Developed at DKMS LSL and available on GitHub (`https://github.com/DKMS-LSL/dr2s`)

## Results

We successfully generated 74 distinct KIR2DL1 allele sequences from 58 specifically selected samples. These include 54 novel alleles, 15 distinct alleles that were previously only partially characterised and confirmed 5 extant alleles. The sequence analysis revealed in total 301 SNPs, of which 143 are separating the sequences into two deep allelic groups (**Figure 3**). We took the nucleotide at position 151 of exon 5, being either A or C, as a marker for defining the groups. The two groups are mostly separated by non-coding SNPs, indels, and a T homopolymer of fixed length in one and variable length in the other group. Some separating variants exist in most exons and alter the amino acid sequence of the leader peptide, the extracellular D1- and D2-domains containing the HLA binding sites as well as the cytoplasmatic domain (**Figure 2**). The C group (76.34%) dominantes our large European population sample relative to the A group (16.10%). 7.56% of alleles were not assigned to a group because they were identified as novel alleles during our routine genotyping (**Figure 4**.

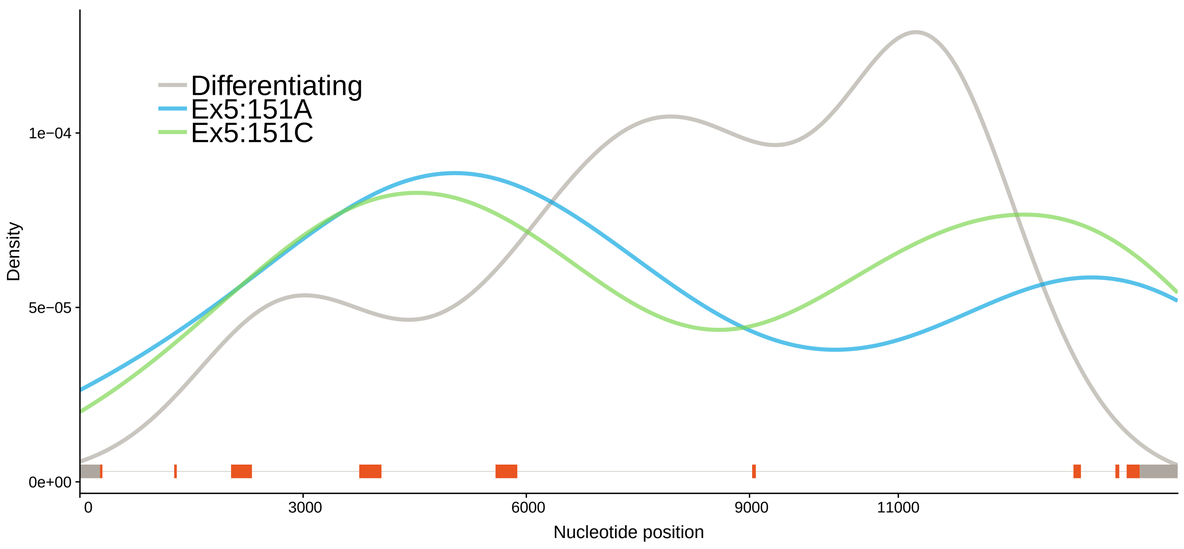


**Figure 2.** *Probability density of SNPs along the gene sequence of KIR2DL1. "Differentiating" describes positions that differ only between groups, but not within either of groups.*

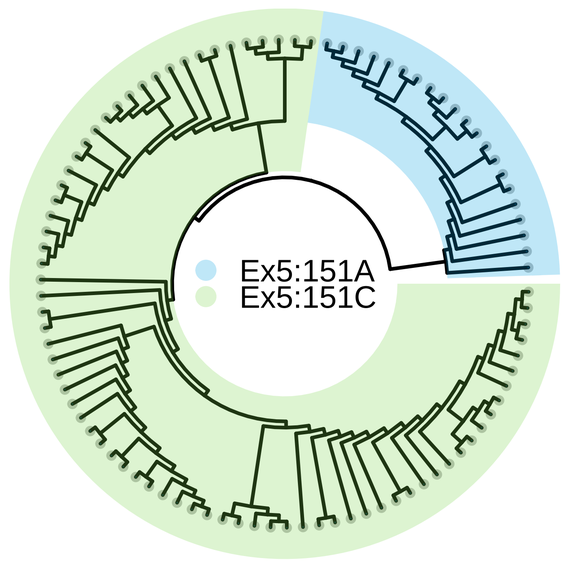


**Figure 3.** *Phylogenetic distribution of all KIR2DL1 full length alleles. The Ex5:151A and Ex5:151C clades are clearly separated.*

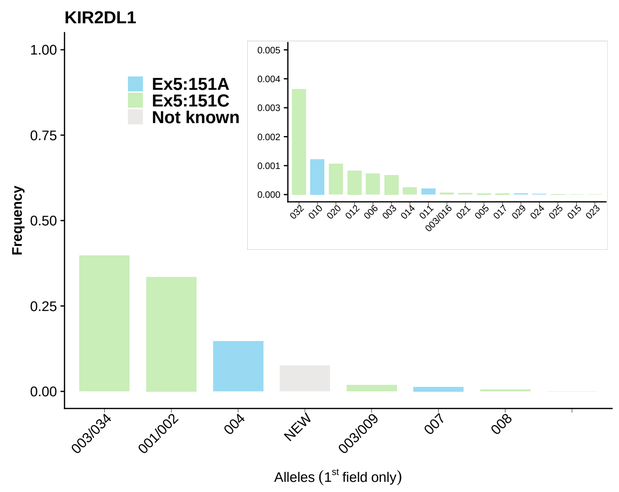


**Figure 4.** *Allele frequencies observed in our routine genotyping. The `NEW` category describes a mixture of several unknown KIR2DL1 alleles, which will to a great extent be resolved by future database releases.*

## Conclusions

Our sequencing efforts resulted in a more than 3-fold increase in known full-length sequences of KIR2DL1, enabling further research on this specific KIR gene. We gained insights into systematic differences at the sequence level which might be responsible for or indicative of medically relevant allelic differences. Especially variation in the D2-domain has been shown to be involved in binding to HLA proteins and may as such be clinically relevant.

DKMS Life Science Lab
Blasewitzerstr. 46
01307 Dresden, Germany
www.dkms-lab.de