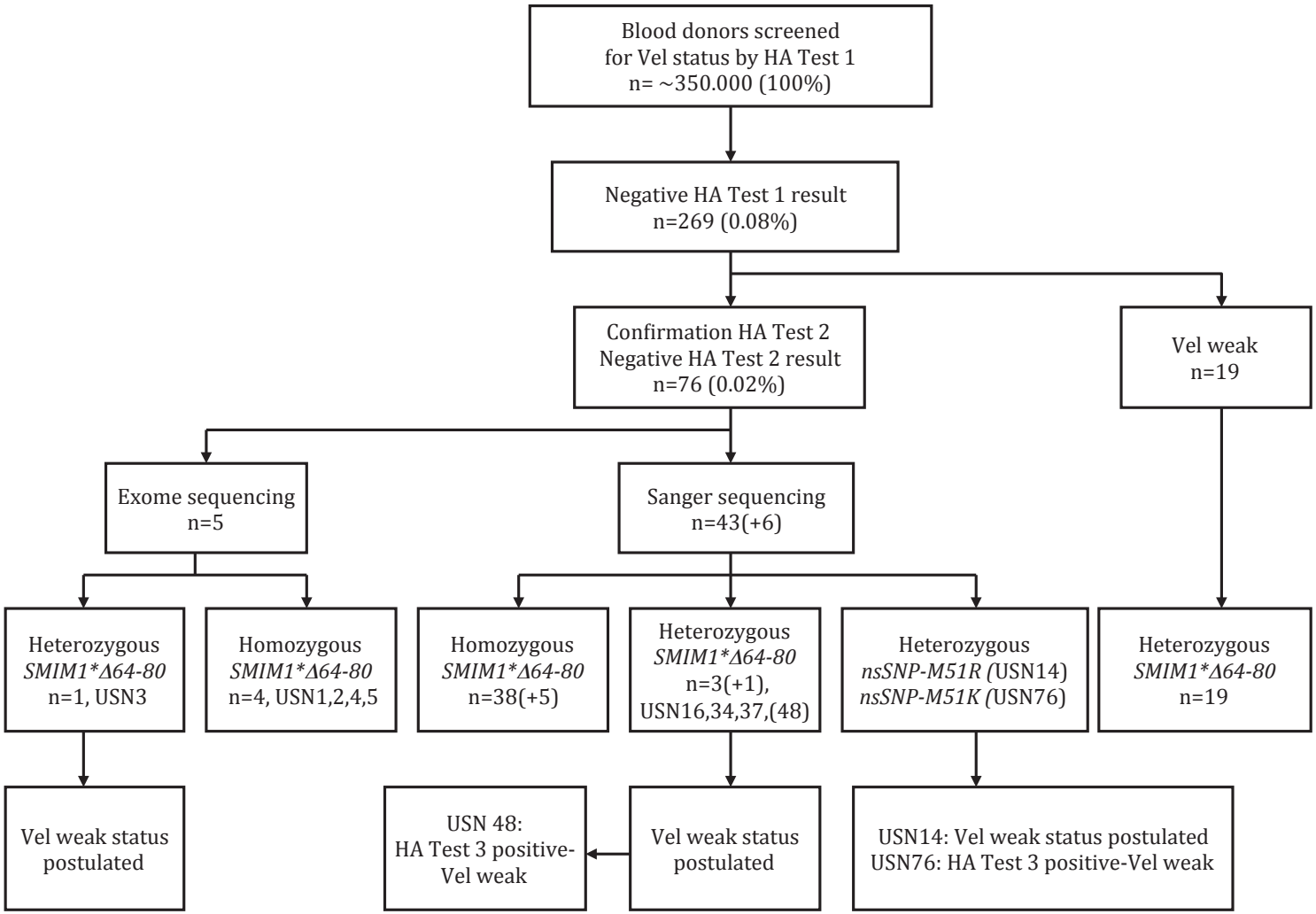
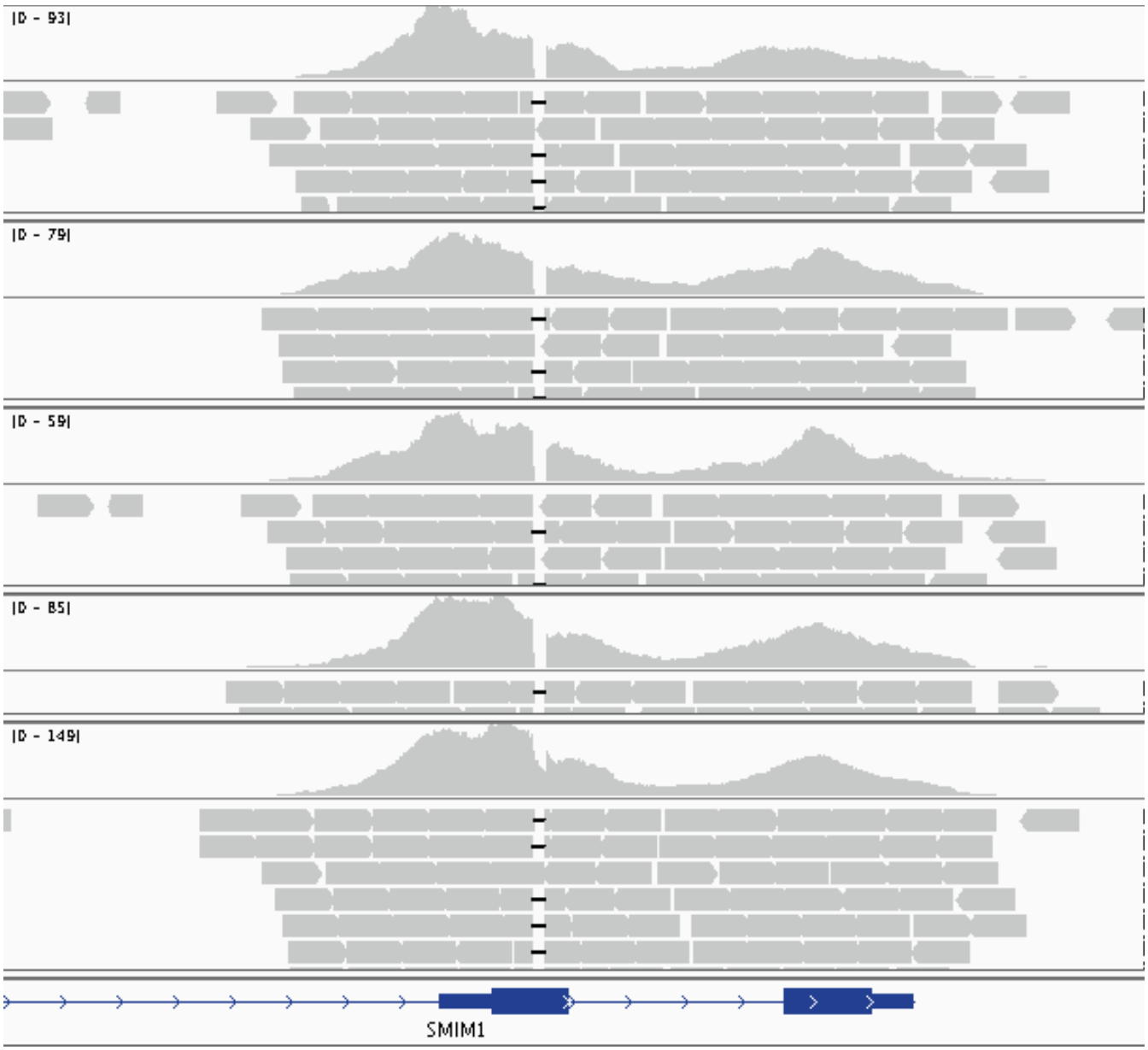
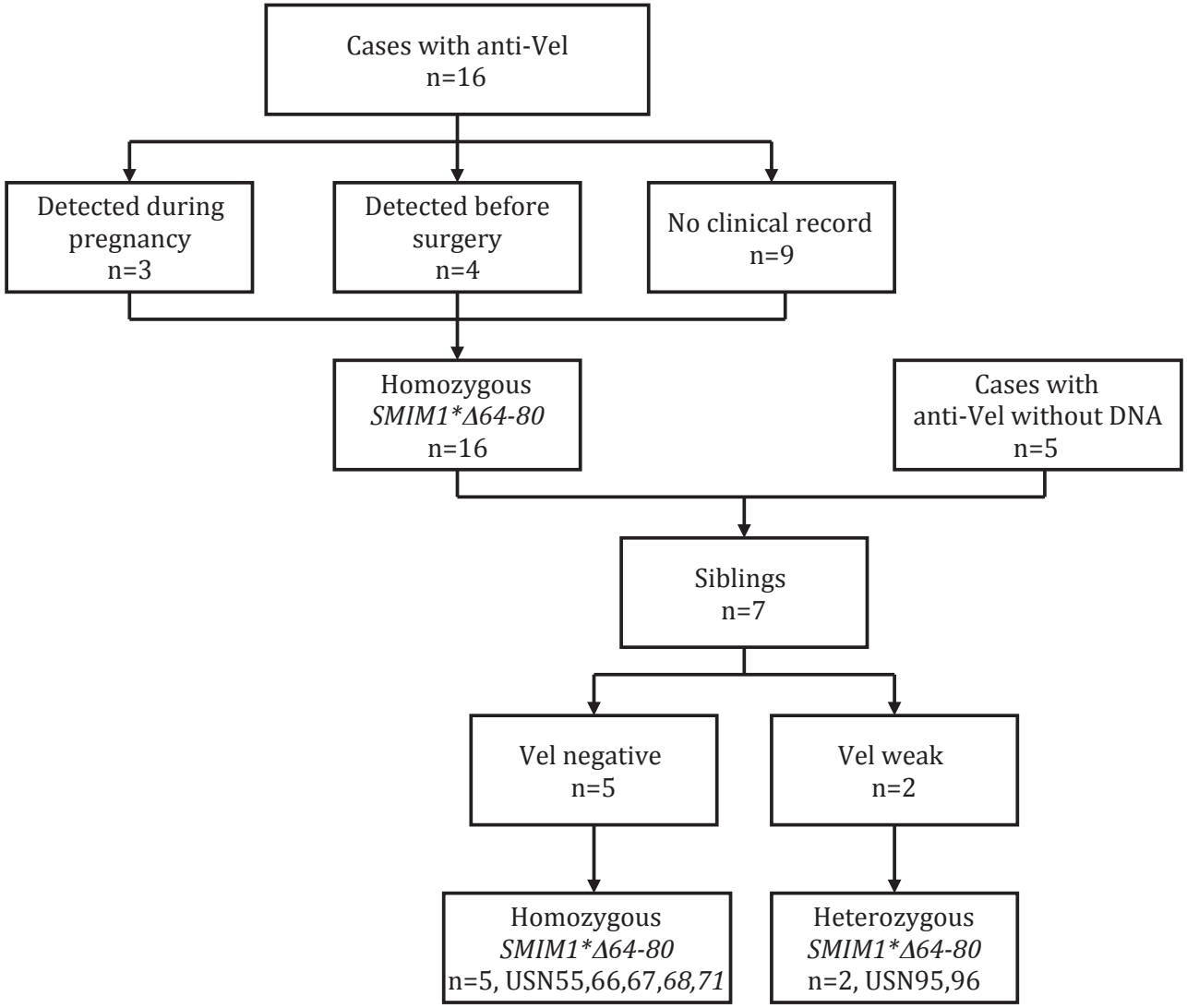


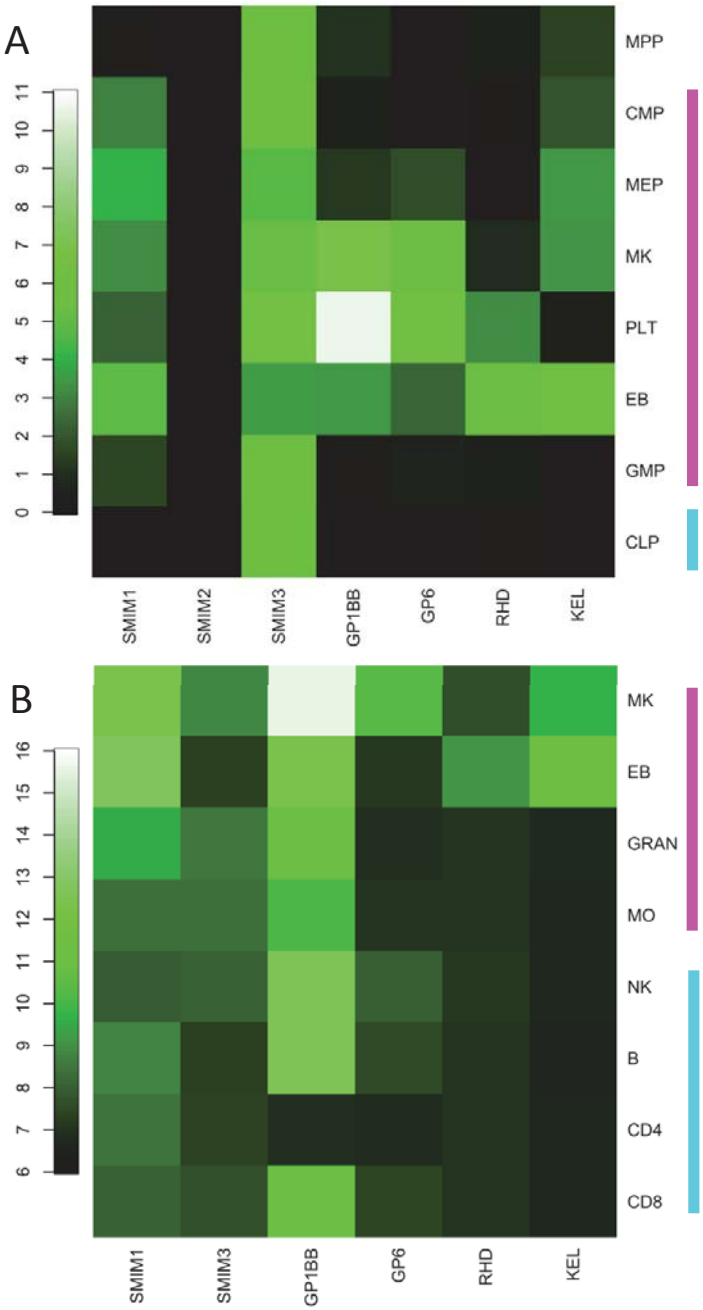

**Supplementary Fig. 1** *Vel-negative and Vel-weak blood donors*. Vel-typed donors were identified by phenotyping their red blood cells by a haemagglutination with an anti-Vel (HA test 1) and the Vel status of apparent Vel-negative donors was confirmed by HA Test 2. This identified 76 Vel-negative and 19 Vel-weak individuals; 48 of the Dutch and English Vel-negative donors were available for this study and another 6 Danish Vel-negative blood donors were included (numbers for Danish donors and related genotyping data are between brackets). The observed *SMIM1* genotypes for the 73 samples are presented. USN,Unique Study Number (see **Supplementary Table 1**).




**Supplementary Fig. 2.** *Exome sequencing of 5 Vel-negative individuals.* The top four individuals are homozygous for the 17 nucleotide framehift deletion (hg19, chr1:Δ3691998-3692014:GTCAGCCTAGGGGCTGT/-) in *SMIM1*, the bottom individual is heterozygous. DNA was enriched using the Roche Nimblegen SeqCap EZ Human Exome v3.0 protocol. Data was visualized using IGV.

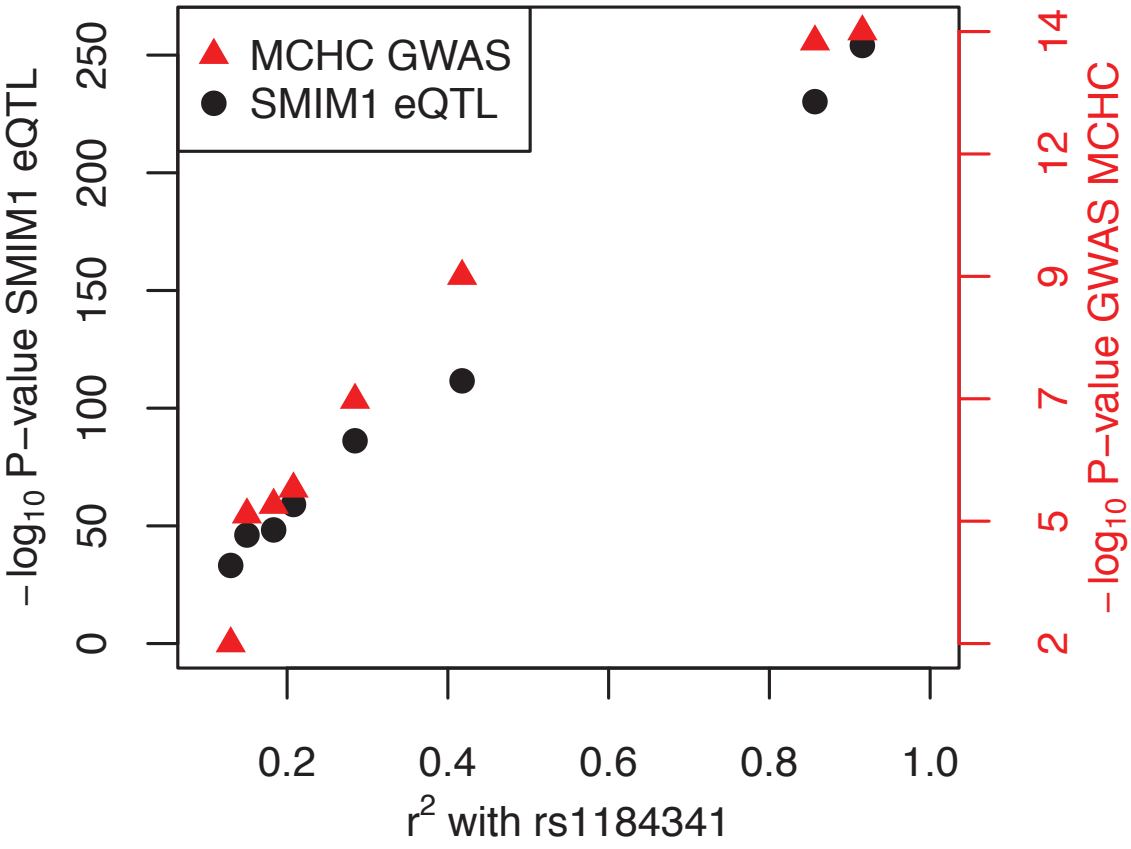



**Supplementary Fig. 3**: *Vel-negative clinical cases with anti-Vel immune antibodies and their siblings*. Twenty-one clinical cases with anti-Vel were identified and 7 siblings from 7 of the 21 cases were also enrolled. DNA samples were obtained from 16 clinical cases with anti-Vel and from all 7 siblings. The observed *SMIM1* genotypes of the 16 cases with anti-Vel and DNA (top of figure) are presented in the main text; the 7 siblings were not included there. USN, Unique Study Number (see **Supplementary Table 1**).

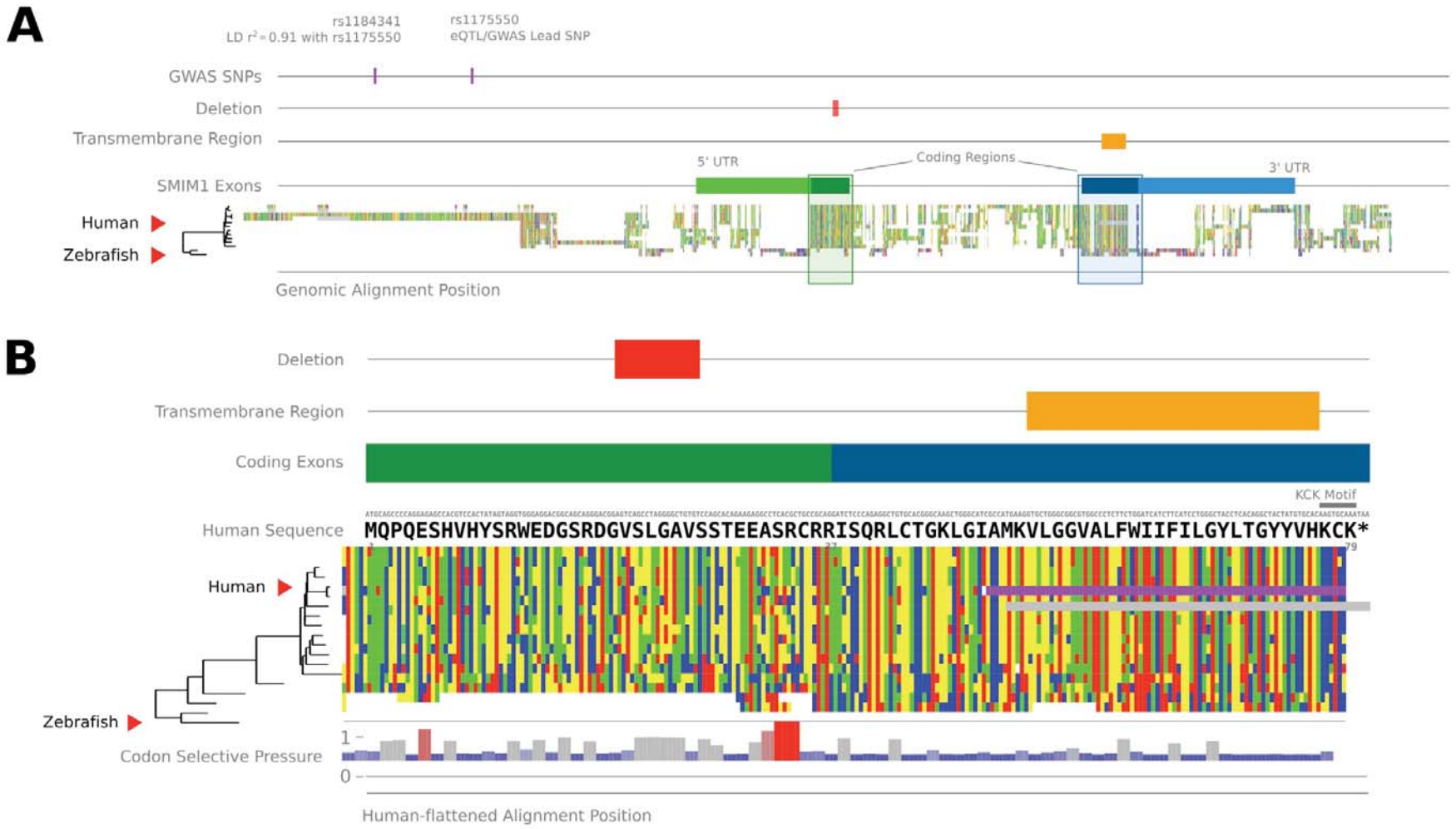



**Supplementary Fig. 4**. *Gene expression of the SMIM family members, two megakaryocyte-specific genes (*GP1BB *and* GP6*) and two lymphoid-specific genes (*RHD *and* KEL*) in haematopoiesis.* A) Heatmap of gene expression of the *SMIM* genes in addition to red cell and platelet lineage markers in haematopoietic precursor cells. Scale bar represents $\log_2$(FPKM+1) as data was quantified from RNA-seq experiments. B) Heatmap for the same genes in mature blood cell types, including the precursors of red cells and platelets using gene expression microarrays (Illumina WGE V2). Note that no probe was available for *SMIM2*. Scale bar represents log2 Signal Intensity (SI). Cell types: MPP, Multipotent Progenitor; CMP, Common Myeloid Progenitor; MEP, Megakaryocyte-Erythroblast Precursor; MK, Megakaryocyte; PLT, Platelets; EB, Erythroblast; GMP, Granulocyte-Monocyte Progenitor, CLP, Common Lymphoid Progenitor; GRAN, Granulocyte; MO, Monocyte; NK, Natural Killer cell; B, CD19+ B cell; CD4, CD4+ T cell; CD8, CD8+ T cell. Magenta: Cell of myeloid lineage; Cyan: Cells of lymphoid lineage.

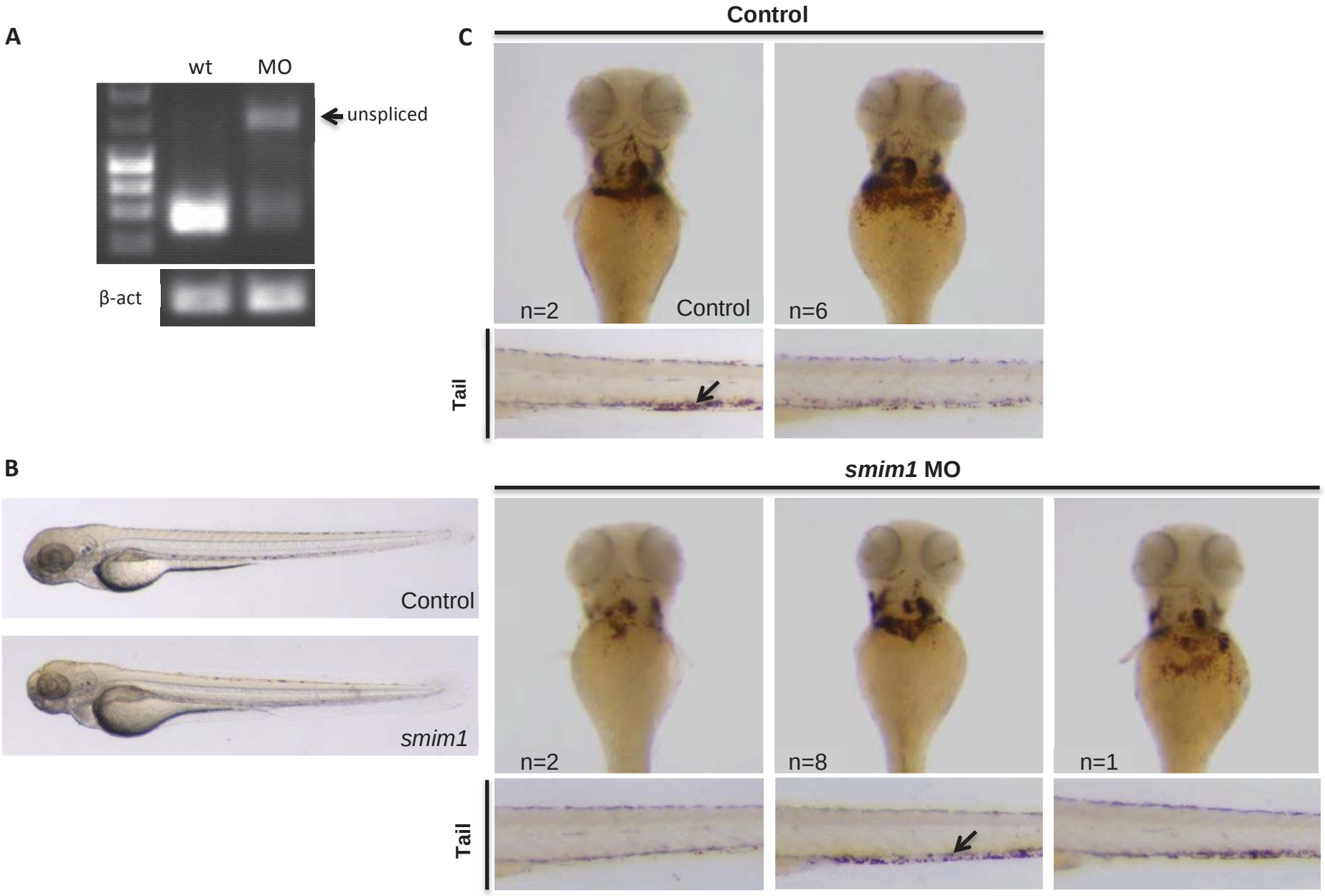

**Supplementary Fig. 5**. *Differential binding of nuclear proteins at the intronic SNP rs1175550.* Electrophoretic mobility shift assays (EMSA) in nuclear protein extracts from the myeloid-erythroid K562 cell line cells showed higher protein affinity of the probe containing the A-allele (major allele of the *SMIM1* intronic SNP rs1175550, lane 1) than the G-allele (minor allele, lane 2).




**Supplementary Fig. 6**. *Co-localization of the* SMIM1 *eQTL association signal and the red blood cell mean corpuscular concentration (MCHC) association signal.* Each circle or triangle represents a SNP. Each SNP has both an SMIM1 eQTL association p-value (black, left axis) and an MCHC association p-value from the meta-analysis of GWAS of red blood cell parameters (red, right axis)(van der Harst et al, Nature, 2012). The position of a SNP on the horizontal axis is determined by the LD $r^2$ with the common SNP rs1184341. rs1184341 is in strong LD ($r^2$=0.92) with rs1175550, the lead eQTL and GWAS SNP. rs1184341 is 288 bp upstream of rs1175550, and is also nucleosome depleted in erythroblasts, the red blood cell precursor (see Figure 1 in main text).




**Supplementary Fig. 7**. *Alignment of human SMIM1 and its vertebrate orthologs.* A) Genomic DNA alignment, including Ensembl orthologs of human *SMIM1* and its manually-identified zebrafish homolog. Only the alignment region surrounding the two exons with protein-coding DNA are shown (two upstream noncoding exons are not shown). Four species (turtle, coelacanth, rat and tenrec) with large intronic insertions were included in the multiple alignment calculation, but are excluded from the plot for clarity. Sequence features were mapped via the human sequence and are shown above the alignment (GWAS SNP rs1175550, Deletion nucleotide 64-80, Transmembrane Region, *SMIM1* Exons). B) Coding sequence DNA alignment, extracted from the genomic DNA alignment by keeping only alignment columns present in the human protein-coding transcript. Ambiguous or missing characters are colored gray. Codon-specific ω estimates from the PAML M3 model analysis are shown below the alignment (Codon Selective Pressure) with a vertical axis ranging from ω=0 to 1; blocks are scaled according to the ω estimate, with sites under purifying selection colored blue, neutral selection colored gray, and positive selection colored red. The human *SMIM1* CDS and protein sequence are shown directly above the alignment, with amino acid residue numbers shown below the sequence and a horizontal line marking the location of the cytoplasmic KCK protein motif. ). The estimated site-wise selective pressures across *SMIM1* (bottom track) showed that most sites (~95%) are under strong purifying selection. The second exon shows especially strong conservation, with most sites showing evidence of strong purifying selection and no sites under apparent positive selection.




**Supplementary Fig. 8**. *Zebrafish studies of SMIM1.* A) Splice modification caused by *smim1* splice MO was assayed by RT-PCR and is seen as a band shift after gel electrophoresis of RT-PCR products (black arrow). RT-PCR and cDNA sequencing results showed presence of unspliced form of *smim1* in MO-injected embryos. RT-PCR analysis of β-actin (β-act) was used as a loading control. B) *Smim1* MO injected embryos showed normal morphology and no off-target defects. C) O-dianisidine staining was used to assess the number of mature red cells in control and *smim1* MO injected embryos at 3 days post fertilisation (pdf). *Smim1* MOs had a mild effect on the formation of red cells. For both control and *smim1* MO injected embryos, MO effect is presented as a set of representative images illustrating different experimental outcomes, with the number of embryos in each group indicated in the lower, left corner. Black arrow indicates the hemoglobin staining in the tail.