LUND UNIVERSITY

Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype.

Storry, Jill; Jöud, Magnus; Christophersen, Mikael Kronborg; Thuresson, Britt; Åkerström, Bo; Sojka, Birgitta Nilsson; Nilsson, Björn; Olsson, Martin L

Published in:
Nature Genetics

DOI:
10.1038/ng.2600

2013

Link to publication

Citation for published version (APA):
Storry, J., Jöud, M., Christophersen, M. K., Thuresson, B., Åkerström, B., Sojka, B. N., Nilsson, B., & Olsson, M. L. (2013). Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype. *Nature Genetics*, *45*(5), 537-U109. https://doi.org/10.1038/ng.2600

Total number of authors:
8

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

LETTER

Homozygosity for a null allele of *SMIM1* defines the Vel– blood group phenotype

Jill R. Storry[1,2,*], Magnus Jöud[1,2,*], Mikael Kronborg Christophersen[1], Britt Thuresson[1,2], Bo Åkerström[3], Birgitta Nilsson Sojka[4], Björn Nilsson[1,2,5,§], Martin L. Olsson[1,2,§]

[1] Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, Sweden. [2] Clinical Immunology and Transfusion Medicine, University and Regional Laboratories, Lund, Sweden. [3] Infection Medicine, Department of Clinical Medicine, Lund University, Lund, Sweden. [4] Clinical Immunology and Transfusion Medicine, Laboratory Medicine, Umeå University Hospital, Umeå, Sweden. [5] Broad Institute, Cambridge, MA, USA.

[*,§] These authors contributed equally to the manuscript (shared first and senior authorships, respectively).

Correspondence to: Jill R. Storry (jill.storry@med.lu.se), Björn Nilsson (bjorn.nilsson@med.lu.se), or Martin L. Olsson (Martin_L.Olsson@med.lu.se).

No. of words in introductory paragraph: 150

No. of words in main manuscript: 1529

No. of display items: 5

Keywords: Vel blood group antigen, erythrocyte, integrative genomics.

The Vel antigen is present on red blood cells (RBCs) from all humans except rare Vel negative (Vel–) individuals who can form anti-Vel in response to transfusion or pregnancy. These antibodies may cause severe hemolytic reactions in blood recipients. We combined SNP profiling and transcriptional network modeling to link the Vel– phenotype to *SMIM1*, located in a 97-kb haplotype block at chromosome 1p36. This gene encodes a previously undiscovered, evolutionarily conserved transmembrane protein expressed on RBCs. Remarkably, 35/35 Vel– individuals were homozygous for a frameshifting 17-bp deletion in exon 3. Functional studies using antibodies raised against SMIM1 peptides confirmed a null phenotype in RBC membranes, and *SMIM1* overexpression induced Vel expression. Genotype screening estimated that ~1/17 Swedish blood donors are heterozygous deletion carriers, ~1/1,200 homozygous deletion knock-outs, and identified new Vel– donors. Our results establish *SMIM1* as a novel erythroid gene, and Vel as a new blood group system.

Human RBCs carry functionally important proteins/glycoproteins on their surfaces. Allelic variation in the underlying genes determines the presence/absence of antigenic structures on these molecules, and, in turn, transfusion and transplant compatibility between individuals.

The Vel– blood group phenotype was described in 1952 when Sussman *et al.* reported a patient who suffered an acute intravascular hemolytic episode following a blood transfusion[1]. The patient made antibodies that reacted with all RBCs tested except her own. The newly defined antigen was named Vel. Subsequent reports of hemolytic reactions upon transfusion of Vel+ RBCs to Vel– patients with anti-Vel, and of hemolytic disease in newborns of Vel– mothers, have established Vel as a clinically important blood group antigen[2,3].

Since its description, Vel has eluded molecular identification. Population studies indicate a prevalence of Vel– individuals at ~1/4,000 in Europe[2], but somewhat higher (~1/1,700) in northern Scandinavia[2,4]. Family studies suggest that the Vel– phenotype is inherited as an autosomal recessive trait, but no locus has been found[4,5]. Similarly, efforts to identify a Vel carrier molecule have been inconclusive[6]. Hence, identification of Vel– blood donors who can provide compatible blood to Vel– patients with anti-Vel relies on typing with scarce polyclonal sera from immunized individuals – monoclonal reagents do not exist.

Because of the presumed inheritance pattern and higher prevalence in northern Scandinavia (a sparsely populated and historically isolated region), we hypothesized that the Vel– phenotype could be caused by a homozygous founder mutation located in a chromosomal region with an identifiable SNP signature. We therefore SNP-profiled 20 Vel– individuals (predominantly Swedish; **Supplementary Table 1**) using microarrays (2.44×10^6 SNPs).

We identified SNPs that segregated with the Vel– phenotype by requiring a homozygous and identical SNP genotype in five Vel– siblings from two families, and a different genotype in their Vel+ siblings (**Fig. 1a**). This filter identified 8,780 SNPs distributed across the genome. To test which of the segregating SNPs were enriched among Vel– individuals, we compared minor allele frequencies among the 15 remaining/non-related Vel– individuals to background frequencies calculated across 379 whole-genome-sequenced individuals of European ethnicity from the 1000 Genomes Project[7]. Strikingly, this revealed 25 highly enriched SNPs in a 97-kilobase haplotype block at chromosome 1p36, where all 20 Vel– individuals showed the same SNP signature (**Fig. 1b** and **Supplementary Table 2**).

The region at 1p36 contains 5 annotated genes (**Fig. 1c** and **Supplementary Table 3**). To prioritize, we inferred a genome-wide gene-regulatory network from 2,096 pre-existing gene expression profiles[8] of human bone marrow samples. Here, *SMIM1* (Small integral

membrane protein 1) stood out as its network neighborhood was dominated by known blood group genes and erythroid genes (**Fig. 2**). Furthermore, SMIM1 has been predicted as a transmembrane protein (www.uniprot.org), contains binding sites for erythroid transcription factors like GATA-1[9] (**Supplementary Table 4**), is highly expressed in erythroleukemia cell lines (**Supplementary Fig. 1**), and is up-regulated in human CD34+ hematopoietic progenitors cultured towards RBCs (**Supplementary Fig. 2**). These results implicated SMIM1 as an erythroid transmembrane protein, an ideal candidate gene.

According to the current human genome annotation, *SMIM1* has four exons and encodes a 78-amino-acid protein (calculated weight ~8.7 kDa). Computational sequence analysis predicted a type 1 transmembrane protein with an extracellular portion of ~50-amino-acid portion containing potential O-glycosylation sites, a transmembrane domain containing a GxxxG motif, a short intracellular tail, but no other functional motifs (**Fig. 3a, b** and **Supplementary Table 5**). Interestingly, protein sequence alignment found no homologs in humans, but 56 orthologs in 45 other species, from primates to sea squirt. Cross-species genome sequence alignment uncovered a conserved GATA motif in primates and rodents, suggesting that *SMIM1* is expressed in the erythroid lineage in higher mammals (**Fig. 3c** and **Supplementary Fig. 3**). Consistent with the informatics, we detected *SMIM1* transcripts in total blood RNA by qPCR, both from Vel+ and Vel– individuals (not shown).

Sequencing of *SMIM1* revealed homozygosity for a 17-bp deletion in the coding region of exon 3 in all 20 Vel– individuals whereas Vel+ controls showed consensus sequence (**Supplementary Fig. 4**). Because the deletion and its surrounding SNP signature is identical in all cases, it is a likely founder mutation. To ascertain that the deletion is truly associated with the Vel– phenotype, we collected DNA from 15 additional Vel– individuals (predominantly European; **Supplementary Table 1**). Remarkably, all cases in this independent validation cohort were homozygous for the same 17-bp deletion. Directed

genotyping (**Fig. 3d**) of 520 Swedish blood donors revealed 30 heterozygous deletion carriers but no homozygotes. These results establish that the 17-bp deletion in *SMIM1* is associated with the Vel– phenotype, and that it is not restricted to the Swedish population.

Assuming Hardy-Weinberg equilibrium, an estimated heterozygote frequency of 30/520 (~1/17) predicts ~1/1,200 homozygotes. Importantly, this figure does not differ significantly from the reported prevalence of ~1/1,700 Vel– individuals[4] (binomial test; P=0.26). Further, while the deletion is not called in 1000 Genomes, the NHLBI Exome Sequencing Project (ESP) lists allele frequencies of 57/5,763 in European Americans and 6/3,198 in African Americans, yielding heterozygote frequencies of ~1/50 and ~1/267. While indel detection by high-throughput sequencing has algorithm-dependent sensitivity, a lower frequency in Europeans than in Swedes agrees with lower prevalences of Vel– individuals[2]. ESP lists the 17-bp deletion as the most common non-consensus coding variant in *SMIM1* (**Supplementary Table 6**), although existence of alternative null alleles cannot be excluded.

The 17-bp deletion introduces a frameshift 5′ of the encoded transmembrane domain. With Vel+ blood, 5′/3′-RACE showed that the wildtype transcript contains the complete coding sequence, and an unannotated non-coding exon 1 (**Supplementary Fig. 5**). With Vel– blood, no capped and polyadenylated mRNA was captured despite repeated attempts. With a known heterozygote, RACE yielded only the wildtype sequence. These data suggest that the mutant transcripts could be subject to degradation by mRNA surveillance mechanisms, as expected from the frameshift[10,11].

To explore a null/knock-out phenotype, we synthesized peptides corresponding to the predicted extracellular domain of SMIM1, and raised rabbit polyclonal antibodies. While these sera did not hemagglutinate RBCs, Western blotting under reducing conditions showed bands at ~9–10 and ~20 kDa with RBC membranes from Vel+ blood, and Glycophorin A (GPA) positive bone marrow cells. Human anti-Vel showed similar reactivity with reduced

Vel+ membranes. Neither band was present in Vel– membranes or Vel+ non-erythroid bone marrow cells (**Fig. 4a–c**). Hence, SMIM1 is present on Vel+, but not Vel–, erythroid cells. Moreover, analysis of microarray profiles of human tissues showed high *SMIM1* expression in bone marrow, and lower levels in non-hematopoietic tissues (**Supplementary Fig. 6**).

The detection of two bands with anti-SMIM1 (at ~1× and ~2× the calculated SMIM1 weight) suggests that the protein forms reduction-resistant homodimers. This is consistent with the presence of the GxxxG motif which mediates transmembrane dimerization via helix-helix association[12,13], as shown for GPA[14]. Additionally, Western blotting under non-reducing conditions showed broad reactivity, suggesting that SMIM1 forms disulfide-dependent complexes with itself or other, as yet unidentified, proteins (**Fig. 4c**). Further biochemical studies confirmed that SMIM1 orients its N-terminal extracellularly, and indicated that O-glycans may not be involved in the Vel epitope (**Fig. 5**).

To verify that Vel depends on *SMIM1*, we over-expressed its cDNA in the erythroleukemic cell line K562. In multiple experiments, we observed increased cell surface expression of Vel in *SMIM1*-transfected, but not mock- or mutant-transfected, cells by flow cytometry with human anti-Vel, accompanied by increased protein expression as shown by anti-SMIM1 (**Fig. 4d,e**). This confirms *SMIM1* as the gene underlying Vel antigen expression.

Finally, we compared Vel expression on RBCs from individuals heterozygous or negative for the deletion. Consistent with the transfections, this revealed a strong correlation between Vel expression and *SMIM1* deletion zygosity, most likely as a gene dose effect (**Fig. 4f,g**).

The Vel antigen has been elusive for 60 years. We show that allelic variation in *SMIM1* defines the Vel blood group system, and link the Vel– phenotype to a founder null allele. It is remarkable that the genetic basis of the Vel– phenotype is identical in 35 individuals, as this contrasts with the heterogenous backgrounds of other recently resolved blood groups[15,16,17].

Importantly, this discovery enables identification of rare Vel– blood donors by genotyping[18]. Screening for the 17-bp deletion was implemented at the Regional Blood Center in Lund, Sweden. To date, two new homozygotes were found, and confirmed serologically as Vel–. This demonstrates how our findings facilitate procurement of compatible blood for Vel-immunized patients, who risk delayed transfusion support and worse outcomes[19].

The identification of *SMIM1* as an erythroid gene – encoding a novel transmembrane protein – raises intriguing questions. Firstly, the protein's functional and molecular context awaits further exploration. Strikingly, there are similarities with glycophorins, especially GPA, a short, type 1, O-glycosylated transmembrane protein that forms reduction-resistant dimers via its GxxxG motif[14]. Secondly, conservation across species indicates an important function, particularly in higher-mammal erythropoiesis. This raises the question whether Vel– individuals, who are in fact human knock-outs, exhibit as yet unrecognized hematologic features. Thirdly, it is conceivable that pathogens utilize SMIM1 as an invasion receptor, similarly to how malarial parasites exploit other blood-group carriers[20,21,22]. In fact, given our data in **Fig. 5** and the similarities with GPA, it is possible that SMIM1 represents the long-sought α-chymotrypsin-sensitive *Plasmodium falciparum* receptor[23]. Future studies will uncover the roles of SMIM1 in hematopoiesis and human disease.

URLs

The 1000 Genomes Project, http://1000genomes.org; gene expression microarray data from ref. [8], http://www.ncbi.nlm.nih.gov/geo (accession no. GSE13159); The NHLBI Exome Sequencing Project (ESP) data, http://evs.gs.washington.edu/EVS; Ultranet, http://www.broadinstitute.org/ultranet; UCSC Genome Browser, http://genome.ucsc.edu/

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes: The genomic *SMIM1* mutant sequence containing the 17-bp deletion, and the new erythroid cDNA wildtype sequence determined by RACE experiments, have been deposited in Genbank under accession numbers KC152643 and KC152644, respectively.

ACKNOWLEDGEMENTS

The study was supported by research grants from the Swedish Foundation for Strategic Research (no. ICA08-0057 to B.N.), the Swedish Research Council (no. 14251 to M.L.O.), governmental ALF grants to the University and Regional Laboratories (Labmedicin Skåne; to J.R.S., M.J., B.N., M.L.O.), the Medical Faculty at Lund University (B.N., M.L.O.), the Skåne County Council's Research and Development foundation, Sweden (M.L.O.), Marianne and Marcus Wallenberg's foundation (no. 2010.0112 to B.N.), Harald Jeansson's Foundation (B.N.), Swedish Society of Medicine (B.N.), the National Blood Foundation (J.R.S.), and Glashofs Legat (no. 302499 to M.K.C.). We thank Joyce Poole, Jan Hamilton, Hein Hustinx, and the SCARF exchange scheme for kindly providing samples for this study. We thank Jörgen Adolfsson, Linnea Järvstråt, and Annika Hult for technical assistance. We are indebted to the individuals and their families who participated in this project.

AUTHOR CONTRIBUTIONS

J.R.S., B.N., M.L.O. conceived, designed, and coordinated the study. M.J. and B.N. performed genomic, bioinformatic, and computational analyses. J.R.S. performed DNA

sequence analysis and protein chemistry. B.T. designed, and M.K.C. performed RACE, qPCR and over-expression experiments. J.R.S. and M.K.C. performed FACS analysis. B.N.S. contributed background knowledge, performed serological screening, and enrolled blood donors. B.Å. contributed to experimental design and the early conception of the project. J.R.S., M.J., B.N. and M.L.O. discovered the 17-bp deletion in *SMIM1* and wrote the paper. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. A patent application covering Vel typing using genetic information from *SMIM1* has been filed (J.R.S., M.J., B.N., and M.L.O.).

FIGURE LEGENDS

Figure 1: Genome-wide SNP profiling maps the locus of the Vel– phenotype to chromosome 1p36. (a) Because of a reported enrichment of Vel– individuals in northern Sweden[4] and anticipated autosomal recessive inheritance, we hypothesized that the Vel– phenotype could be caused by a founder mutation. We analyzed 15 non-familial (predominantly Swedish; **Supplementary Table 1**) and 5 familial Vel– individuals (A1, A2, B6, B7 and B8 in pedigrees; families from Sweden) and their siblings (A3 and B9) on Illumina Human-Omni 2.5M BeadChip microarrays (2.44×10^6 SNPs). **(b)** Based on the familial genotypes, we identified 8,780 SNPs that segregated homozygously with the Vel– phenotype. Each of these SNPs was tested for over-representation among Vel– individuals by comparing their allele frequencies across the 15 non-familial cases with background frequencies calculated from whole-genome-sequence data for 379 individuals of European ethnicity from the 1000 Genomes Project[7]. This identified 25 strongly significant SNPs at 1p36. The *x*-axis indicates SNP locations, the *y*-axis $-\log_{10} P$ values for Fisher's exact test (red bar indicates the $P<10^{-8}$ threshold). **(c)** All 25 significant SNPs were located inside a 97-kb haplotype block that contains five genes (*CCDC27*, *SMIM1*, *LRRC47*, *CEP104*, and *DFFB*; see **Supplementary Table 3**), and is flanked by recombination hotspots in *CCDC27* and *DFFB*. Blue line indicates local recombination rates estimated from the 1000 Genomes data. Red color reflects association with the most significant SNP rs12729572. All Vel– individuals showed the same SNP signature in the identified region (**Supplementary Table 2**).

Figure 2: Identification of SMIM1 as a previously uncharacterized transmembrane protein expressed in erythroid cells. Because most blood group genes are expressed in erythroid cells, we prioritized the genes in the region at 1p36 by computing a genome-wide gene-regulatory network for human hematopoiesis from 2,096 gene expression profiles of bone marrow samples[8]. Using a new computational tool, Ultranet[24], we created a Gaussian Graphical Model which is a sparse graph where nodes represent genes (n=20,958), and vertices correlations in expression between genes that cannot be explained by confounding co-correlation with other observed genes. The network neighborhood of each gene was defined as the set of first- and second-degree neighbor genes in the graph. Strikingly, the neighborhood of *SMIM1* contained 17 of 35 known blood group genes (red; χ^2 test for enrichment, $P<1.0\times10^{-10}$), and was dominated by erythroid genes, including the erythropoietin receptor (*EPOR*), heme synthesis genes (*UROD*, *HMBS*, *BLVRB*, *ALAD*, *ALAS1*, *CPOX*, *FECH*), hemoglobin genes (*HBB*, *HBM*, *HBQ1*), erythroid transcription factors (*GATA1*, *KLF1*, *GFI1B*, *TAL1*, *SOX6*), and erythroid membrane proteins (numerous). By contrast, the neighborhoods of the other four candidate genes contained fewer genes and no erythroid genes or blood group genes (not shown). This finding, along with protein sequence analysis (UniProt accession no. B2RUZ4 and **Fig. 3b**) and additional gene expression microarray data (**Supplementary Figs. 1 and 2**), predicted *SMIM1* as a previously undiscovered transmembrane protein expressed in erythroid cells.

Figure 3: Founder mutation in SMIM1 explains a null/knock-out phenotype in Vel+ RBC membranes. (a) In the current (GRCh37) genome annotation, *SMIM1* has four exons (reference sequence NM_001163724.1). All 20 Vel– individuals in the discovery cohort were homozygous for a 17-bp deletion (red) in exon 3, as were an additional 15 Vel– individuals. **(b)** *SMIM1* encodes a protein of 78 amino acids, ~50 of which constitute the extracellular domain. Potential O-glycosylation (S/T residues) and oligomerization (C residues and GxxxG transmembrane motif) sites are indicated The 17-bp deletion introduces a frameshift from serine 22. **(c)** Cross-species sequence alignment found no protein homologs in humans, but 56 orthologs in 45 other species (11 shown; see also **Supplementary Fig. 4**). The GATA motif in intron 2 (**Supplementary Table 4**) is conserved in primates and rodents, implicating *SMIM1* in higher-mammal erythropoiesis. Conserved amino acids are color coded according to their properties (green: hydrophobic nonpolar; yellow: polar uncharged; orange: acidic; blue: basic). Bars indicate degree of conservation. **(d)** To exclude a common variant, we designed a PCR assay enabling discrimination of consensus/wildtype (w) and deletion/mutant (m) alleles. Genotyping of 520 random Swedish blood donors revealed 30 heterozygous deletion carriers and no homozygotes, showing that the homozygous 17-bp deletion is linked to the Vel– phenotype. Analysis of whole-exome sequence data from several thousand individuals revealed that the 17-bp deletion is the most common coding non-consensus variant in *SMIM1* (**Supplementary Table 6**).

Figure 4: Confirmation of *SMIM1* as the gene underlying Vel expression. (a) Western blotting of reduced solubilized RBC membranes with rabbit polyclonal anti-SMIM1. Two bands at ~9-10 and ~20 kDa seen in Vel+ samples. Weaker bands were observed in the deletion-heterozygous sample (Vel+^{w}). Vel– membranes were non-reactive. **(b)** Western blotting of sorted bone marrow cells with rabbit polyclonal anti-SMIM1, compared with Vel+ and Vel– RBCs. Only CD235a (GPA) positive cells were reactive, indicating that SMIM1 is restricted to the erythroid lineage. Anti-GAPDH loading controls shown in the bottom panels. **(c)** Western blotting with three human anti-Vel sera (designated 1–3) with nonreduced Vel+ and Vel– membranes as indicated, and reduced Vel+ membranes (right-most lane). Thus, anti-Vel and anti-SMIM1 identify similar-sized bands in reduced membranes. Anti-SMIM1 reacts broadly with non-reduced membranes (not shown). **(d)** *SMIM1* over-expression in K562 cells. Cell surface expression of Vel, measured by flow-cytometry with human anti-Vel (ratio of PE positivity/GFP expression; normalized against mock-transfected cells). Mock=pEF1αIRES-ZsGreen1 empty/control vector; SMIM1-mut=pEF1α vector with *SMIM1* containing 17-bp deletion; SMIM1-wt=pEF1α vector with wildtype *SMIM1*). **(e)** Western blotting with anti-SMIM1 shows parallel increase in SMIM1-protein expression. **(f)** Vel expression correlates strongly with *SMIM1* deletion zygosity. To semi-quantify Vel, we used mean fluorescence intensity given as a ratio of human anti-Vel reactivity vs. AB serum in m/m (n=1), w/m (n=16) and w/w (n=21) individuals. Asterisks indicate $P<0.001$ (unpaired Student's *t*-test); the error bars reflect standard error of the mean (SEM). **(g)** The representative histogram shows the differences in antigen expression.

Figure 5: Enzymatic characterization of the extracellular domain of SMIM1. To show that SMIM1 was located on the extracellular surface, anti-SMIM1 reactivity following protease- or glycosidase-treatment of intact RBCs was investigated. Reduced staining of protease-treated RBCs at **(a)** 10 seconds exposure, **(b)** 1 minute exposure, is consistent with overall peptide degradation of the extracellular proteins. This was dose-dependent with both papain (P) and trypsin (T) since 1/100 dilutions of both enzymes (P 1:100 and C 1:100) did not affect reactivity; however, SMIM1 demonstrated an increased sensitivity to α-chymotrypsin (C and C 1:100). **(c)** Reprobing with anti-GPA/B (clone E3, Sigma; 1:10,000) shows sensitivity to papain and α-chymotrypsin, (although a relative resistance to trypsin) and also, the expected shift in molecular weight following treatment with O-glycosidase whereas no effect on SMIM1 reactivity is observed with O-glycosidase treatment. Additionally, we performed hemagglutination experiments with human anti-Vel and RBCs treated with neuraminidase, O-glycosidase or combined neuraminidase + O-glycosidase to assess if Vel reactivity was altered by removing O-glycans. In line with *in silico* predictions (**Supplementary Fig. 7**), and consistent with the Western blot with anti-SMIM1, there was no change in hemagglutination titer/score (not shown). While this does not exclude the presence of minor O-glycan components on SMIM1, it suggests that O-glycans do not play a major role in Vel antigenicity. **(d)** Commassie staining of treated membranes. **(e)** Available cleavage sites in SMIM1 for trypsin (red) and α-chymotrypsin (blue) showing the relative abundance of α-chymotrypsin sites that might explain the unexpected sensitivity. Predicted transmembrane sequence is underlined.

ONLINE METHODS

Samples

Subject to ethics approval and informed consent, anonymized bone marrow samples from healthy volunteers, buffy coat waste material from routine blood donations, and blood samples were obtained from various sources (**Supplementary Table 1**). DNA panels for screening were anonymized collections of 520 unselected Swedish blood donors. No donors were approached solely for the purpose of this study.

Nucleic acid preparation

Genomic DNA was prepared using a modified salting-out procedure[25] and diluted to 100 ng/μL; or the QIAamp DNA Blood Mini Kit (Qiagen) and used undiluted. Total RNA was extracted from whole blood or cell lines using Trizol® LS (Life Technologies Europe).

SNP genotyping and SNP data analysis

A total of 20 Vel– and 8 Vel+ individuals were genotyped on HumanOmni 2.5M BeadChip microarrays probing 2,443,179 SNPs (Illumina; analysis performed at the SCIBLU genomics facility at Lund University). Genotypes were called in Illumina GenomeStudio (call rate >99.5% for all samples; four samples genotyped twice with >99.3% concordance). To estimate background allele frequencies, we used whole-genome sequence data from 379 individuals of European ethnicity from the 1000 Genomes Project[7] (because a high-resolution Swedish SNP data set was unavailable). Allele frequencies were compared by Fisher's exact test with $P<10^{-8}$ as significance threshold. The linkage disequilibrium boundaries of the detected haplotype block were estimated from the unfiltered genotype calls (**Supplementary Table 2**). These boundaries coincided with regions of high recombination rate in the 1000

Genomes Project CEU population. While SNP data for the identified region at 1p36 is given in **Supplementary Table 2**, the complete SNP data are not publicly available to protect the identity of the study participants, as individuals can be identified from SNP profiles.

Network modeling

To infer gene-regulatory networks, we used a new high-performance tool (Ultranet[24]), which enables computation of Graphical Gaussian Models at a genome scale. As input data, we used 2,096 Affymetrix U133A Plus 2 gene expression profiles of human bone marrow from different conditions[8] (NCBI Gene Expression Omnibus; accession no. GSE13159). Ultranet's parameter λ controls the sparsity of the network. In **Fig. 2**, λ=0.75. Other reasonable values yielded agreeing results.

Bioinformatic characterization of *SMIM1*

To predict the location of the transmembrane domain, we used nine different computational tools (**Supplementary Table 5**). To find homologs, we searched the Refseq non-redundant, Ensembl (release 68) and UniProtKB data bases, and retrieved alignment data from UCSC Genome Browser (table knownGene.exonAA). BLASTP was performed using an *E*-value cut-off of 1. To search for conserved protein domains, we used NCBI Conserved Domains. To perform cross-species genome sequence alignment, we used UCSC Genome Browser.

ChIP-seq data from ENCODE[9] obtained via UCSC Genome Browser were examined for transcription factor binding sites in *SMIM1* (**Supplementary Table 4**). Whole-exome sequence data from the Exome Sequencing Project were obtained from NHLBI Exome Variant Server (**Supplementary Table 6**).

Genetic analysis of *SMIM1*

Genomic DNA from 35 Vel– and 10 Vel+ samples were amplified with primers 388588int2f and 388588ex4r which flank the *SMIM1* open reading frame (**Supplementary Table 7**). Amplified products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Life Technologies) on an ABI 3500 Dx Genetic Analyzer. Analyses were performed using CodonCode Aligner.

We developed genotyping methods to allow discrimination of consensus/wildtype and deletion/mutant *SMIM1* alleles (**Supplementary Fig. 8**). To screen blood donors, we used exon 3-specific PCR based on primers LOCex3f_screen and LOCex3r_screen (**Supplementary Table 7**). Allele-specific PCR products of 178 bp (wildtype) or 161 bp (mutant) were discriminated by agarose-gel electrophoresis (**Fig. 3c**).

Transcript analysis

Total RNA from whole blood or cell lines were converted to cDNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). mRNA was amplified with the Expand High Fidelity PCR system (Roche Diagnostics) using primers 388588cDNAf and 388588cDNAr, and the products were sequenced.

A TaqMan® Gene Expression Assay (Hs01369635_g1, Applied Biosystems, binding to the exon 3–4 boundary) was used in real-time qPCR for the detection of *SMIM1* transcripts. Transcript quantities were normalized to human β-actin RNA (Invitrogen assay 4333762F). All samples were run in triplicate, and calibrated against the sample with the lowest C_T value. Triplicates with ≥2 C_T values <40 were considered positive.

Messenger RNA was isolated from total blood RNA by an mRNA Isolation Kit (Roche Diagnostics). RACE was performed with the FirstChoice RLM-RACE kit (Ambion). In 5′-RACE experiments, cDNA was synthesized with random or oligo-dT primers. Gene-specific

primers Vel 332R, Vel 355R and Vel 376R were used for nested PCR amplification together with the 5′-RACE forward primers provided in the kit. For the 3′-RACE, cDNA was obtained with the provided adapter, and nested PCR was performed with primers Vel 59F, Vel 176F and Vel 280F, together with the 3′-RACE reverse primers. PCR products were analyzed by gel electrophoresis, eluted with QIAquick Gel Extraction Kit (Qiagen) and sequenced as described above. Primer sequences are shown in **Supplementary Table 7**.

Generation of rabbit antibodies

Four overlapping peptides were designed from the predicted extracellular domain of SMIM1 (**Supplementary Table 8**). Peptides 1 and 3 were selected to produce rabbit polyclonal antibodies to the predicted SMIM1 protein (Innovagen AB). Partial inhibition of human anti-Vel reactivity with a pool of the peptides was demonstrated (**Supplementary Fig. 9**). While the antibodies were highly specific by Western blotting (**Fig. 4**), they did not hemagglutinate RBCs, nor were they successful in immunoprecipitation protocols. This suggests that they recognized a linear epitope but not the conformational Vel antigen on intact RBCs.

SDS-PAGE and Western Blot analysis

RBC membranes were prepared as described previously[26], and solubilized with 1% Nonidet P-40. The solubilizate was mixed with equal volumes of Laemmli sample buffer with or without the addition of 5% mercaptoethanol depending on the conditions required and run on NuPAGE® 12% or 4–12% Bis-Tris gels (Novex, Life Technologies). The gel was either stained with SimplyBlue™ SafeStain (Life Technologies), or the proteins transferred onto polyvinylidene difluoride (PVDF) membranes. Following blocking in 5% non-fat milk/phosphate-buffered saline (PBS), the membranes were incubated with primary antibody (anti-SMIM1, 1:20,000; affinity-purified human anti-Vel) for 2 hours at room temperature (RT). Following washing in PBS/Tween (PBST), incubation followed for 1 hour at RT with

horseradish peroxidase (HRP)-labelled goat anti-IgG (rabbit, Agrisera; human BioRad) diluted 1:10,000 in PBS-T. The membranes were developed with ECL reagent (Life Technologies), exposed to film for an appropriate time as determined by an initial 60 second exposure, and developed. The membranes were reprobed with a murine anti-GAPDH (clone 6C5, Millipore), diluted 1:5,000, washed and incubated with HRP-labelled goat anti-mouse IgG (BioRad) diluted 1:10,000 in PBS-T; and visualized as above (**Fig. 4**).

Over-expression experiments

The full coding sequence of *SMIM1* was amplified from cDNA using primers VelF and VelR2 (**Supplementary Table 7**), and inserted into a green fluorescent protein-positive plasmid pEF1α-IRES2-ZsGreen1 vector (Clontech) following *Eco*RI/*Bam*HI digestion. NEB5-α competent *E. coli* (New England Biolabs) were transformed, and positive colonies were PCR-amplified and sequenced to confirm the insert. pEF1α vectors containing the wildtype sequence (pEF1α-SMIM1wt), mutant sequence (pEF1α-SMIM1mut) or empty vector (pEF1α-mock) were selected and prepared using Plasmid Midi Kit (Qiagen).

K562 cells were cultured in RPMI 1640 medium (Gibco, Life Technologies) supplemented with 10% fetal bovine serum (Gibco) at 37°C and 5% CO_2. Upon transfection, 5×10^6 K562 cells were mixed with 10 μg pEF1α-vectors, electroporated using the Gene Pulser II electroporation system (BioRad) and incubated for 48 hours in 6-well culture plates (Nunc).

Flow cytometry

Flow cytometry was performed on RBCs, cell lines and transfected K562 cells. A suspension of ~0.5×10^6 RBCs or 10^6 K562 cells in PBS containing 1% bovine serum albumin (PBS/BSA) were tested with human group AB anti-Vel (selected after screening of several anti-Vel reagents) or normal AB serum (isotype control). Either FITC-conjugated $F(ab')_2$

Fragment Rabbit Anti-Human IgG (Dako, Electra-Box Diagnostica AB) (RBCs) or R-Phycoerythrin(PE)-conjugated AffiniPure F(ab′)$_2$ Fragment Goat Anti-Human IgG (Jackson Immunoresearch Europe Ltd) (cell lines and transfected cells) were used as secondary antibodies. Analysis (10,000 events/RBCs; 50,000 events/transfected cells) was performed with the FACSCalibur cell cytometer (Becton Dickinson) using CellQuest™ 3.3 software (Becton Dickinson). PE-conjugated anti-CD33 (Dako, Life Technologies) and 7-AAD (BD Pharmingen) were used as positive compensation controls in tricolor FACS analysis (transfected cells). Viable cells were gated for double-positive signal in the GFP- and PE-channels and assumed to be Vel+ (transfected K562 cells).

Cell populations were sorted from human bone marrow with the following antibodies: CD3-V500, clone UCHT1; CD16-APC-Cy7, clone V NK89; CD19-AlexaFluor 488, clone HIB19; CD34-Fitc, clone 581; CD56-V450, clone B159 (all from Becton Dickinson); CD235ab-APC, Clone HIR2 (Biolegend).

Serologic analysis

Standard agglutination techniques were used for identification of anti-Vel, RBC phenotyping and hemagglutination/inhibition tests[27]. Anti-Vel was affinity-purified by adsorption onto and elution from Vel+ RBCs. Eluates were prepared using the Elu-Kit II kit (Immucor).

Enzyme treatment

RBCs were treated with enzymes (Sigma-Aldrich), including papain (0.1%/0.001%), trypsin (1%/0.01%), α-chymotrypsin (0.5%/0.005%), neuraminidase + O-glycosidase, and O-glycosidase only. For protease treatment, 100 µL washed packed RBCs were incubated with 100 µL enzyme for one hour at 37°C. Glycosidase treatment was performed by (a) incubating 100 µL washed packed RBCs with 0.05 U neuraminidase diluted in 100 µL PBS for one hour at 37°C, then adding 0.02 U O-glycosidase diluted in 100 µL PBS for a further hour's

incubation; and (b) washed packed RBCs were incubated only with O-glycosidase. RBC membranes were prepared directly following enzyme treatment.

Statistical analysis

Statistical tests were performed in R (http://www.r-project.org/) or Excel (Microsoft). For single-hypothesis tests, *P*-values <0.05 were considered significant.

Establishing Vel as a blood group system

The information presented in this paper fulfills the formal requirements to establish Vel as a novel human blood group system. Based on the available data, the current chair (J.R.S.) and two members (current secretary Christine Lomas-Francis, M.Sc, FIBMS, and former chair Dr. Geoff Daniels) of the International Society of Blood Transfusion (ISBT) Working Party for Red Cell Immunogenetics and Blood Group Terminology have assigned a provisional blood group system name and number, "Vel, system no. 34", and promoted Vel from a genetically unresolved blood group antigen (ISBT no. 212001) to the new system defined by the *SMIM1* locus (as antigen no. 034001).

References

1. Sussman, L.N. & Miller, E.B. New blood factor: Vel. *Rev. Hematol.* **7,** 368–371 (1952).
2. Daniels, G. *Human Blood Groups*. (Blackwell Science, Oxford,UK, 2002).
3. Reid, M.E., Lomas-Francis, C. & Olsson, M.L. *The Blood Group Antigen FactsBook* (Academic Press, London, UK, 2012).
4. Cedergren, B., Giles, C.M. & Ikin, E.W. The Vel blood group in northern Sweden. *Vox. Sang* **31,** 344–355 (1976).
5. Levine, P., White, J.A. & Stroup, M. Seven Ve-a (Vel) negative members in three generations of a family. *Transfusion* **1,** 111–115 (1961).
6. Storry, J.R., Åkerström, B. & Olsson, M.L. Investigation into the carrier molecule of the Vel blood group antigen. *Transfusion* **50,** 28 (2010).
7. The 1000 Genomes Project Consortium.An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
8. Haferlach, T. *et al.* Clinical utility of microarray-based gene expression profiling in the diagnosis and classification of leukemia. *J. Clin. Oncol.* **28,** 2529–2537 (2010).
9. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
10. van Hoof, A. & Wagner, E.J. A brief survey of mRNA surveillance. *Trends Biochem. Sci.* **36,** 585–592 (2011).
11. Shoemaker, C.J. & Green, R. Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* **19,** 594–601 (2012).
12. Russ, W.P. & Engelman, D.M. The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296,** 911–919 (2000).

13. Curran, A.R. & Engelman, D.M. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* **13,** 412–417 (2003).

14. MacKenzie, K.R. & Engelman, D. M. Structure-based prediction of the stability of transmembrane helix-helix interactions: the sequence dependence of glycophorin A dimerization. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 3583–3590 (1998).

15. Zelinski, T., Coghlan, G., Liu, X.Q. & Reid, M.E. ABCG2 null alleles define the Jr(a–) blood group phenotype. *Nat. Genet.* **44,** 131–132 (2012).

16. Helias, V. *et al.* ABCB6 is dispensable for erythropoiesis and specifies the new blood group system Langereis. *Nat. Genet.* **44,** 170–173 (2012).

17. Saison, C. *et al.* Null alleles of ABCG2 encoding the breast cancer resistance protein define the new blood group system Junior. *Nat. Genet.* **44,** 174–177 (2012).

18. Quill, E. Blood-matching goes genetic. *Science* **319,** 1478–1479 (2008).

19. Seltsam, A., Wagner, F.F., Salama, A. & Flegel, W.A. Antibodies to high-frequency antigens may decrease the quality of transfusion support: an observational study. *Transfusion* **43,** 1563–1566 (2003).

20. Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature* **480,** 534–537 (2011).

21. Rowe, J.A., Opi, D.H. & Williams, T.N. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. *Curr. Opin. Hematol.* **16,** 480–487 (2009).

22. Moulds, J.M. & Moulds, J.J. Blood group associations with parasites, bacteria, and viruses. *Transfus. Med. Rev.* **14,** 302–311 (2000).

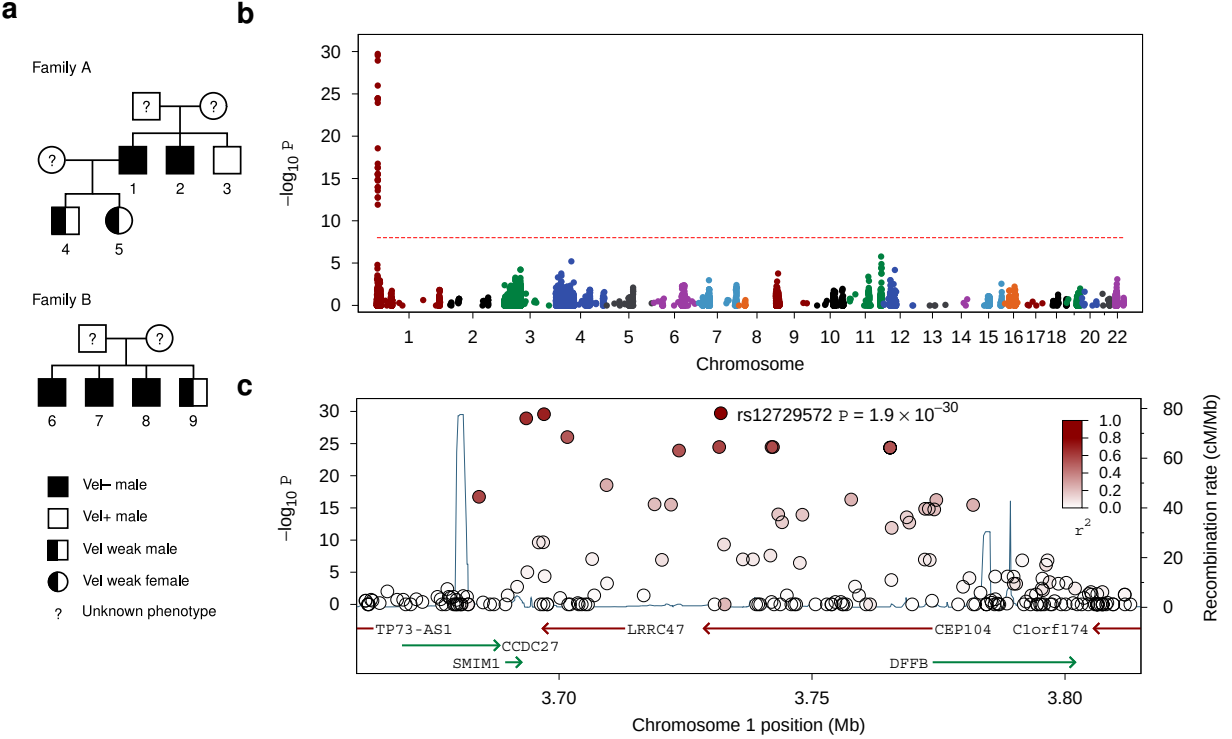
23. Gilberger, T.W. *et al.* A novel erythrocyte binding antigen-175 paralogue from Plasmodium falciparum defines a new trypsin-resistant receptor on human erythrocytes. *J. Biol. Chem.* **278**, 14480–14486 (2003).

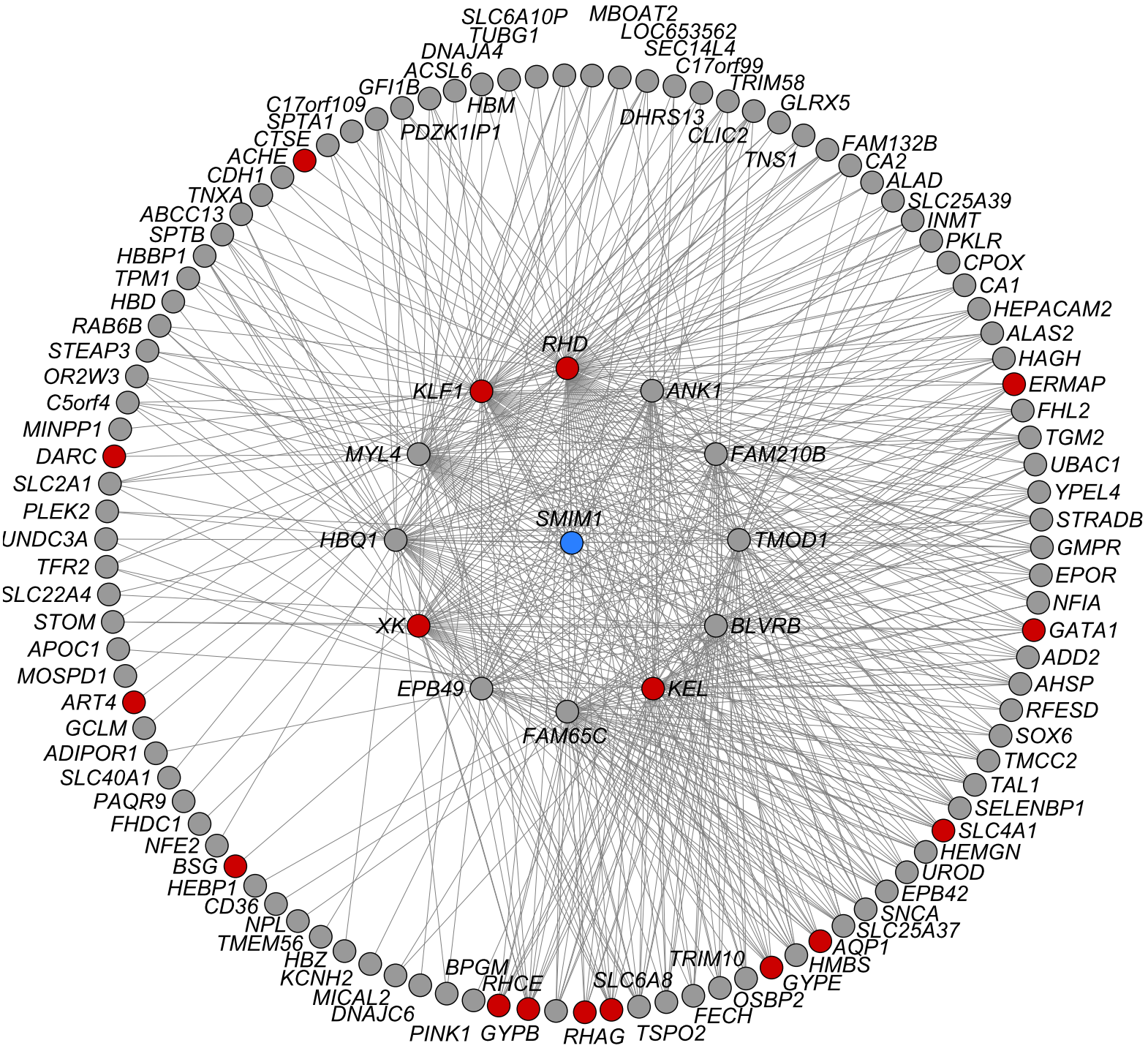
24. Järvstråt, L., Johansson, M., Gullberg, U. & Nilsson, B. Ultranet: high-performance solver for the sparse inverse covariance selection problem in gene network modeling. *Bioinformatics* Epub ahead of print, doi:10.1093/bioinformatics/bts717 (2012).

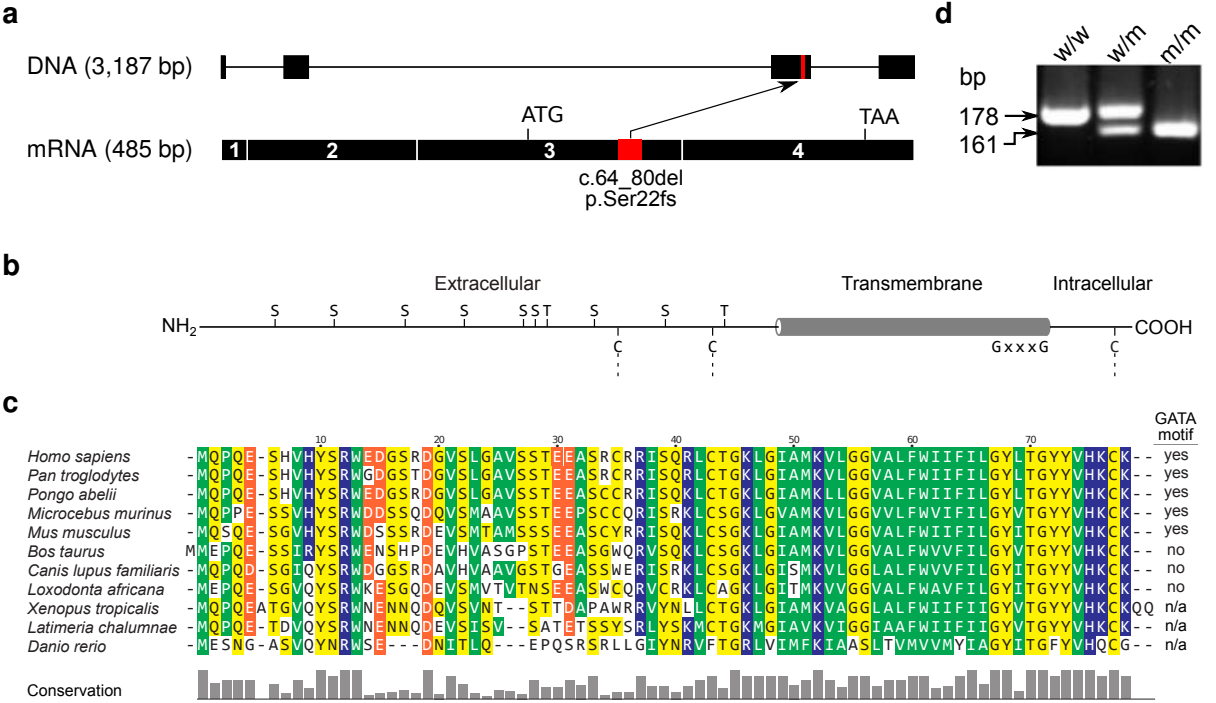
25. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16,** 1215 (1988).

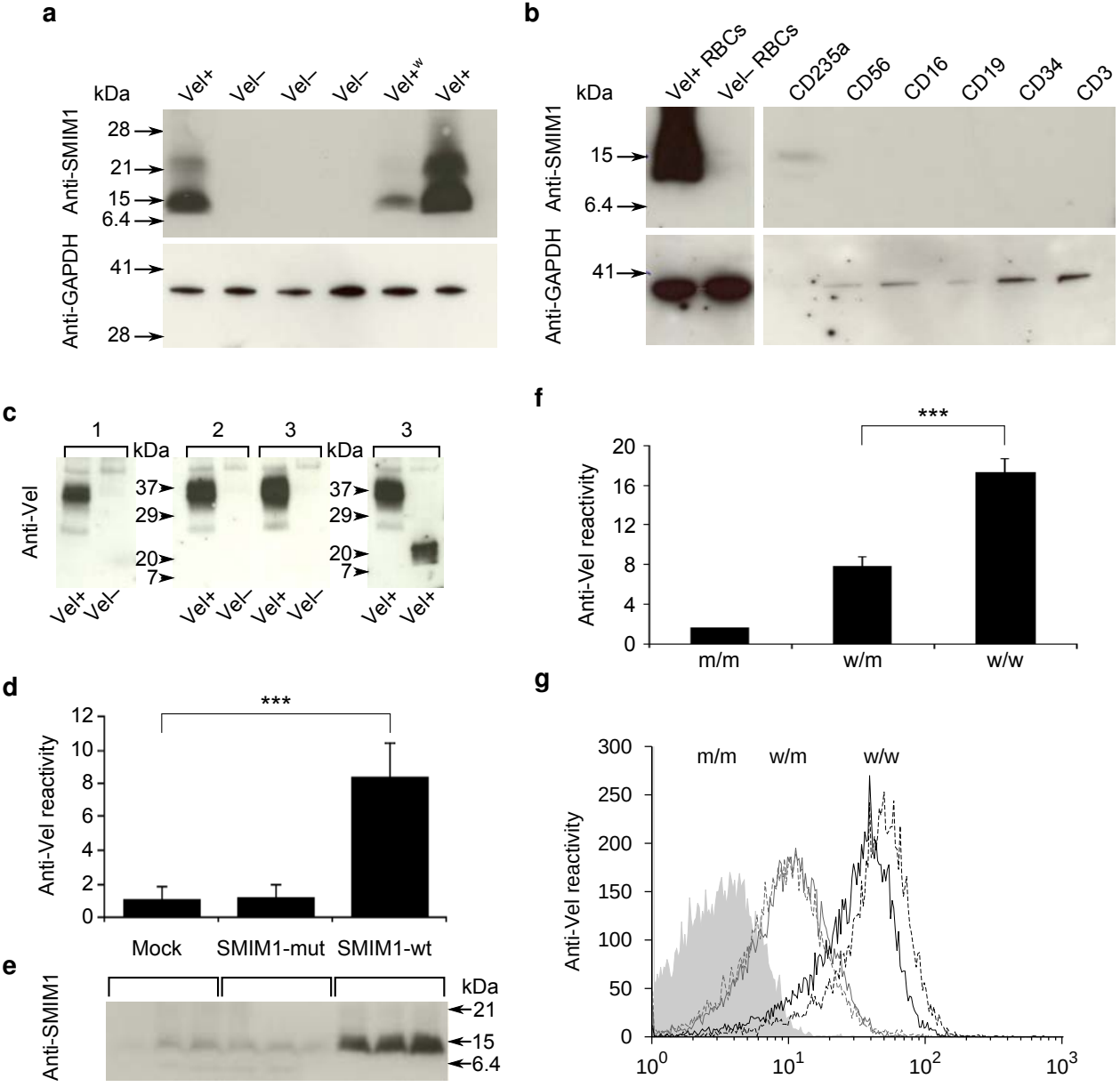
26. Dodge, J.T., Mitchell, C. & Hanahan, D.J. The preparation and chemical characteristics of hemoglobin-free ghosts of human erythrocytes. *Arch. Biochem. Biophys.* **100,** 119–130 (1963).

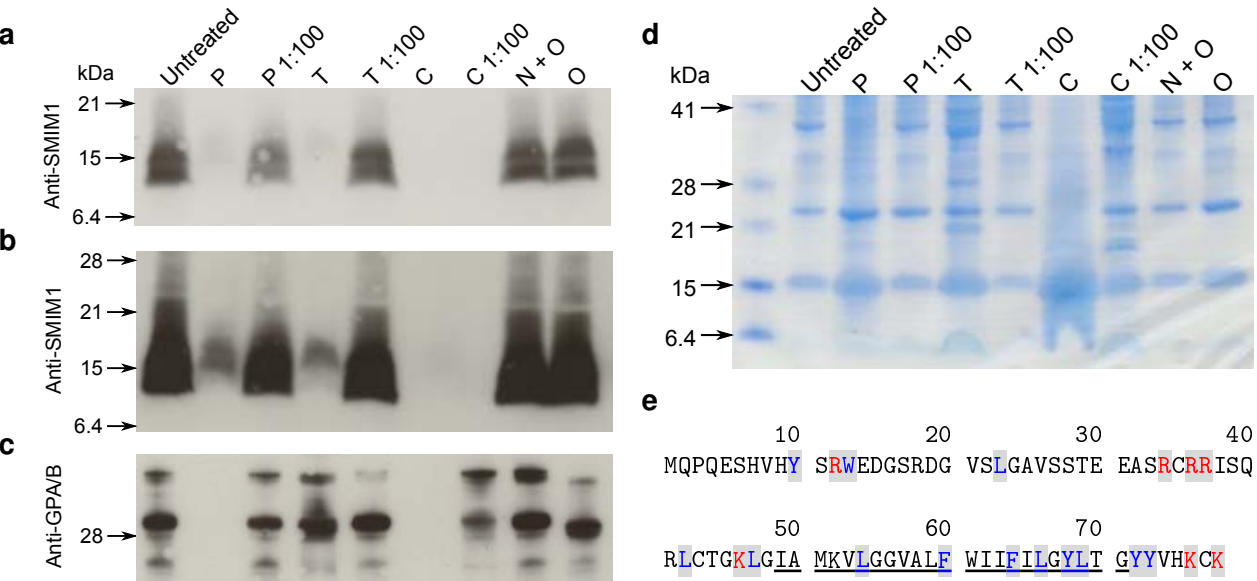
27. Judd, W.J., Johnson, S.T. & Storry, J.R. *Judd's Methods in Immunohematology* (AABB Press, Bethesda, MD, 2008).


a
Family A
?
?
?
1
2
3
4
5
Family B
?
?
6
7
8
9
Vel− male
Vel+ male
Vel weak male
Vel weak female
? Unknown phenotype
b
−log10 P
Chromosome
c
rs12729572 P = 1.9 × 10−30
r2
Recombination rate (cM/Mb)
TP73-AS1
CCDC27
SMIM1
LRRC47
CEP104
C1orf174
DFFB
Chromosome 1 position (Mb)


SLC6A10P
MBOAT2
TUBG1
LOC653562
DNAJA4
SEC14L4
ACSL6
C17orf99
GFI1B
TRIM58
HBM
GLRX5
C17orf109
DHRS13
SPTA1
PDZK1IP1
CLIC2
CTSE
FAM132B
TNS1
ACHE
CA2
CDH1
ALAD
TNXA
SLC25A39
ABCC13
INMT
SPTB
PKLR
HBBP1
CPOX
TPM1
CA1
HBD
HEPACAM2
RAB6B
ALAS2
STEAP3
RHD
HAGH
OR2W3
KLF1
ANK1
ERMAP
C5orf4
FHL2
MINPP1
TGM2
MYL4
FAM210B
DARC
UBAC1
SLC2A1
YPEL4
PLEK2
SMIM1
STRADB
UNDC3A
HBQ1
TMOD1
GMPR
TFR2
EPOR
SLC22A4
NFIA
STOM
XK
BLVRB
GATA1
APOC1
ADD2
MOSPD1
EPB49
KEL
AHSP
ART4
RFESD
FAM65C
GCLM
SOX6
ADIPOR1
TMCC2
SLC40A1
TAL1
PAQR9
SELENBP1
FHDC1
SLC4A1
NFE2
HEMGN
BSG
UROD
HEBP1
EPB42
CD36
SNCA
NPL
SLC25A37
TMEM56
AQP1
HBZ
TRIM10
HMBS
KCNH2
BPGM
SLC6A8
GYPE
MICAL2
RHCE
OSBP2
DNAJC6
FECH
PINK1
GYPB
RHAG
TSPO2


a
DNA (3,187 bp)
ATG
TAA
mRNA (485 bp)
1
2
3
4
c.64_80del
p.Ser22fs
d
w/w
w/m
m/m
bp
178
161
b
Extracellular
Transmembrane
Intracellular
NH2
COOH
GxxxG
c
GATA motif
Homo sapiens yes
Pan troglodytes yes
Pongo abelii yes
Microcebus murinus yes
Mus musculus yes
Bos taurus no
Canis lupus familiaris no
Loxodonta africana no
Xenopus tropicalis n/a
Latimeria chalumnae n/a
Danio rerio n/a
Conservation


a
kDa
Vel+
Vel–
Vel–
Vel–
Vel+w
Vel+
Anti-SMIM1
28
21
15
6.4
Anti-GAPDH
41
28
b
kDa
Vel+ RBCs
Vel– RBCs
CD235a
CD56
CD16
CD19
CD34
CD3
Anti-SMIM1
15
6.4
Anti-GAPDH
41
c
1
2
3
3
kDa
kDa
Anti-Vel
37
29
20
7
Vel+
Vel–
Vel+
Vel–
Vel+
Vel–
Vel+
Vel+
d
***
Anti-Vel reactivity
12
10
8
6
4
2
0
Mock
SMIM1-mut
SMIM1-wt
e
Anti-SMIM1
kDa
21
15
6.4
f
***
Anti-Vel reactivity
20
16
12
8
4
0
m/m
w/m
w/w
g
m/m
w/m
w/w
Anti-Vel reactivity
300
250
200
150
100
50
0
10^0
10^1
10^2
10^3


a
kDa
Untreated
P
P 1:100
T
T 1:100
C
C 1:100
N + O
O
21
15
6.4
Anti-SMIM1
b
28
21
15
6.4
Anti-SMIM1
c
28
Anti-GPA/B
d
kDa
41
28
21
15
6.4
e
10 20 30 40
MQPQESHVHY SRWEDGSRDG VSLGAVSSTE EASRCRRISQ
50 60 70
RLCTGKLGIA MKVLGGVALF WIIFILGYLT GYYVHKCK

Supplementary Information

Homozygosity for a null allele of *SMIM1* defines the Vel– blood group phenotype

Jill R. Storry[1,2,*], Magnus Jöud[1,2,*], Mikael Kronborg Christophersen[1], Britt Thuresson[1,2], Bo Åkerström[3], Birgitta Nilsson Sojka[4], Björn Nilsson[1,2,5,§], Martin L. Olsson[1,2,§]

[1] Hematology and Transfusion Medicine, Department of Laboratory Medicine, Lund University, Sweden. [2] Clinical Immunology and Transfusion Medicine, University and Regional Laboratories, Lund, Sweden. [3] Infection Medicine, Department of Clinical Medicine, Lund University, Lund, Sweden. [4] Clinical Immunology and Transfusion Medicine, Laboratory Medicine, Umeå University Hospital, Umeå, Sweden. [5] Broad Institute, Cambridge, MA, USA.

[*,§] These authors contributed equally to the manuscript (shared first and senior authorships, respectively).

Supplementary Table 1	List of samples
Supplementary Table 2	Unfiltered genotype calls for all subjects in the chromosome 1 candidate region
Supplementary Table 3	Genes in the identified region at 1p36
Supplementary Table 4	ChIP-seq peaks
Supplementary Table 5	Transmembrane region predictions
Supplementary Table 6	*SMIM1* variants in larger populations
Supplementary Table 7	List of primers
Supplementary Table 8	Peptides

Supplementary Figure 1	*SMIM1* transcript expression in cell lines
Supplementary Figure 2	*SMIM1* expression in CD34+ hematopoietic progenitors cultured towards RBCs
Supplementary Figure 3	Cross-species alignment
Supplementary Figure 4	Sequence data
Supplementary Figure 5	RACE on Vel+ blood samples
Supplementary Figure 6	*SMIM1* expression in normal human tissues
Supplementary Figure 7	Predictive analysis of potential O-glycosylation sites in SMIM1
Supplementary Figure 8	Genotyping assays
Supplementary Figure 9	Inhibition of anti-Vel

Supplementary Table 1: List of samples

This table lists all samples in the discovery cohort (above bar) and validation cohort (below bar).

Sample number	Vel phenotype	Comments	Origin
1	Vel–	Blood donor	Southern Sweden
2	Vel–	Brother of #1	Southern Sweden
3	Vel+	Brother of #1	Southern Sweden
4	Vel+	Son of #1	Southern Sweden
5	Vel+	Daughter of #1	Southern Sweden
6	Vel–	Blood donor	Northern Sweden
7	Vel–	Brother of #6	Northern Sweden
8	Vel–	Brother of #6	Northern Sweden
9	Vel+	Brother of #6	Northern Sweden
10	Vel–	Unrelated blood donor	Southern Sweden
11	Vel–	Unrelated blood donor	Southern Sweden
12	Vel–	Unrelated blood donor	Southern Sweden
13	Vel+	Unrelated blood donor	Northern Sweden
14	Vel–	Unrelated blood donor	Northern Sweden
15	Vel–	Unrelated blood donor	Northern Sweden
16	Vel–	Unrelated blood donor	Northern Sweden
17	Vel–	Unrelated blood donor	Northern Sweden
18	Vel–	Unrelated blood donor	Northern Sweden
19	Vel–	Unrelated blood donor	Northern Sweden
20	Vel–	Unrelated blood donor	Northern Sweden
21	Vel–	Unrelated blood donor	Northern Sweden
22	Vel–	Unrelated blood donor	Southern Sweden
23	Vel+	Unrelated blood donor	Northern Sweden
24	Vel–	Unrelated blood donor	Detroit, USA
25	Vel–	Unrelated blood donor	Minneapolis, USA
26	Vel–	Unrelated blood donor	Minneapolis, USA
27	Vel+	Unrelated blood donor	Houston, USA
28	Vel+	Unrelated blood donor	Southern Sweden
29	Vel–	Blood donor	Southern Sweden
30	Vel–	Patient with anti-Vel	England
31	Vel–	SCARF sample	England
32	Vel–	Patient with anti-Vel	England
33	Vel–	Patient with anti-Vel	England
34	Vel–	Patient with anti-Vel	England
35	Vel–	SCARF antiserum	USA
36	Vel–	SCARF antiserum	USA
37	Vel–	SCARF antiserum	USA
38	Vel–	Blood donor	Switzerland
39	Vel–	Blood donor	Switzerland
40	Vel–	Blood donor	Switzerland
41	Vel–	Blood donor	Switzerland
42	Vel–	Blood donor	Switzerland
43	Vel–	SCARF antiserum	Eastern Europe

Abbreviations: SCARF (Serum, cells and rare fluids exchange; `http://scarfex.jove.prohosting.com/`).

Supplementary Table 2: Unfiltered genotype calls for all subjects in the chromosome 1 candidate region

Unfiltered genotype calls for all subjects (n=28) in the chromosome 1 candidate region. Coordinates are GRCh37. Genotypes of Vel– subjects (n=20) are printed in bold font. SNPs where all Vel– subjects have identical genotype are highlighted in yellow. Genes in the region are shown in the right margin.

		Family A					Family B				Vel– samples (unrelated)															Vel+ controls				
Sample id		1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18	19	20	21	22	24	25	26	13	23	27	28	
Vel phenotype		–	–	+	+	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	
Chr	Pos																													Gene
1	3651126	TG	TG	TG	TT	TG	TT	TT	TT	TT	TG	TG	TG	TT	TG	TT	TT	TG	TG	TG	TT	TG	TT	TT	TG	TT	TT	TT	TT	
1	3651318	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	TC	TC	CC	TC	CC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3652626	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3652818	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3653412	CC	CC	CC	CC	CC	CC	CC	CC	CC	AC	AC	AC	CC	AC	CC	CC	CC	CC	CC	CC	CC	AC	CC	CC	CC	CC	CC	CC	
1	3653426	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	AG	AG	GG	AG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	
1	3654595	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3654747	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3655141	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	
1	3655698	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3655719	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3655818	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3656941	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3657470	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3657759	AG	AG	AG	GG	AA	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AA	AG	GG	GG	AA	GG	AG	AG	GG	GG	GG	AG	
1	3659657	AG	AG	AG	GG	AA	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AA	AG	GG	GG	AA	GG	AG	AG	GG	GG	GG	AG	
1	3659742	TC	TC	TC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	TC	CC	CC	TC	CC	CC	TC	CC	CC	CC	CC	
1	3661751	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3661806	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3662208	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3662309	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3662409	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3662845	AG	AG	AG	AA	GG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	GG	AG	AA	AA	GG	AA	AG	AG	AA	AA	AA	AG	
1	3663128	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3664552	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3666079	AG	AG	AG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	AG	GG	GG	AG	GG	GG	AG	GG	GG	GG	GG	
1	3668004	TC	TC	TC	TT	CC	TT	TT	TT	TC	TT	TT	TT	TC	TT	TT	TT	CC	TC	TT	TT	CC	TT	TC	TC	TT	TT	TT	TC	
1	3669201	AG	AG	AG	AA	GG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	GG	AG	AA	AA	GG	AA	AG	AG	AA	AA	AA	AG	*CCDC27*
1	3669703	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3670691	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	
1	3671791	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	
1	3672992	CC	CC	AC	AC	CC	CC	CC	CC	CC	AC	CC	AC	AC	AC	AC	AA	CC	CC	AC	AC	CC	AC	AC	CC	AC	AC	CC	CC	
1	3675959	TT	TT	TT	TT	TG	TT	TT	TT	TG	TT	TT	TT	TG	TT	TT	TT	TG	TT	TT	TT	TG	TT	TG	TT	TT	TT	TT	TG	
1	3676598	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	
1	3676868	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	

1	3677579	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3677933	AA	AA	AA	AA	AG	AA	AA	AA	AG	AA	AG	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AG	AA	AA	AA	AA	AG		
1	3678231	AA	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AG	AA	AA	AA	AA	AG		
1	3678511	AA	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AG	AA	AA	AA	AA	AG		
1	3679146	GG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AG	GG	AG	GG	GG	GG	GG	AG		
1	3679203	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3679477	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3679775	CC	CC	CC	CC	TC	CC	CC	CC	TC	CC	CC	CC	TC	CC	CC	CC	TC	CC	CC	CC	TC	CC	TC	CC	CC	CC	CC	TC		
1	3679783	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
1	3679886	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3680333	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3680643	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC	TC	CC	CC	CC	TC	CC	CC	CC	CC	TC	TC	CC	TC	CC	TC	CC	TC	TC		
1	3680729	AG	AG	AA	GG	AG	GG	GG	GG	AG	GG	AG	GG	GG	AG	GG	GG	AG	AG	GG	GG	AA	GG	AG	AG	AG	GG	AG	GG		
1	3681063	AG	AG	AG	AA	GG	AA	AA	AA	AA	AG	AG	AA	AG	AG	AA	GG	AG	GG	AA	AA	AG	AG	AA	AG	AG	AG	AA	0		
1	3681831	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AG	AA		
1	3681985	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
1	3682013	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3682070	TT	TT	TT	TG	TG	TT	TT	TT	TG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GG	TT	TG	GG		
1	3684185	AG	AG	GG	AG	GG	AA	AA	AA	AG	AA	AA	AG	AA	AG	AA	AA	GG	AG	AA	AA	AG	AG	AA	AG	GG	AG	GG	GG		
1	3684954	AC	AC	AA	CC	AC	CC	CC	CC	CC	CC	CC	AC	CC	AC	CC	CC	AA	AC	CC	CC	AC	AC	CC	AC	CC	CC	CC	CC		
1	3686361	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3687140	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3688030	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
1	3689408	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC		
1	3689852	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC		
1	3691727	TT	TT	TG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TG		
1	3693511	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TC	CC	CC		
1	3693755	GG	GG	TG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	TG	GG	TG		
1	3696000	GG	GG	GG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	GG	AA	AG		
1	3696356	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3696890	CC	CC	CC	AC	AC	CC	CC	CC	AC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	CC	AA	AC		
1	3696925	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3697035	GG	GG	AG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AG	AA	AA		
1	3697137	GG	GG	GG	GG	TG	GG	GG	GG	TG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	TG	GG	GG	TG		
1	3697663	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3697700	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3700216	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC		
1	3701663	GG	GG	AG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AG	AA	AA		
1	3701795	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3702188	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		

CCDC27 (cont.)

SMIM1

LRRC47

1	3703337	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	*LRRC47 (cont.)*
1	3703710	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3703903	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3704479	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3705211	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	
1	3705711	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3706538	CC	CC	CC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	CC	TT	TC	
1	3706969	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AC	CC	CC	
1	3709412	CC	CC	TC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	CC	TT	TT	
1	3709487	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AG	
1	3716737	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	AA	GG	
1	3718864	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TT	CC	CC	
1	3720323	AA	AA	AA	AC	AC	AA	AA	AA	AC	AA	0	0	AA	AA	0	AA	0	AA	AA	AA	AA	AA	AA	AA	AA	CC	AA	CC	AC	
1	3720466	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3722130	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG	
1	3723708	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AG	GG	GG	
1	3730729	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	*CEP104*
1	3731657	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TC	CC	CC	
1	3731959	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TC	CC	CC	
1	3732596	TT	TT	TT	TG	TG	TT	TT	TT	TG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GG	TT	GG	TG	
1	3732729	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	
1	3736319	GG	GG	GG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	GG	AA	AG	
1	3738332	GG	GG	GG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	GG	AA	AG	
1	3738954	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3739382	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3739740	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3741782	TT	TT	TT	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TT	CC	TC	
1	3741985	CC	CC	TC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TC	TT	TT	
1	3742009	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	0	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3742257	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TC	CC	CC	
1	3742498	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3743319	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG	
1	3743703	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	
1	3744084	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG	
1	3745015	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TG	TT	TT	
1	3745924	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3746484	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	
1	3746874	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	
1	3747580	CC	CC	CC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	0	CC	CC	CC	CC	CC	CC	CC	TT	CC	TT	TC	
1	3748085	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG	

1	3750216	CC	CC	CC	AC	AC	CC	CC	CC	AC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	CC	AA	AC	***CEP104 (cont.)***
1	3750686	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3752941	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3753813	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
1	3753892	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3754596	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
1	3755675	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3755965	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3756226	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3756774	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3757749	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG		
1	3758493	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC		
1	3759265	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AG	AA	AA		
1	3760672	AA	AA	AA	AA	AA	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA		
1	3761479	GG	GG	GG	AG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	GG	AA	AG		
1	3765424	TT	TT	TG	TG	TT	TT	TT	TT	TG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	GG	TG	GG	GG		
1	3765678	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC		
1	3765755	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	0	TT	TT	TT	TT	TT	TT	CC	TT	CC	CC		
1	3768719	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TT	CC	CC		
1	3769231	AA	AA	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	AA	GG	GG		
1	3769480	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3769760	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3772275	GG	GG	GG	TG	TG	GG	GG	GG	TG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	TT	GG	TT	TG		
1	3772359	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TT	CC	CC		
1	3773089	CC	CC	TC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	CC	TT	TT		
1	3773252	CC	CC	CC	AC	AC	CC	CC	CC	AC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	CC	AA	AC		
1	3773718	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
1	3774138	CC	CC	TC	TC	TC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	CC	TT	TT		***DFFB***
1	3774549	TT	TT	TC	TC	TC	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	TT	CC	CC		
1	3778873	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA		
1	3779818	AA	AA	AG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AG		
1	3781096	AA	AA	AC	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AC		
1	3781830	AA	AA	AC	AC	AC	AA	AA	AA	AC	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AC	CC	AA	CC	CC		
1	3781856	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3782228	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
1	3783278	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	CC	CC		
1	3784454	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT	TC	TC		
1	3784464	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	AG	GG		
1	3785375	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG		
1	3785538	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		

DFFB (cont.)

1	3786055	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3786189	GG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3786245	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG
1	3786612	CC	CC	CC	CC	TC	TT	TT	TT	TT	TC	TC	CC	TC	TC	TT	CC	TC	TC	TT	TT	CC	TC	TC	CC	TT	CC	TC	TC
1	3786794	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3787024	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3788499	AA	AA	AA	AA	AG	GG	GG	GG	GG	AG	AG	AA	AG	AG	GG	AA	AG	AG	GG	GG	AA	AG	AG	AA	AG	AA	AA	GG
1	3788638	GG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3789814	AA	AA	AA	AA	AC	AA	AA	AA	AC	AA	AA	AA	AA	AA	AA	AA	AC	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
1	3789852	CC	CC	CC	CC	CC	TT	TT	TT	TC	TC	TC	CC	TC	TC	TT	CC	CC	TC	TT	TT	CC	TC	TC	CC	TC	CC	CC	TC
1	3790332	TT	TT	TT	TT	TT	CC	CC	CC	TC	TC	TC	TT	TC	TC	CC	TT	TT	TC	CC	CC	TT	TC	TC	TT	TC	TT	TT	TC
1	3791533	GG	GG	GG	GG	GG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3791837	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
1	3792310	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	CC
1	3793180	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
1	3793324	AA	AA	AA	AA	AA	0	GG	GG	AG	AG	AG	AA	GG	AG	GG	AA	AA	AG	GG	GG	AA	AG	AG	AA	AG	AG	AA	AG
1	3794273	TC	TC	TC	TC	TC	CC	CC	CC	CC	CC	CC	CC	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TC	TC	TC	CC
1	3794421	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3794504	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3795092	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3795321	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
1	3795391	TC	TC	TC	TC	TC	CC	CC	CC	CC	TC	TC	TT	CC	TC	CC	TT	TC	TC	CC	CC	TT	TC	TT	TT	CC	TC	TC	TC
1	3795485	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
1	3795516	GG	GG	GG	GG	GG	GG	GG	GG	TG	GG	GG	GG	GG	GG	GG	GG	TG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3795666	AA	AA	AA	AA	AA	GG	GG	GG	AG	AG	AG	AA	AG	AG	GG	AA	AA	AG	GG	GG	AA	AG	AA	AA	AG	AA	AA	AG
1	3795855	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3796237	AA	AA	AA	AA	AA	GG	GG	GG	AG	AG	AG	AA	AG	AG	GG	AA	AA	AG	GG	GG	AA	AG	AA	AA	AA	AA	AA	AA
1	3796542	AG	AG	AG	AG	AG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	GG	GG	GG	GG	GG	AG	AG	AG	GG
1	3796948	GG	GG	GG	GG	GG	TT	TT	TT	TG	TG	TG	GG	TG	TG	TT	GG	TT	TG	TT	TT	GG	TG	GG	GG	TG	GG	GG	TG
1	3797119	GG	GG	GG	GG	GG	GG	0	0	GG	GG	GG	GG	GG	GG	0	GG	GG	GG	GG	0	GG	GG	GG	GG	GG	GG	GG	GG
1	3797428	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3797919	TT	TT	TC	TT	TT	TT	TT	TT	TT	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TC	TC	TT	TT	TT
1	3798600	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
1	3798674	TC	TC	TC	TC	TC	CC	CC	CC	CC	TC	TC	TT	0	TC	0	TT	TC	TC	CC	CC	TT	TC	TT	TT	CC	TC	TC	TC
1	3799252	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
1	3800118	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
1	3800242	AG	AG	AG	AG	AG	AA	AA	AA	AG	AA	AA	AA	AG	AA	AA	AA	AG	AG	AA	AA	AA	AA	AA	AA	AG	AG	AG	AA
1	3801611	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT
1	3801973	TT	TT	TT	TT	TT	CC	CC	CC	TC	TC	TC	TT	TC	TC	CC	TT	TT	TT	CC	CC	TT	TC	TT	TT	TC	TT	TT	TC
1	3802392	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA

1	3803278	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
1	3803755	CC	CC	TC	CC	TC	CC	CC	CC	CC	TC	TC	TT	CC	TC	CC	CC	TC	TC	CC	CC	TC	TC	TT	TT	CC	CC	CC	TC			
1	3804543	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA			
1	3804683	TC	TC	CC	TC	CC	CC	CC	CC	TC	CC	CC	CC	TC	CC	CC	TT	TC	TC	CC	CC	TC	CC	CC	CC	CC	TT	TC	CC			
1	3805098	CC	CC	CC	CC	CC	TT	TT	TT	TC	TC	TC	CC	TC	TC	TT	CC	CC	CC	TT	TT	CC	TC	CC	CC	TC	CC	CC	TC			
1	3805255	AA	AA	AA	AA	AA	GG	GG	GG	AG	AG	AG	AA	AG	AG	GG	AA	AA	AA	GG	GG	AA	AG	AA	AA	AG	AA	AA	AG			
1	3805868	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA			*C1orf174*
1	3806286	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3806335	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3806335	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3806358	TC	TC	TC	TC	TC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TC	TT			
1	3806548	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3806973	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA			
1	3807258	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
1	3807360	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3807388	GG	GG	GG	GG	GG	AA	AA	AA	AG	AG	AG	GG	AG	AG	AA	GG	GG	GG	AA	AA	GG	AG	GG	GG	AG	GG	GG	AG			
1	3807450	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3807593	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
1	3808127	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3808362	AG	AG	AG	AG	AG	AA	AA	AA	AG	AG	AG	GG	AG	AG	AA	GG	GG	GG	AA	AA	GG	AG	GG	GG	AA	GG	AG	AG			
1	3809520	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3809528	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3811188	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3811790	GG	GG	GG	GG	GG	TT	TT	TT	TG	TG	TG	GG	TG	TG	TT	GG	GG	GG	TT	TT	GG	TG	GG	GG	TG	GG	GG	TG			
1	3811810	TT	TT	TT	TT	TT	CC	CC	CC	TC	TC	TC	TT	TC	TC	CC	TT	TT	TT	CC	CC	TT	TC	TT	TT	TC	TT	TT	TC			
1	3811848	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3812920	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3814607	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
1	3816950	CC	CC	CC	CC	CC	TT	TT	TT	TC	TC	TC	CC	TC	TC	TT	CC	CC	CC	TT	TT	CC	TC	CC	CC	TC	CC	CC	TC			*LOC100133612*
1	3817252	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TC	TT	TT	TT	TT	TT	TC	TT			
1	3818099	TT	TT	TT	TT	TT	GG	GG	GG	TG	TG	TG	TT	TG	TG	GG	TT	TT	TT	GG	GG	TT	TG	TT	TT	TG	TT	TT	TG			
1	3821216	AG	AG	AG	AG	AG	AA	AA	AA	AG	AG	AG	GG	AG	AG	AA	GG	GG	GG	AA	AA	GG	AG	GG	GG	AA	GG	AG	AG			
1	3821782	AG	AG	GG	AG	GG	GG	GG	GG	AG	GG	GG	GG	AG	GG	GG	AA	AG	AG	GG	GG	GG	GG	GG	GG	GG	AA	GG	GG			
1	3822283	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3822314	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC			
1	3822703	TC	TC	TC	TC	TC	TT	TT	TT	TC	TC	TC	CC	TC	TC	TT	CC	CC	CC	TT	TT	CC	TC	CC	CC	TT	CC	TC	TC			
1	3823150	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG			
1	3823395	TC	TC	CC	TC	CC	CC	CC	CC	TC	CC	CC	CC	TC	CC	CC	TT	TC	TC	CC	CC	CC	CC	CC	CC	CC	TT	CC	CC			

Supplementary Table 3: Genes in the identified region at 1p36

This table lists the five genes in the identified region at 1p36, along with their predicted number of amino acids, predicted molecular weights, and annotated functions as based on the information in the Entrez Gene database.

Symbol	Name	AAs* (kDa)	Functional annotation
CCDC27	Coiled-coil domain-containing protein 27	656 (75.4)	no annotation
SMIM1	Small integral membrane protein 1	78 (8.7)	single-pass transmembrane protein
LRRC47	Leucine-rich repeat-containing protein 47	583 (63.4)	RNA-binding, phenylalanine-tRNA ligase
CEP104	Centrosomal protein 104kDa	925 (104.4)	centriole, cytoskeleton
DFFB	DNA fragmentation factor, 40kDa, beta polypeptide	338 (39.1)	nucleus, apoptosis factor

*) predicted/annotated number of amino acids.

Supplementary Table 4: ChIP-seq peaks

The table lists all transcription sites detected by ChIP-seq (chromatin immunoprecipitation followed by massively parallel sequencing) in *SMIM1*, including 500 bp upstream and downstream of the gene, in the ENCODE project[1] (data obtained through UCSC Genome Browser).

Antibody target	Antibody	Peak span(s)
BCL3	BCL3	Intron 2
		3′ region
BHLHE40	BHLHE40	Exon 4 – 3′ region
CCNT2	CCNT2	5′ region – Exon 2
		Intron 2
		Exon 3 – Intron 3
CEBPB	CEBPB	Intron 2
CTCF	CTCF	Exon 4 – 3′ region
	CTCF_(C-20)	3′ region
	CTCF_(SC-5916)	3′ region
CTCFL	CTCFL_(SC-98982)	3′ region – 3′ region
E2F6	E2F6	Intron 2
		Exon 4 – 3′ region
	E2F6_(H-50)	Intron 2
		Intron 2 – Intron 3
		Exon 4 – 3′ region
EBF1	EBF1_(C-8)	3′ region – 3′ region
EGR1	Egr-1	5′ region – Intron 1
		Intron 2
		Exon 3 – Intron 3
ELF1	ELF1_(SC-631)	5′ region
		Intron 3 – Exon 4
		3′ region
ETS1	ETS1	Intron 2
		Intron 2 – Intron 3
ETS1	ETS1	Exon 4 – 3′ region
FOXA1	FOXA1_(C-20)	Intron 2
	FOXA1_(SC-101058)	Intron 2
FOXA2	FOXA2_(SC-6554)	Intron 2
GABP	GABP	Intron 2
GATA1	GATA-1	Intron 2 – Exon 3
GATA2	eGFP-GATA2	Intron 2
	GATA-2	Intron 2
	GATA2_(CG2-96)	Intron 2
HEY1	HEY1	5′ region
		5′ region – Intron 1
		Exon 3 – Intron 3
HMGN3	HMGN3	Intron 2
		Exon 3 – Intron 3
HNF4A	HNF4A	Intron 3 – Exon 4
	HNF4A_(H-171)	Intron 3 – Exon 4
HNF4G	HNF4G_(SC-6558)	Intron 3 – Exon 4
IRF1	IRF1	Exon 3 – Intron 3
JUND	eGFP-JunD	Intron 2

		Exon 4 – 3′ region
MAX	Max	3′ region
MXI1	Mxi1_(bHLH)	Intron 2
		3′ region
MYC	c-Myc	3′ region
		Intron 2
NFYB	NF-YB	Intron 3 – Exon 4
NR3C1	GR	Intron 2
POLR2A	Pol2	5′ region
		Intron 2
		Exon 3 – Exon 4
	Pol2(b)	Intron 2
	Pol2-4H8	5′ region
		Intron 2
		3′ region
RAD21	Rad21	3′ region
REST	NRSF	Intron 2
		Exon 4 – 3′ region
RXRA	RXRA	Intron 3 – Exon 4
SETDB1	SETDB1	3′ region
SIN3AK20	Sin3Ak-20	5′ region
		Intron 2
		3′ region
SMC3	SMC3_(ab9263)	Exon 4 – 3′ region
SP1	SP1	Intron 2 – Intron 2
		Intron 3 – Exon 4
TAF1	TAF1	5′ region
		Intron 2
TAL1	TAL1_(SC-12984)	Intron 2
		Exon 3 – Intron 3
		3′ region
TBP	TBP	5′ region
		Intron 2
		Intron 2
TCF12	TCF12	3′ region
TCF4	TCF4	3′ region
USF1	USF-1	Intron 2
		3′ region
	USF1_(SC-8983)	3′ region
YY1	YY1_(C-20)	3′ region
ZBTB7A	ZBTB7A_(SC-34508)	5′ region
		Exon 1 – Intron 1
		Intron 2
		Intron 2 – Exon 3

References

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

Supplementary Table 5: Transmembrane region predictions

Locations of the transmembrane domain in SMIM1, as predicted by nine different computational sequence analysis tools.

Tool	TM domain start	TM domain stop	Reference
HMMTOP	47	69	1
PHOBIUS	49	74	2
PRED-TMR	53	69	3
SOSUI	47	69	4
SPLIT	48	71	5
TMAP	36	64	6
TMHMM	53	75	7
TMPRED	53	74	8
TOPPRED	53	73	9
Median value	49	71	

References

1. Tusnady, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).

2. Kall, L., Krogh, A. & Sonnhammer, E. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).

3. Pasquier, C. & Hamodrakas, S. J. An hierarchical artificial neural network system for the classification of transmembrane proteins. *Protein Eng.* **12**, 631–634 (1999).

4. Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379 (1998).

5. Juretic, D., Zoranic, L. & Zucic, D. Basic charge clusters and predictions of membrane protein topology. *J Chem Inf Comput Sci* **42**, 620–632 (2002).

6. Persson, B. & Argos, P. Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.* **237**, 182–192 (1994).

7. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

8. Hofmann, K. TMbase-a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993).

9. von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487–494 (1992).

Supplementary Table 6: *SMIM1* variants in larger populations

This table lists all variants in *SMIM1* (formerly *LOC388588*) detected by whole-exome sequencing in the NHLBI Exome Sequencing Project (ESP) in European Americans (EA) and African Americans (AA). As shown, the 17-bp deletion is the most common non-consensus variant detected in EA, and the only variant whose allele frequency predicts a homozygote frequency that is on par with the prevalence of Vel– individuals among individuals of European ethnicity. The most common variant among AA is a noncoding GC/G indel of unknown significance in the three-prime untranslated region. The remaining variants occur at extremely low frequencies.

Chr 1 pos	Alleles (minor/major)	EA allele counts	AA allele counts	PolyPhen Prediction	Comment
3691980	A/G	1/3181	0/1384	benign	Asn15Asp
3691997	A/AGTCAGCCTAGGGGCTGT	57/5763	6/3198	unknown	frameshift
3692333	C/T	1/3181	0/1384	unknown	intronic
3692435	A/ATCT	7/5061	0/2834	unknown	in-frame deletion
3692460	G/GC	1/5021	0/2812	unknown	frameshift
3692468	A/G	1/3181	0/1384	benign	Met74Val
3692504	GC/G	1/5051	68/2728	unknown	3′-UTR

Data obtained through Exome Variant Server (http://evs.gs.washington.edu/EVS/; version November 2012). Fractions denote the number of variant alleles over the number of sequenced/covered alleles.

Supplementary Table 7: List of primers

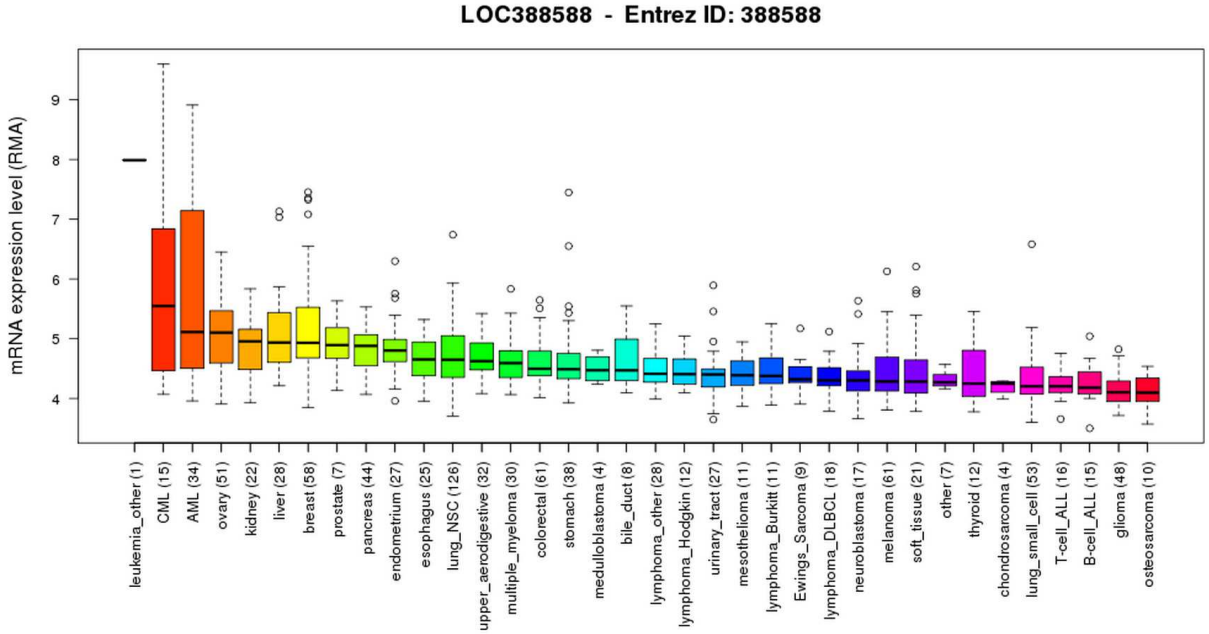
Primer name	Sequence (5′ to 3′)	Purpose	Product size (bp) wt/mut
388588cDNAf	CGCACCTCAGCCCACGAC	Amplification and sequencing	472/455
388588cDNAr	TCCAGGCCTGTGCTCTCAC		
VelF	GCCGAATTCGCCACCATGCAGCCCCAGGAGAGC	Cloning and expression	257/240
VelR2	GCCGGATCCCCCTTATTTGCACTTGTGCACATA		
388588int2f	TCTCCTAACAGCAGCCTCAG	Amplification and sequencing	745/728
388588ex4r	TGTCTCCAGGCCTGTGCTC		
388588int3f	CAGCTCAGCAAACCGCAGC	Sequencing	
388588int3r	GGCGCTCTGCTGGAGTCA		
388588int3r2	CTGGGCGCTCTGCTGGAG	Sequencing and ASP	
388588wtex3f	CGGAGTCAGCCTAGGGGC	ASP screening	266/–
388588mutex3f	GGACGGAGTCCAGCACAG	ASP screening	–/249
LOCex3f_screen	ACAGCCTGGCCACCTGTCTTG	GSP Screening	178/161
LOCex3r_screen	CTGCGGCAGCGTGAGGC		
Vel 59F	GGCCGCAGTGGGCAGGCTC	3′-RACE	
Vel 176F	CTCAGGCAAGGTTCTCCGGTGA		
Vel 280F	AGGAGAGCCACGTCCACTATAG		
Vel 332R	GCAGCGTGAGGCCTCTTCTGTG	5′-RACE	
Vel 355R	CACAGCCTCTGGGAGATCCTGC		
Vel 376R	GATGCCCAGCTTGCCCGTGC		

Supplementary Table 8: Peptides

This table lists four peptides designed from the predicted extracellular domain of SMIM1. Peptides 1 and 3 were selected to produce rabbit polyclonal antibodies to the predicted SMIM1 protein.

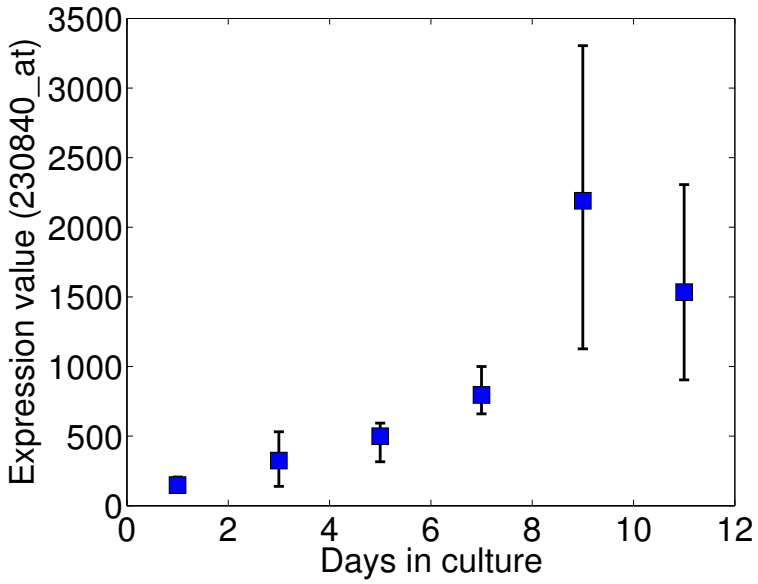
Name	Sequence
Peptide 1	MQPQESHVHYSRWED
Peptide 2	SRWEDGSRDGVSLGA
Peptide 3	GVSLGAVSSTEEASR
Peptide 4	EASRCRRISQRLCTG

Supplementary Figure 1: *SMIM1* transcript expression in cell lines



Expression of the *SMIM1* (formerly *LOC388588*) transcript, as measured by microarray across 974 cell lines in the Broad-Novartis Cancer Cell Line Encyclopedia (`http://www.broadinstitute/ccle/home`; version June 2012). As shown, *SMIM1* is highest expressed in cell lines derived from examples of Chronic Myeloid Leukemia (CML) or Acute Myeloid Leukemia (AML). The highest-expressing lines were HEL and JK-1, which have erythroid differentiation. The erythroleukemic cell line K562 used for transfection experiments expresses *SMIM1* at moderate levels (not shown).

Supplementary Figure 2: *SMIM1* expression in CD34+ hematopoietic progenitors cultured towards RBCs



This figure shows the expression of *SMIM1* as measured on Affymetrix microarrays. In this experiment, which was performed by Keller and coworkers[1], adult differentiating CD34+ hematopoietic progenitor cells were cultured towards RBCs, and gene expression-profiled at various time points up to eleven days of growth in serum-free medium containing erythropoietin, interleukin 3 and stem cell factor. The purpose of the experiment was to identify transcriptional regulators involved in erythroid development at a global level. We specifically extracted the *SMIM1*-targeting probe from this data set (Affymetrix probe set identifier 230840_at). As shown, *SMIM1* is induced around day 9. The original data set is available in NCBI Gene Expression Omnibus, accession no. GSE4655. The experiment was performed in triplicate. Error bars indicate min to max. In accordance with the above, our data based on cDNA amplification of RNA derived from bone marrow cells cultured towards erythropoiesis[2] confirms the presence of *SMIM1* transcripts on days 3 and 9 in culture, with a tendency towards increased levels on day 9 (Data not shown).

References

1. Keller *et al.* Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators. *Physiol Genomics* **28**, 114–28 (2006).

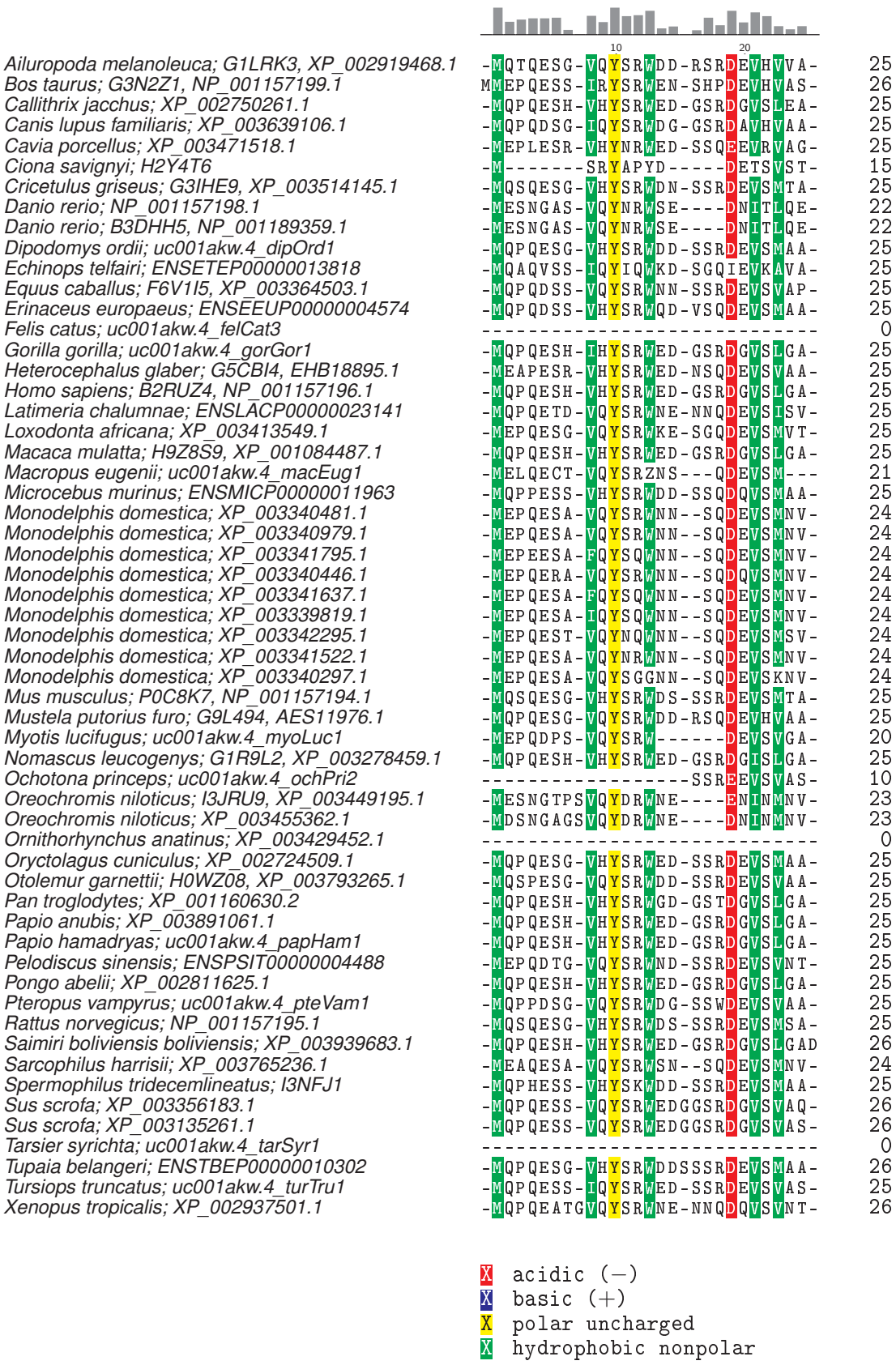
2. Edvardsson, L., Dykes, J., Olsson, M. L. & Olofsson, T. Clonogenicity, gene expression and phenotype during neutrophil versus erythroid differentiation of cytokine-stimulated CD34+ human marrow cells in vitro. *Br J Haematol* **127**, 451–463 (2004).

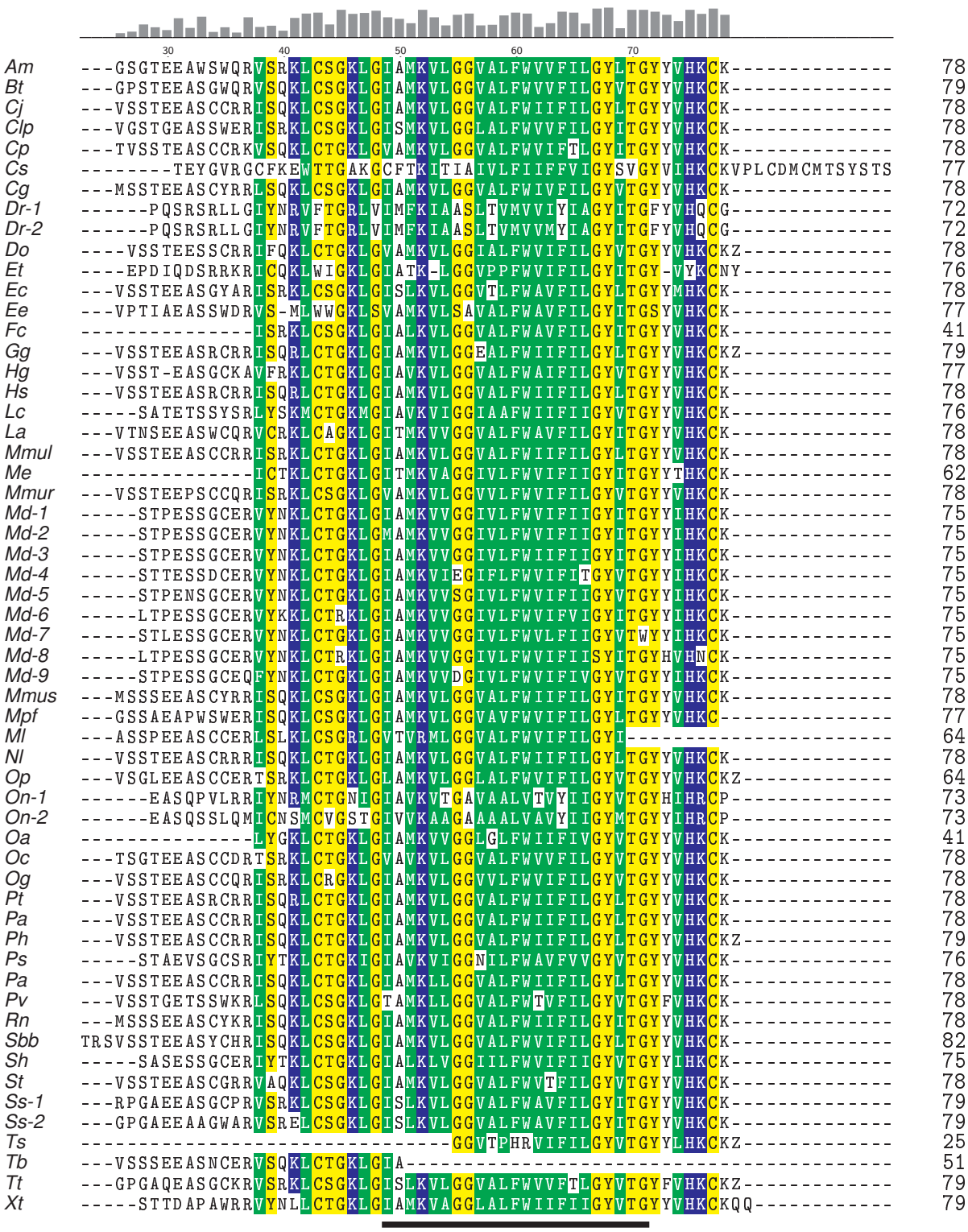
Supplementary Figure 3: Cross-species alignment

Multiple sequence alignment for all homologous proteins found with BLASTP in the Refseq non-redundant and UniProtKB databases, in alignment data from UCSC Genome Browser (table knownGene.exonAA) or in Ensembl (release 68). An *E*-value of 1 was used as cut-off in BLASTP searches. No human homologs were found (human SMIM1 is included for reference), but a total of 56 protein orthologs with significant sequence similarity in 45 additional species.

Rows are labelled with species name and source (description below). Multiple sequence alignment was performed using MUSCLE v. 3.8.31 with default parameters. Bars at top indicate degree of conservation. Conserved amino acids are color coded according to their hydrophobic properties. The black bar at the bottom indicate the position of the predicted transmembrane domain in human, as determined by the computational tools listed in **Supplementary Table 5**.

Entry starting with	**Source**
Six alphanumeric characters	UniProtKB
*XP_* or *NP_*	Refseq non-redundant
uc001akw	UCSC knownGene.exonAA
ENS	Ensembl

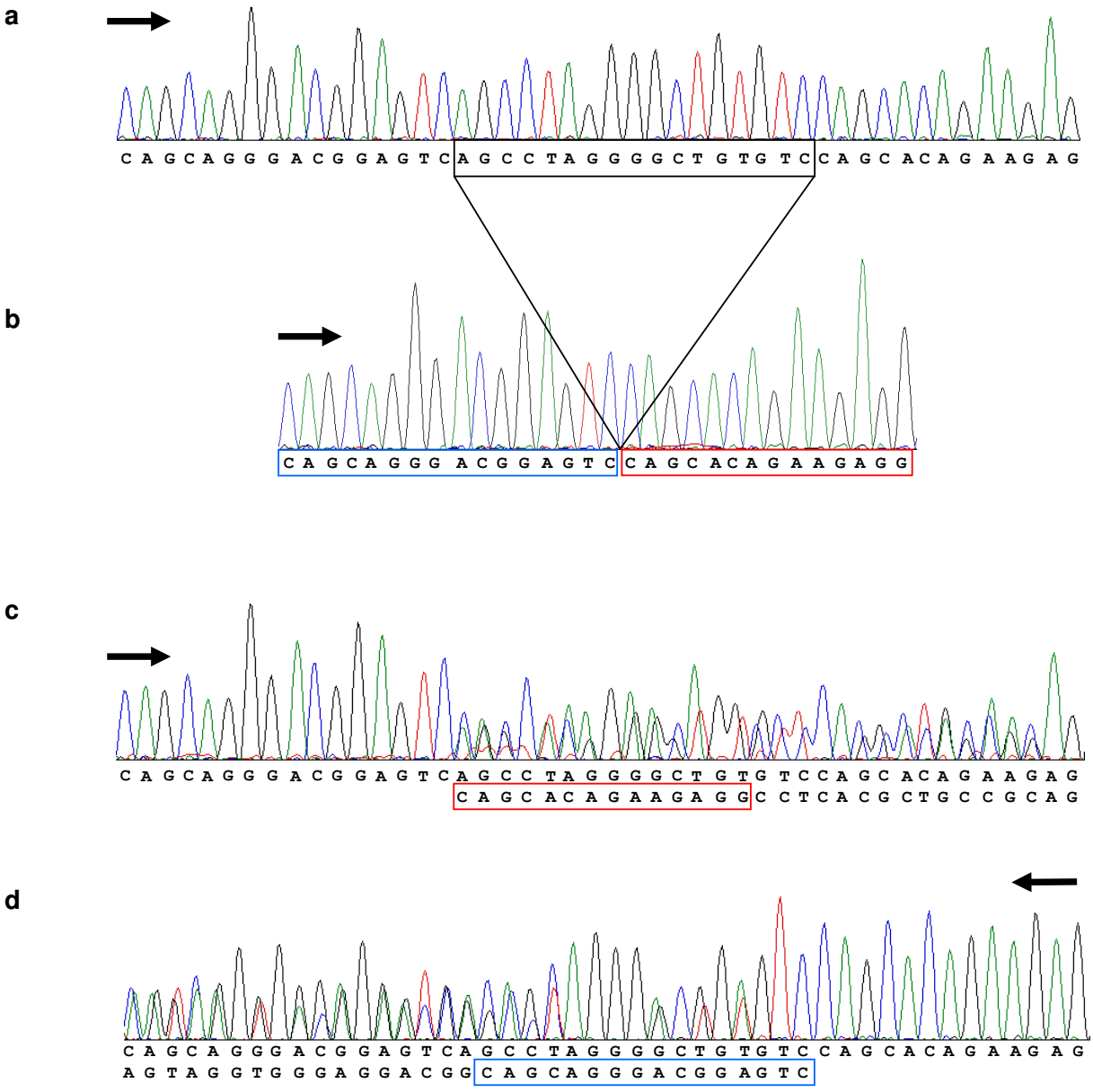

10
20
Ailuropoda melanoleuca; G1LRK3, XP_002919468.1 -MQTQESG-VQYSRWDD-RSRDEVHVVA- 25
Bos taurus; G3N2Z1, NP_001157199.1 MMEPQESS-IRYSRWEN-SHPDEVHVAS- 26
Callithrix jacchus; XP_002750261.1 -MQPQESH-VHYSRWED-GSRDGVSLEA- 25
Canis lupus familiaris; XP_003639106.1 -MQPQDSG-IQYSRWDG-GSRDAVHVAA- 25
Cavia porcellus; XP_003471518.1 -MEPLESR-VHYNRWED-SSQEEVRVAG- 25
Ciona savignyi; H2Y4T6 -M-------SRYAPYD-----DETSVST- 15
Cricetulus griseus; G3IHE9, XP_003514145.1 -MQSQESG-VHYSRWDN-SSRDEVSMTA- 25
Danio rerio; NP_001157198.1 -MESNGAS-VQYNRWSE----DNITLQE- 22
Danio rerio; B3DHH5, NP_001189359.1 -MESNGAS-VQYNRWSE----DNITLQE- 22
Dipodomys ordii; uc001akw.4_dipOrd1 -MQPQESG-VHYSRWDD-SSRDEVSMAA- 25
Echinops telfairi; ENSETEP00000013818 -MQAQVSS-IQYIQWKD-SGQIEVKAVA- 25
Equus caballus; F6V1I5, XP_003364503.1 -MQPQDSS-VQYSRWNN-SSRDEVSVAP- 25
Erinaceus europaeus; ENSEEUP00000004574 -MQPQDSS-VHYSRWQD-VSQDEVSMAA- 25
Felis catus; uc001akw.4_felCat3 ---------- 0
Gorilla gorilla; uc001akw.4_gorGor1 -MQPQESH-IHYSRWED-GSRDGVSLGA- 25
Heterocephalus glaber; G5CBI4, EHB18895.1 -MEAPESR-VHYSRWED-NSQDEVSVAA- 25
Homo sapiens; B2RUZ4, NP_001157196.1 -MQPQESH-VHYSRWED-GSRDGVSLGA- 25
Latimeria chalumnae; ENSLACP00000023141 -MQPQETD-VQYSRWNE-NNQDEVSISV- 25
Loxodonta africana; XP_003413549.1 -MEPQESG-VQYSRWKE-SGQDEVSMVT- 25
Macaca mulatta; H9Z8S9, XP_001084487.1 -MQPQESH-VHYSRWED-GSRDGVSLGA- 25
Macropus eugenii; uc001akw.4_macEug1 -MELQECT-VQYSRZNS---QDEVSM--- 21
Microcebus murinus; ENSMICP00000011963 -MQPPESS-VHYSRWDD-SSQDQVSMAA- 25
Monodelphis domestica; XP_003340481.1 -MEPQESA-VQYSRWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003340979.1 -MEPQESA-VQYSRWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003341795.1 -MEPEESA-FQYSQWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003340446.1 -MEPQERA-VQYSRWNN--SQDQVSMNV- 24
Monodelphis domestica; XP_003341637.1 -MEPQESA-FQYSQWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003339819.1 -MEPQESA-IQYSQWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003342295.1 -MEPQEST-VQYNQWNN--SQDEVSMSV- 24
Monodelphis domestica; XP_003341522.1 -MEPQESA-VQYNRWNN--SQDEVSMNV- 24
Monodelphis domestica; XP_003340297.1 -MEPQESA-VQYSGGNN--SQDEVSKNV- 24
Mus musculus; P0C8K7, NP_001157194.1 -MQSQESG-VHYSRWDS-SSRDEVSMTA- 25
Mustela putorius furo; G9L494, AES11976.1 -MQPQESG-VQYSRWDD-RSQDEVHVAA- 25
Myotis lucifugus; uc001akw.4_myoLuc1 -MEPQDPS-VQYSRW------DEVSVGA- 20
Nomascus leucogenys; G1R9L2, XP_003278459.1 -MQPQESH-VHYSRWED-GSRDGISLGA- 25
Ochotona princeps; uc001akw.4_ochPri2 ----------SSREEVSVAS- 10
Oreochromis niloticus; I3JRU9, XP_003449195.1 -MESNGTPSVQYDRWNE----ENINMNV- 23
Oreochromis niloticus; XP_003455362.1 -MDSNGAGSVQYDRWNE----DNINMNV- 23
Ornithorhynchus anatinus; XP_003429452.1 ---------- 0
Oryctolagus cuniculus; XP_002724509.1 -MQPQESG-VHYSRWED-SSRDEVSMAA- 25
Otolemur garnettii; H0WZ08, XP_003793265.1 -MQSPESG-VQYSRWDD-SSRDEVSVAA- 25
Pan troglodytes; XP_001160630.2 -MQPQESH-VHYSRWGD-GSTDGVSLGA- 25
Papio anubis; XP_003891061.1 -MQPQESH-VHYSRWED-GSRDGVSLGA- 25
Papio hamadryas; uc001akw.4_papHam1 -MQPQESH-VHYSRWED-GSRDGVSLGA- 25
Pelodiscus sinensis; ENSPSIT00000004488 -MEPQDTG-VQYSRWND-SSRDEVSVNT- 25
Pongo abelii; XP_002811625.1 -MQPQESH-VHYSRWED-GSRDGVSLGA- 25
Pteropus vampyrus; uc001akw.4_pteVam1 -MQPPDSG-VQYSRWDG-SSWDEVSVAA- 25
Rattus norvegicus; NP_001157195.1 -MQSQESG-VHYSRWDS-SSRDEVSMSA- 25
Saimiri boliviensis boliviensis; XP_003939683.1 -MQPQESH-VHYSRWED-GSRDGVSLGAD 26
Sarcophilus harrisii; XP_003765236.1 -MEAQESA-VQYSRWSN--SQDEVSMNV- 24
Spermophilus tridecemlineatus; I3NFJ1 -MQPHESS-VHYSKWDD-SSRDEVSMAA- 25
Sus scrofa; XP_003356183.1 -MQPQESS-VQYSRWEDGGSRDGVSVAQ- 26
Sus scrofa; XP_003135261.1 -MQPQESS-VQYSRWEDGGSRDGVSVAS- 26
Tarsier syrichta; uc001akw.4_tarSyr1 ---------- 0
Tupaia belangeri; ENSTBEP00000010302 -MQPQESG-VHYSRWDDSSSRDEVSMAA- 26
Tursiops truncatus; uc001akw.4_turTru1 -MQPQESS-IQYSRWED-SSRDEVSVAS- 25
Xenopus tropicalis; XP_002937501.1 -MQPQEATGVQYSRWNE-NNQDQVSVNT- 26
X acidic (−)
X basic (+)
X polar uncharged
X hydrophobic nonpolar


30
40
50
60
70
Am ---GSGTEEAWSWQRVSRKLCSGKLGIAMKVLGGVALFWVVFILGYLTGYYVHKCK---------- 78
Bt ---GPSTEEASGWQRVSQKLCSGKLGIAMKVLGGVALFWVVFILGYVTGYYVHKCK---------- 79
Cj ---VSSTEEASCCRRISQKLCSGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Clp ---VGSTGEASSWERISRKLCSGKLGISMKVLGGLALFWVVFILGYITGYYVHKCK---------- 78
Cp ---TVSSTEASCCRKVSQKLCTGKLGVAMKVLGGVALFWVIFTLGYITGYYVHKCK---------- 78
Cs --------TEYGVRGCFKEWTTGAKGCFTKITIAIVLFIIFFVIGYSVGYVIHKCKVPLCDMCMTSYSTS 77
Cg ---MSSTEEASCYRRLSQKLCSGKLGIAMKVLGGVALFWIVFILGYVTGYYVHKCK---------- 78
Dr-1 ------PQSRSRLLGIYNRVFTGRLVIMFKIAASLTVMVVIYIAGYITGFYVHQCG---------- 72
Dr-2 ------PQSRSRLLGIYNRVFTGRLVIMFKIAASLTVMVVMYIAGYITGFYVHQCG---------- 72
Do ----VSSTEESSCRRIFQKLCTGKLGVAMKVLGGIALFWVIFILGYVTGYYVHKCKZ---------- 78
Et ----EPDIQDSRRKRICQKLWIGKLGIATK-LGGVPPFWVIFILGYITGY-VYKCNY---------- 76
Ec ---VSSTEEASGYARISRKLCSGKLGISLKVLGGVTLFWAVFILGYLTGYYMHKCK---------- 78
Ee ---VPTIAEASSWDRVS-MLWWGKLSVAMKVLSAVALFWAVFILGYITGSYVHKCK---------- 77
Fc ---------------ISRKLCSGKLGIALKVLGGVALFWAVFILGYITGYYVHKCK---------- 41
Gg ---VSSTEEASRCRRISQRLCTGKLGIAMKVLGGEALFWIIFILGYLTGYYVHKCKZ---------- 79
Hg ---VSST-EASGCKAVFRKLCTGKLGIAVKVLGGVALFWAIFILGYVTGYYVHKCK---------- 77
Hs ---VSSTEEASRCRRISQRLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Lc -----SATETSSYSRLYSKMCTGKMGIAVKVIGGIAAFWIIFIIGYVTGYYVHKCK---------- 76
La ---VTNSEEASWCQRVCRKLCAGKLGITMKVVGGVALFWAVFILGYITGYYVHKCK---------- 78
Mmul ---VSSTEEASCCRRISRKLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Me ---------------ICTKLCTGKLGITMKVAGGIVLFWVIFIIGYITGYYTHKCK---------- 62
Mmur ---VSSTEEPSCCQRISRKLCSGKLGVAMKVLGGVVLFWVIFILGYVTGYYVHKCK---------- 78
Md-1 -----STPESSGCERVYNKLCTGKLGIAMKVVGGIVLFWVIFIIGYITGYYIHKCK---------- 75
Md-2 -----STPESSGCERVYNKLCTGKLGMAMKVVGGIVLFWVIFIIGYITGYYIHKCK---------- 75
Md-3 -----STPESSGCERVYNKLCTGKLGIAMKVVGGIVLFWIIFIIGYITGYYIHKCK---------- 75
Md-4 -----STTESSDCERVYNKLCTGKLGIAMKVIEGIFLFWVIFITGYVTGYYIHKCK---------- 75
Md-5 -----STPENSGCERVYNKLCTGKLGIAMKVVSGIVLFWVIFIIGYVTGYYIHKCK---------- 75
Md-6 -----LTPESSGCERVYKKLCTRKLGIAMKVVGGIVLFWVIFVIGYITGYYIHKCK---------- 75
Md-7 -----STLESSGCERVYNKLCTGKLGIAMKVVGGIVLFWVLFIIGYVTWYYIHKCK---------- 75
Md-8 -----LTPESSGCERVYNKLCTRKLGIAMKVVGGIVLFWVIFIISYITGYHVHNCK---------- 75
Md-9 -----STPESSGCEQFYNKLCTGKLGIAMKVVDGIVLFWVIFIVGYVTGYYIHKCK---------- 75
Mmus ---MSSSEEASCYRRISQKLCSGKLGIAMKVLGGVALFWIIFILGYITGYYVHKCK---------- 78
Mpf ---GSSAEAPWSWERISQKLCSGKLGIAMKVLGGVAVFWVIFILGYLTGYYVHKC---------- 77
Ml ---ASSPEEASCCERLSLKLCSGRLGVTVRMLGGVALFWVIFILGYI---------- 64
Nl ---VSSTEEASCRRRISQKLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Op ---VSGLEEASCCERTSRKLCTGKLGLAMKVLGGLALFWVIFILGYVTGYYVHKCKZ---------- 64
On-1 ------EASQPVLRRIYNRMCTGNIGIAVKVTGAVAALVTVYIIGYVTGYHIHRCP---------- 73
On-2 ------EASQSSLQMICNSMCVGSTGIVVKAAGAAAALVAVYIIGYMTGYYIHRCP---------- 73
Oa ---------------LYGKLCTGKLGIAMKVVGGLGLFWIIFIVGYVTGYYVHKCK---------- 41
Oc ---TSGTEEASCCDRTSRKLCTGKLGVAVKVLGGVALFWVVFILGYVTGYYVHKCK---------- 78
Og ---VSSTEEASCCQRISRKLCRGKLGIAMKVLGGVVLFWVIFILGYVTGYYVHKCK---------- 78
Pt ---VSSTEEASRCRRISQRLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Pa ---VSSTEEASCCRRISQKLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Ph ---VSSTEEASCCRRISQKLCTGKLGIAMKVLGGVALFWIIFILGYLTGYYVHKCKZ---------- 79
Ps -----STAEVSGCSRIYTKLCTGKIGIAVKVIGGNILFWAVFVVGYVTGYYVHKCK---------- 76
Pa ---VSSTEEASCCRRISQKLCTGKLGIAMKLLGGVALFWIIFILGYLTGYYVHKCK---------- 78
Pv ---VSSTGETSSWKRLSQKLCSGKLGTAMKLLGGVALFWTVFILGYVTGYFVHKCK---------- 78
Rn ---MSSSEEASCYKRISQKLCSGKLGIAMKVLGGVALFWIIFILGYITGYYVHKCK---------- 78
Sbb TRSVSSTEEASYCHRISQKLCSGKLGIAMKVLGGVALFWVIFILGYLTGYYVHKCK---------- 82
Sh -----SASESSGCERIYTKLCTGKLGIALKLVGGIILFWVIFIIGYVTGYYIHKCK---------- 75
St ---VSSTEEASCGRRVAQKLCSGKLGIAMKVLGGVALFWVTFILGYVTGYYVHKCK---------- 78
Ss-1 ---RPGAEEASGCPRVSRKLCSGKLGISLKVLGGVALFWAVFILGYITGYYVHKCK---------- 79
Ss-2 ---GPGAEEAAGWARVSRELCSGKLGISLKVLGGVALFWAVFILGYITGYYVHKCK---------- 79
Ts ------------------------------GGVTPHRVIFILGYVTGYYLHKCKZ---------- 25
Tb ---VSSSEEASNCERVSQKLCTGKLGIA---------- 51
Tt ---GPGAQEASGCKRVSRKLCSGKLGISLKVLGGVALFWVVFTLGYVTGYFVHKCKZ---------- 79
Xt -----STTDAPAWRRVYNLLCTGKLGIAMKVAGGLALFWIIFIIGYVTGYYVHKCKQQ---------- 79

Am	----------	78
Bt	----------	79
Cj	----------	78
Clp	----------	78
Cp	----------	78
Cs	SMVTVTSSNSTT	89
Cg	----------	78
Dr-1	----------	72
Dr-2	----------	72
Do	----------	78
Et	----------	76
Ec	----------	78
Ee	----------	77
Fc	----------	41
Gg	----------	79
Hg	----------	77
Hs	----------	78
Lc	----------	76
La	----------	78
Mmul	----------	78
Me	----------	62
Mmur	----------	78
Md-1	----------	75
Md-2	----------	75
Md-3	----------	75
Md-4	----------	75
Md-5	----------	75
Md-6	----------	75
Md-7	----------	75
Md-8	----------	75
Md-9	----------	75
Mmus	----------	78
Mpf	----------	77
Ml	----------	64
Nl	----------	78
Op	----------	64
On-1	----------	73
On-2	----------	73
Oa	----------	41
Oc	----------	78
Og	----------	78
Pt	----------	78
Pa	----------	78
Ph	----------	79
Ps	----------	76
Pa	----------	78
Pv	----------	78
Rn	----------	78
Sbb	----------	82
Sh	----------	75
St	----------	78
Ss-1	----------	79
Ss-2	----------	79
Ts	----------	25
Tb	----------	51
Tt	----------	79
Xt	----------	79

Supplementary Figure 4: Sequence data

This figure shows the DNA sequencing results for **(a)** a homozygous Vel+ individual; **(b)** a Vel– individual; **(c)**, **(d)** a Vel+ heterozygous individual. Panel **(c)** shows sequencing results initiated with a primer in intron 2; **(d)** shows sequencing initiated by a primer in intron 3. The direction of the sequencing primer is indicated by the black arrow. The nucleotides deleted in the Vel– individual are shown by the black box. The sequence 5′ of the deletion is boxed in blue. The sequence 3′ of the deletion is boxed in red. In (c) and (d), the wildtype sequence is shown on top of the mutant sequence. The Vel– genomic sequence has been deposited in GenBank (accession number KC152643).



(e) Genomic sequence alignment of a 745 bp fragment of *SMIM1* spanning exons 3 and 4 compared with the same region sequenced from Vel– individuals. The exons are highlighted in gray. The differences between the sequences are indicated by asterisks; the initiation (blue) and stop (red) codons are underlined.

```
SMIM1     TCTCCTAACAGCAGCCTCAGAGGGGGTCTTGACTGCCGCCCTCCATCCGCTTGTTTTACA
Vel-      TCTCCTAACAGCAGCCTCAGAGGGGGTCTTGACTGCCGCCCTCCATCCGCTTGTTTTACA
          ------------------------------------------------------------

SMIM1     GTGAAGCCACAGCCTGGCCACCTGTCTTGATCTCCCCACCGAGAAGGCCCCGCCCCTCCC
Vel-      GTGAAGCCACAGCCTGGCCACCTGTCTTGATCTCCCCACCGAGAAGGCCCCGCCCCTCCC
          ------------------------------------------------------------

SMIM1     GCTGCAGCCCCACAGCATGCAGCCCCAGGAGAGCCACGTCCACTATAGTAGGTGGGAGGA
Vel-      GCTGCAGCCCCACAGCATGCAGCCCCAGGAGAGCCACGTCCACTATAGTAGGTGGGAGGA
          ------------------------------------------------------------

SMIM1     CGGCAGCAGGGACGGAGTCAGCCTAGGGGCTGTGTCCAGCACAGAAGAGGCCTCACGCTG
Vel-      CGGCAGCAGGGACGGAGTC-----------------CAGCACAGAAGAGGCCTCACGCTG
          ------------------**********----------------------------

SMIM1     CCGCAGGTGAGGGGCCTGAGGGCAGCCTGCCAGCCATAGCAGGCTGGTGTCTCCCTCCAG
Vel-      CCGCAGGTGAGGGGCCTGAGGGCAGCCTGCCAGCCATAGCAGGCTGGTGTCTCCCTCCAG
          ------------------------------------------------------------

SMIM1     AGACGCCTGCCCTAACCCCTGCTACCGGCCCCATCACCCTCCACCCCATCCTGGCTGGGA
Vel-      AGACGCCTGCCCTAACCCCTGCTACCGGCCCCATCACCCTCCACCCCATCCTGGCTGGGA
          ------------------------------------------------------------

SMIM1     GCCCACGGTCCAGCAGCTCAGCAAACCGCAGCCTTTGGCCTTCCCTCTGGTTGGCTGTGG
Vel-      GCCCACGGTCCAGCAGCTCAGCAAACCGCAGCCTTTGGCCTTCCCTCTGGTTGGCTGTGG
          ------------------------------------------------------------

SMIM1     GCGGGGAGAGCTTCCTCTTGACTCCAGCAGAGCGCCCAGGCCCCTCCCCCTGACCCAGAC
Vel-      GCGGGGAGAGCTTCCTCTCAACTCCAGCAGAGCGCCCAGGCCCCTCCCCCTGACCCAGAC
          ------------------**------------------------------------------

SMIM1     CAACGGCCACAGTCCACTTAGGGGGCCCCTCATGCGGCCCTGGCCTGGGGCTCACCTCCA
Vel-      CAACGGCCACAGTCCACTTAGGGGGCCCCTCATGCGGCCCTGGCCTGGGGCTCACCTCCA
          ------------------------------------------------------------

SMIM1     GTTGGTTCTCACCCCAGGATCTCCCAGAGGCTGTGCACGGGCAAGCTGGGCATCGCCATG
Vel-      GTTGGTTCTCACCCCAGGATCTCCCAGAGGCTGTGCACGGGCAAGCTGGGCATCGCCATG
          ------------------------------------------------------------

SMIM1     AAGGTGCTGGGCGGCGTGGCCCTCTTCTGGATCATCTTCATCCTGGGCTACCTCACAGGC
Vel-      AAGGTGCTGGGCGGCGTGGCCCTCTTCTGGATCATCTTCATCCTGGGCTACCTCACAGGC
          ------------------------------------------------------------

SMIM1     TACTATGTGCACAAGTGCAAATAAATGCTGCCCCGCATGCACGCGGGGGGCTGGCCGCAC
Vel-      TACTATGTGCACAAGTGCAAATAAATGCTGCCCCGCATGCACGCGGGGGGCTGGCCGCAC
          ------------------------------------------------------------

SMIM1     ACGTGAGAGCACAGGCCTGGAGACA
Vel-      ACGTGAGAGCACAGGCCTGGAGACA
          -------------------------
```

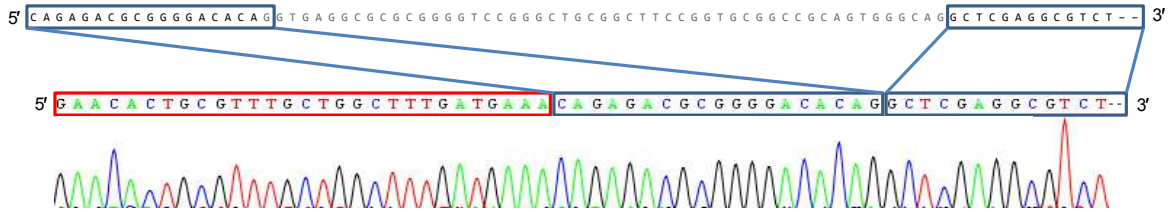
(f) The predicted nucleotide and amino acid sequence of the extended open reading frame in the mutated *SMIM1* sequence. Gray indicates the SMIM1 amino acid sequence until the frameshift. Thereafter, the sequence differs and transmembrane prediction analysis (**Supplementary Table 5**) does not indicate the presence of a transmembrane domain.

```
  1 ATGCAGCCCCAGGAGAGCCACGTCCACTATAGTAGGTGGGAGGACGGCAGCAGGGACGGA 60
  1 M  Q  P  Q  E  S  H  V  H  Y  S  R  W  E  D  G  S  R  D  G   20

 61 GTCCAGCACAGAAGAGGCCTCACGCTGCCGCAGGATCTCCCAGAGGCTGTGCACGGGCAA 120
 21 V  Q  H  R  R  G  L  T  L  P  Q  D  L  P  E  A  V  H  G  Q   40

121 GCTGGGCATCGCCATGAAGGTGCTGGGCGGCGTGGCCCTCTTCTGGATCATCTTCATCCT 180
 41 A  G  H  R  H  E  G  A  G  R  R  G  P  L  L  D  H  L  H  P   60

181 GGGCTACCTCACAGGCTACTATGTGCACAAGTGCAAATAAATGCTGCCCCGCATGCACGC 240
 61 G  L  P  H  R  L  L  C  A  Q  V  Q  I  N  A  A  P  H  A  R   80

241 GGGGGGCTGGCCGCACACGTGAGAGCACAGGCCTGGAGACACACCCCTTGTACACATGGA 300
 81 G  G  L  A  A  H  V  R  A  Q  A  W  R  H  T  P  C  T  H  G   100

301 CCCCCCCACAGACACGGACCCTGCGGCACACACAGCGCACAGGGCACACGCGCTGGCAGC 360
101 P  P  H  R  H  G  P  C  G  T  H  S  A  Q  G  T  R  A  G  S   120

361 CAGGCACACGAAGACACCAGGTGCACAGCTGTCATCGGCCCCACACGGGGGCGCACAAAC 420
121 Q  A  H  E  D  T  R  C  T  A  V  I  G  P  T  R  G  R  T  N   140

421 ACCTGGCACACAGCCCTTCAAAGGACCTACAAACAGCTGGGCACACGTGGCTGGGAGGCC 480
141 T  W  H  T  A  L  Q  R  T  Y  K  Q  L  G  T  R  G  W  E  A   160

481 TGGGCCCAGCCTCAGCAGGAGCTGCAGGACACACCCAGGCTGGGCCCTGCGGCCTGGAGC 540
161 W  A  Q  P  Q  Q  E  L  Q  D  T  P  R  L  G  P  A  A  W  S   180

541 CCCCAGCTACAGCCTCCTCTCTCCCAGGGCCCAGCCCCTTCCCTTGTGAAGGCCAGGATG 600
181 P  Q  L  Q  P  P  L  S  Q  G  P  A  P  S  L  V  K  A  R  M   200

601 AGGGGTTCCTTCAGCGGACAAACCGAGCCCACCTCCCTGGCAGCCCCCCGGGGTGGGATC 660
201 R  G  S  F  S  G  Q  T  E  P  T  S  L  A  A  P  R  G  G  I   220

661 CTCCCGGCTGCTTTCCTCCGTGGGAGCAGTGTGCAGAGCTGTGTGGCCCTGGGCAGGCCC 720
221 L  P  A  A  F  L  R  G  S  S  V  Q  S  C  V  A  L  G  R  P   240

721 CTGTCCTCTCTGGGCCTTTCTGACTCCTGGTTTTGTAAGGGTGGCTATGTGTCCCCCGCC 780
241 L  S  S  L  G  L  S  D  S  W  F  C  K  G  G  Y  V  S  P  A   260

781 CTTGTCTCAGATGCACCATATCTTCCTTAG 810
261 L  V  S  D  A  P  Y  L  P  *   269
```

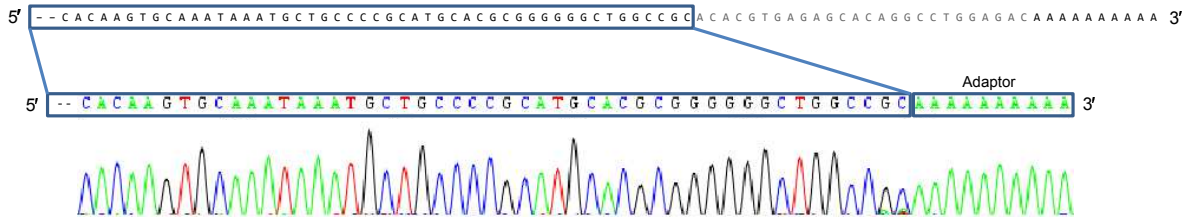
Supplementary Figure 5: RACE on Vel+ blood samples

5′- and 3′-RACE experiments revealed variations compared to the *SMIM1*-mRNA reference sequence (NCBI NM_001163724). An unannotated exon 1 of 19 bp was identified spanning 18 bp upstream of the original start site, and including the first nucleotide of the reference sequence, rendering the original exon 1 obsolete. The 3′ end was found to be 26 bp shorter than the reference. No discrepancies were found in the remaining sequence (not shown). The erythroid cDNA sequence has been deposited in GenBank (accession number KC152644).

a



b

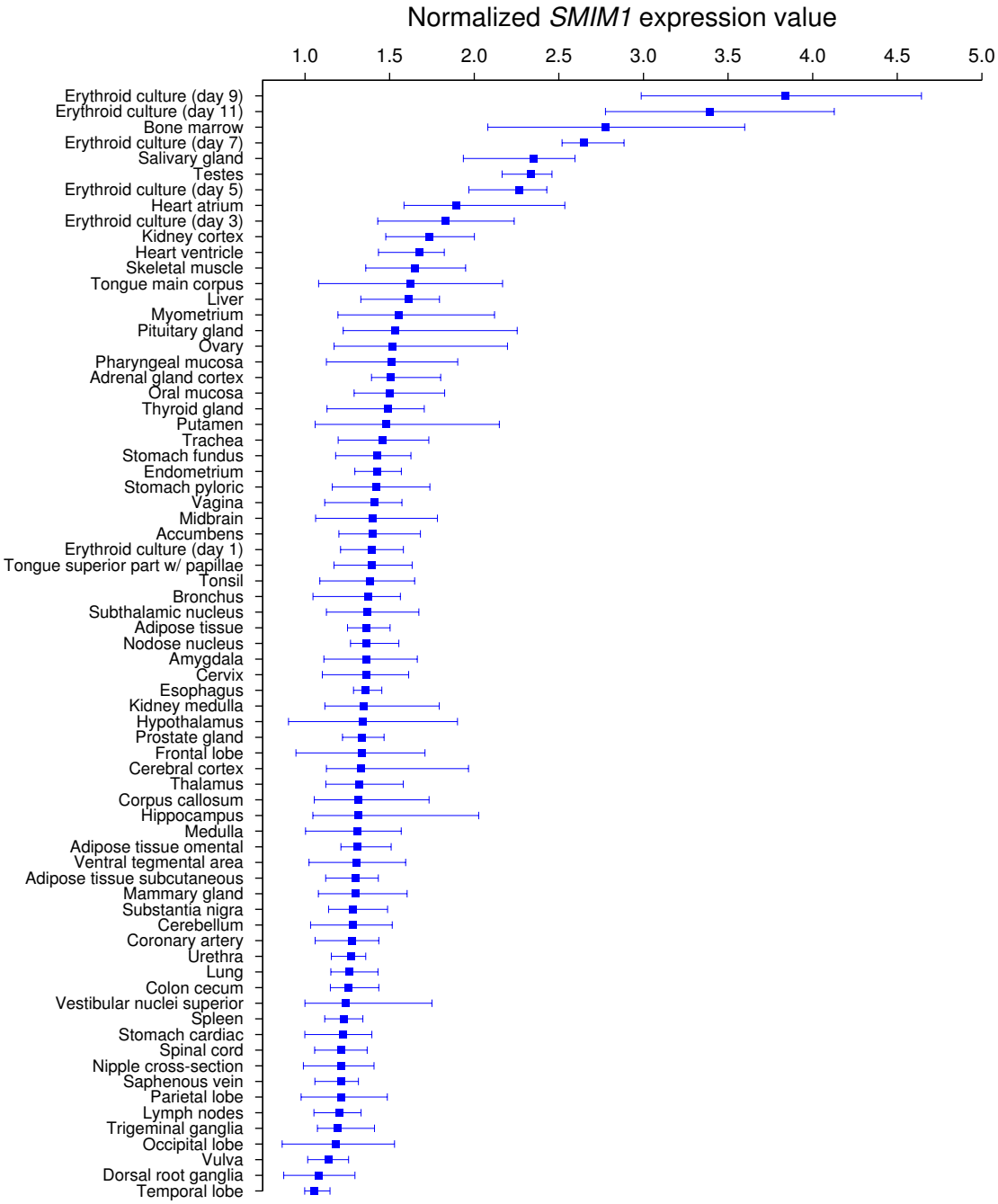


c

```
  1 CAGAGACGCG GGGACACAGG CTCGAGGCGT CTGCCGCACC TCAGCCCACG ACCTGCCCCG CTGGGAGGTG CGGGCCGCTG
 81 GCCAGGCCCT GACCGCAACC TGGCCCAGAG GCCCCAGCCC TCAGGCAAGG TTCTCCGGTG AAGCCACAGC CTGGCCACCT
161 GTCTTGATCT CCCCACCGAG AAGGCCCCGC CCCTCCCGCT GCAGCCCCAC AGCATGCAGC CCCAGGAGAG CCACGTCCAC
241 TATAGTAGGT GGGAGGACGG CAGCAGGGAC GGAGTCAGCC TAGGGGCTGT GTCCAGCACA GAAGAGGCCT CACGCTGCCG
321 CAGGATCTCC CAGAGGCTGT GCACGGGCAA GCTGGGCATC GCCATGAAGG TGCTGGGCGG CGTGGCCCTC TTCTGGATCA
401 TCTTCATCCT GGGCTACCTC ACAGGCTACT ATGTGCACAA GTGCAAATAA ATGCTGCCCC GCATGCACGC GGGGGGCTGG
481 CCGCAAAAAA AAAAAAA
```

(a) 5′-RACE. Top sequence is the *SMIM1*-mRNA reference sequence (gray) plus the 18 bp located directly upstream in the genome (black). At the bottom is the sequence identified using 5′-RACE. Blue boxes: Exons 1 and (start of) 2. Red box: 5′-RACE adaptor added during the experiment only to transcripts with an intact 5′ cap. **(b)** 3′-RACE. Top panel: 3′ end of SMIM1 reference sequence. Bottom panel: Sequence identified by 3′-RACE. Blue boxes: End of exon 4 and adaptor. **(c)** Full sequence of the characterized wildtype SMIM1-mRNA (gray = consensus open reading frame; blue = start codon; red = stop codon).

Supplementary Figure 6: *SMIM1* expression in normal human tissues



This figure illustrates the expression of *SMIM1* across different human tissues, as measured on Affymetrix U133Av2 microarrays (probeset 230840_at). Microarray profiles of normal tissues were generated by Roth *et al.*[1] (data from the NCBI Gene Expression Omnibus, accession no. GSE3526). As positive control, we used microarray profiles of CD34+ hematopoietic progenitors being cultured towards red blood cells (generated by Keller *et al.*[2]; GSE4655; see also **Supplementary Fig. 2**). All data were quantile-normalized per array to a log-normal distribution. Blue indicates average expression values with min–max error bars. Consistent with **Fig. 2** and **Fig. 4**,

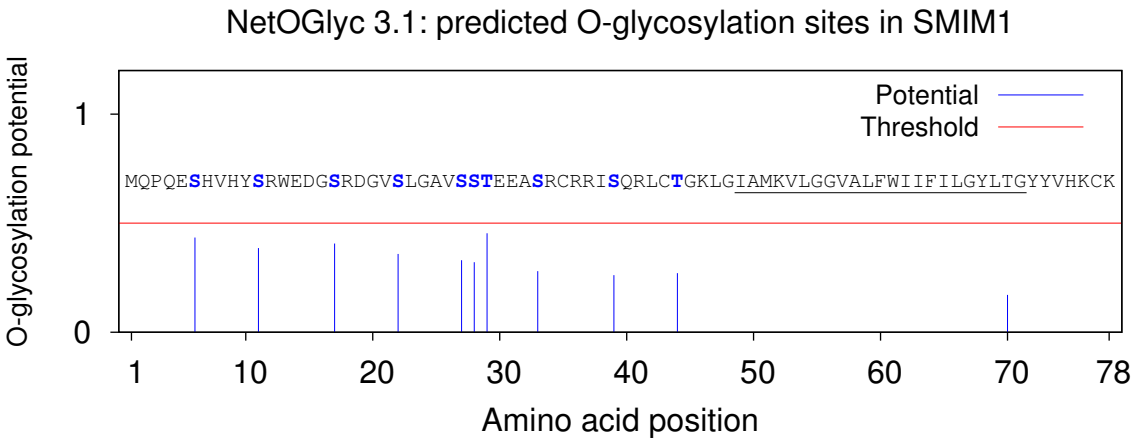
the highest *SMIM1* expression level was observed in bone marrow (where the expression level was on par with that observed in erythroid cultures). Among non-hematopoietic tissues, salivary gland and testes exhibited the highest expression levels.

References

1. Roth, R. *et al.* Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* **7**, 67–80 (2006).

2. Keller, M. *et al.* Transcriptional regulatory network analysis of developing human erythroid progenitors reveals patterns of coregulation and potential transcriptional regulators. *Physiol Genomics* **28**, 114–28 (2006).

Supplementary Figure 7: Predictive analysis of potential O-glycosylation sites in SMIM1



Despite the presence of several extracellular serine and threonine residues (blue), the predictive values for all residues lie under the threshold determined by the prediction tool NetOGlyc 3.1 (`http://www.cbs.dtu.dk/services/NetOGlyc`), suggesting that SMIM1 may lack O-glycans. However, we note that some of the peaks in the above diagram are quite close to the threshold. This may indicate a theoretical possibility that SMIM1 may carry minor O-glycans on certain Ser/Thr residues after all. The human O-glycoproteome has recently been shown to be more complex than previously realized and even the best prediction tools may reveal less than half of all O-glycosylation sites used[1]. Thus, future studies will need to address such potential O-glycosylation.

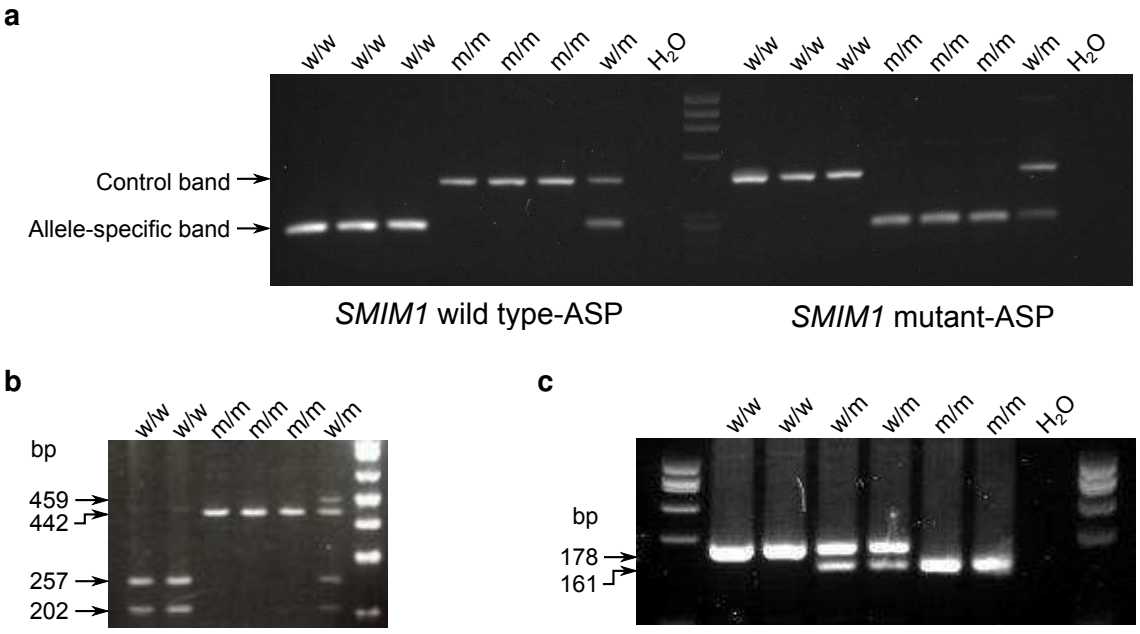
The predicted transmembrane region (**Supplementary Table 5**) is underlined.

References

1. Steentoft, C. *et al.* Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* **8**, 977–982 (2011).

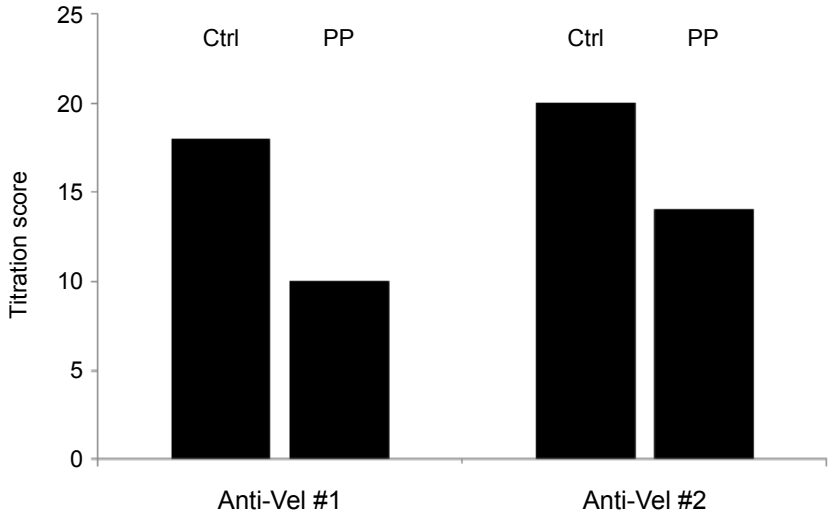
Supplementary Figure 8: Genotyping assays

Assays to determine Vel genotype were designed (primers in **Supplementary Table 7**; used at 10 pM unless stated). Details of PCR assays are available upon request. **(a)** Allele-specific PCR (ASP): gDNA was amplified with primers 388588int2f (1 pM), 388588int3R2 and either 388588wtex3f or 388588mutex3f. An internal control band was provided by the amplification of a 460 bp (wild type) or 443 bp (mutant) product. Allele-specific bands of 266 bp (wild type) or 249 bp (mutant) were specifically amplified by their respective primers. **(b)** PCR-RFLP: gDNA was amplified with primers 388588int2f and 388588int3R2. The products were digested with *Sty*I at 37°C for two hours and analyzed by agarose-gel electrophoresis. The restriction site was abolished in the mutated sequence. **(c)** Gene-specific PCR: gDNA was amplified with primers LOCex3fscreen and LOCex3rscreen which flanked SMIM1 exon 3. Allele-specific PCR products of 178 bp (wild type) or 161 bp (mutant) were discriminated based on size by agarose-gel electrophoresis.



Supplementary Figure 9: Inhibition of anti-Vel

Inhibition tests were performed using synthetic SMIM1 peptides 1–4 with two examples of human anti-Vel. Doubling dilutions of the antibodies were prepared in 1 % BSA/PBS, and 25 μL volumes of diluted plasma and peptide solution (0.1 mg/mL) were mixed and allowed to incubate for 45 minutes in a microplate at room temperature. The mixtures were tested by hemagglutination with Vel+ RBCs and the reactivity determined as a titration score[1]. While tests with individual peptide solutions did not show an effect, a pool of the four peptides demonstrated partial inhibition. Ctrl = 1 % BSA/PBS dilution control; PP = peptide pool.



References

1. Marsh, W. L. Scoring of hemagglutination reactions. *Transfusion* **12**, 352–353 (1972).