haematologica
Journal of the Ferrata Storti Foundation

The insertion/deletion polymorphism rs201494641 at *ITGA9* influences blood CD34^{+} cell levels by altering ZNF384 binding

by Caterina Cafaro, Zain Ali, Antton Lamarca, Daniela Torres Di Bello, Laura Duran Lozano, Ludvig Ekdahl, Malte Thodberg, Maroulio Pertesi, Aitzkoa Lopez de Lapuente Portilla, and Björn Nilsson

Received: February 29, 2024.
Accepted: April 24, 2024.

Citation: Caterina Cafaro, Zain Ali, Antton Lamarca, Daniela Torres Di Bello, Laura Duran Lozano, Ludvig Ekdahl, Malte Thodberg, Maroulio Pertesi, Aitzkoa Lopez de Lapuente Portilla, and Björn Nilsson. The insertion/deletion polymorphism rs201494641 at ITGA9 influences blood CD34^{+} cell levels by altering ZNF384 binding.
Haematologica. 2024 May 9. doi: 10.3324/haematol.2024.285363 [Epub ahead of print]

LETTER TO THE EDITOR

The insertion/deletion polymorphism rs201494641 at *ITGA9* influences blood CD34^{+} cell levels by altering ZNF384 binding

Caterina Cafaro[1,2], Zain Ali[1,2], Antton Lamarca[1,2], Daniela Torres Di Bello[1,2], Laura Duran Lozano[1,2], Ludvig Ekdahl[1,2], Malte Thodberg[1,2], Maroulio Pertesi[1,2], Aitzkoa Lopez de Lapuente Portilla[1,2], Björn Nilsson[1,2,3]

[1]Division of Hematology and Transfusion Medicine, Department of Laboratory Medicine, BMC B13, Lund University, SE-221 84, Lund, Sweden. [2]Lund Stem Cell Center, Lund University, SE-221 84 Lund, Sweden. [3]Broad Institute, 415 Main Street, Cambridge, MA 02142, USA.

Correspondence to:

Björn Nilsson (bjorn.nilsson@med.lu.se)

Keywords: Hematopoietic stem and progenitor cells, stem cell transplantation, genome-wide association study, integrin alpha-9

FUNDING

This work was supported by grants from the European Research Council (CoG-770992), the Swedish Research Council (2017-02023 and 2018-00424), the Swedish Cancer Society (23 2851 Pj), the Swedish Children's Cancer Fund (PR2023-0067 and PR2020-0056), Inga-Britt and Arne Lundberg's Foundation (2017-0055).

AUTHOR CONTRIBUTIONS

C.C. and B.N. designed the experiments. C.C., D.T.D.B., L.D.L, and A.L.D.L.P. carried out experiments. C.C., Z.A., A.L., L.E., M.T., M.P. and B.N. carried out bioinformatic analyses. C.C. and B.N. drafted the manuscript. All authors contributed to the final manuscript.

CONFLICTS-OF-INTEREST STATEMENT

The authors have no conflicts-of-interest to declare.

DATA SHARING STATEMENT

ATAC-seq raw data deposited in Sequence Read Archive, accession number PRJNA1040035.

Stem cell transplantation is a cornerstone in the treatment of blood malignancies. The most common method for harvesting stem cells for transplantation is by leukapheresis. This requires the mobilization of $CD34^+$ hematopoietic stem and progenitor cells (HSPCs) from the bone marrow into the blood. Understanding the genetic factors that influence blood $CD34^+$ cell levels could reveal previously unappreciated genes and mechanisms that control HSPC behavior in humans, as well as potential new drug targets for HSPC mobilization.

Recently, we reported the first large-scale genome-wide association study on blood $CD34^+$ cell levels[1]. Across 13,167 individuals, we identified 11 independent genetic associations with blood $CD34^+$ cell levels. One of the most significant associations maps to the *ITGA9* locus at chromosome 3p22 (lead variant rs201494641:TTT>T; $P = 4.7\times10^{-11}$; $\beta = -0.123$). The ITGA9 protein is the α subunit of the α9β1 integrin receptor, which has been reported to modulate HSPC growth, differentiation, and retention within the bone marrow by interacting with various ligands, like fibronectin[2], tenascin-C[3], and osteopontin[4]. The 3p22 association with blood $CD34^+$ cell levels is represented by 46 variants in high linkage disequilibrium (LD) ($r^2 > 0.8$) with the lead variant. All of these variants are located in *ITGA9* intron 3 and 4 (**Fig. 1a**). However, the causal variants and their mechanisms of action remain unknown. We therefore sought to dissect functionally the association at 3p22 between *ITGA9* and blood $CD34^+$ cell levels.

To identify causal variants, we integrated ATAC-sequencing and mRNA-sequencing data for sorted blood cell types[5,6]. We found a strong positive correlation between *ITGA9* expression and chromatin accessibility in an approximately 1,300 bp-long segment in *ITGA9* intron 3 (**Fig. 1a**)[1]. Notably, this segment is selectively accessible in $CD34^+$ blood cell populations, including hematopoietic stem cells (HSC), multi-potent progenitors (MPP), common myeloid progenitors (CMP), and megakaryocyte-erythroid progenitors (MEP; **Fig. 1b**). The segment encompasses

four variants within the *ITGA9* LD block: three single nucleotide polymorphisms (rs73053290, rs17227369 and rs17227404) and one insertion-deletion polymorphism (rs201494641; **Fig. 1b**). Using promoter capture Hi-C data for primary $CD34^+$ cells[7], we detected a chromatin looping interaction between these four variants and the *ITGA9* promoter (**Fig. 1a**)[1]. Consistent with a gene-regulatory effect, analysis of mRNA-sequencing data for $CD34^+$ cells from 155 blood donors showed association between rs17227369 and *ITGA9* mRNA levels in blood $CD34^+$ cells, with the minor allele yielding lower expression (linear regression $P = 2.0\times10^{-11}$; **Fig. 1c**)[1]. To investigate if this effect translates to the protein level, we quantified ITGA9 surface expression on $CD34^+$ cells in 458 blood donors by flow cytometry, observing a significant association in the same direction (linear regression $P = 8.1\times10^{-15}$; **Fig. 1d** and **Supplementary Fig. 1a-b**).

To confirm the regulatory role of the identified segment on *ITGA9* expression, we used dual-sgRNA CRISPR/Cas9 genome editing[8] to delete a 486-bp region harboring the four candidate variants (**Supplementary Table 1**) in the human erythroleukemia HEL cells, which show an HSPC-like transcriptional profile and are homozygous for the major alleles of the four variants of interest[9]. This led to the downregulation of *ITGA9*, further supporting a regulatory role (**Fig. 2a**). To assess the transcriptional activity of each of the four candidate variants, we carried out luciferase experiments with constructs representing their reference and alternative alleles in HEL cells (**Supplementary Table 2).** We observed higher transcriptional activity with rs201494641-TTT construct than with rs201494641-T construct (one-sided Student's t-test $P = 2.8\text{x}10^{-3}$; **Fig. 2b**), consistent with the direction of the effects on *ITGA9* transcript and protein levels (**Fig. 1c-d**). Similarly, we detected allele-dependent accessibility at rs201494641 (14 *vs* seven reads containing the TTT and T alleles, respectively; Binomial test $P = 2\text{x}10^{-2}$) but not at the other three variants in ATAC-seq data for the acute myeloid leukemia cell line MUTZ-3,

which is heterozygous for all four variants-of-interest. We also noted a DNAase I footprint at rs201494641 in $CD34^+$ cells (**Fig. 2c**)[10]. Collectively, these data identify rs201494641 as a likely causal regulatory variant underlying the 3p22 association with blood $CD34^+$ cell levels.

Further, we searched for differentially binding transcription factors using the FABIAN tool[11]. The strongest differential binding score was seen for the zinc finger protein ZNF384, which binds the rs201494641-harboring region **(Fig. 2c)**[12]. FABIAN predicted higher binding affinity for the minor (T) compared to the major (TTT) allele (**Fig. 2d**). The ZNF384 core binding motif is a poly-A/poly-T sequence (**Fig. 2e**)[13], whose length is affected by rs201494641. siRNA-mediated knockdown of ZNF384 yielded upregulation of *ITGA9* in HEL cells (one-sided Student's t-test $P = 1.0 \times 10^{-2}$; **Fig. 2f**). Additionally, we observed reduced DNAase I accessibility across the repetitive poly-A/poly-T sequence as well as the flanking regions (**Fig. 2g**). Collectively, these observations are consistent with ZNF384 acting as a transcriptional repressor, preferentially binding the minor *ITGA9*-low-expressing allele rs201494641-T.

In conclusion, we functionally dissected the genetic association between *ITGA9* and blood $CD34^+$ cell levels. We show that the association maps to a regulatory region in *ITGA9* intron 3, and identify rs201494641:TTT>T as a likely causal variant. Our data are consistent with rs201494641:TTT>T increasing the affinity of the zinc finger protein ZNF384, which represses *ITGA9* transcription. Previously, *ZNF384* has been reported to undergo somatic rearrangements in B-cell precursor acute lymphoblastic leukemia, including gene fusions with more than ten distinct partner genes, including *TCF3*, *EP300*, *TAF15*, and *CREBBP*[14]. However, its precise role in hematopoiesis remains unexplored. In summary, our findings provide new insight into the genetic factors that influence blood $CD34^+$ cell levels and implicate ITGA9 as a regulator of circulating HSPC levels in humans.

References

1. Lopez de Lapuente Portilla A, Ekdahl L, Cafaro C, et al. Genome-wide association study on 13167 individuals identifies regulators of blood CD34+cell levels. Blood. 2022;139(11):1659-1669.
2. Wirth F, Lubosch A, Hamelmann S, Nakchbandi IA. Fibronectin and Its Receptors in Hematopoiesis. Cells. 2020;9(12):2717.
3. Nakamura-Ishizu A, Okuno Y, Omatsu Y, et al. Extracellular matrix protein tenascin-C is required in the bone marrow microenvironment primed for hematopoietic regeneration. Blood. 2012;119(23):5429-5437.
4. Grassinger J, Haylock DN, Storan MJ, et al. Thrombin-cleaved osteopontin regulates hemopoietic stem and progenitor cell functions through interactions with α9β1and α4β1 integrins. Blood. 2009;114(1):49-59.
5. Ulirsch JC, Lareau CA, Bao EL, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat Genet. 2019;51(4):683-693.
6. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016;48(10):1193-1203.
7. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015;47(6):598-606.
8. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281-2308.
9. Ghandi M, Huang FW, Jané-Valbuena J, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503-508.

10. Vierstra J, Lazar J, Sandstrom R, et al. Global reference mapping of human transcription factor footprints. Nature. 2020;583(7818):729-736.
11. Steinhaus R, Robinson PN, Seelow D. FABIAN-variant: predicting the effects of DNA variants on transcription factor binding. Nucleic Acids Res. 2022;50(W1):W322-W329.
12. Hammal F, De Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: A database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res. 2022;50(D1):D316-D325.
13. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022;50(D1):D165-D173.
14. Hirabayashi S, Butler ER, Ohki K, et al. Clinical characteristics and outcomes of B-ALL with ZNF384 rearrangements: a retrospective analysis by the Ponte di Legno Childhood ALL Working Group. Leukemia. 2021;35(11):3272-3277.
15. Ulirsch JC, Nandakumar SK, Wang L, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. Cell. 2016;165(6):1530-1545.

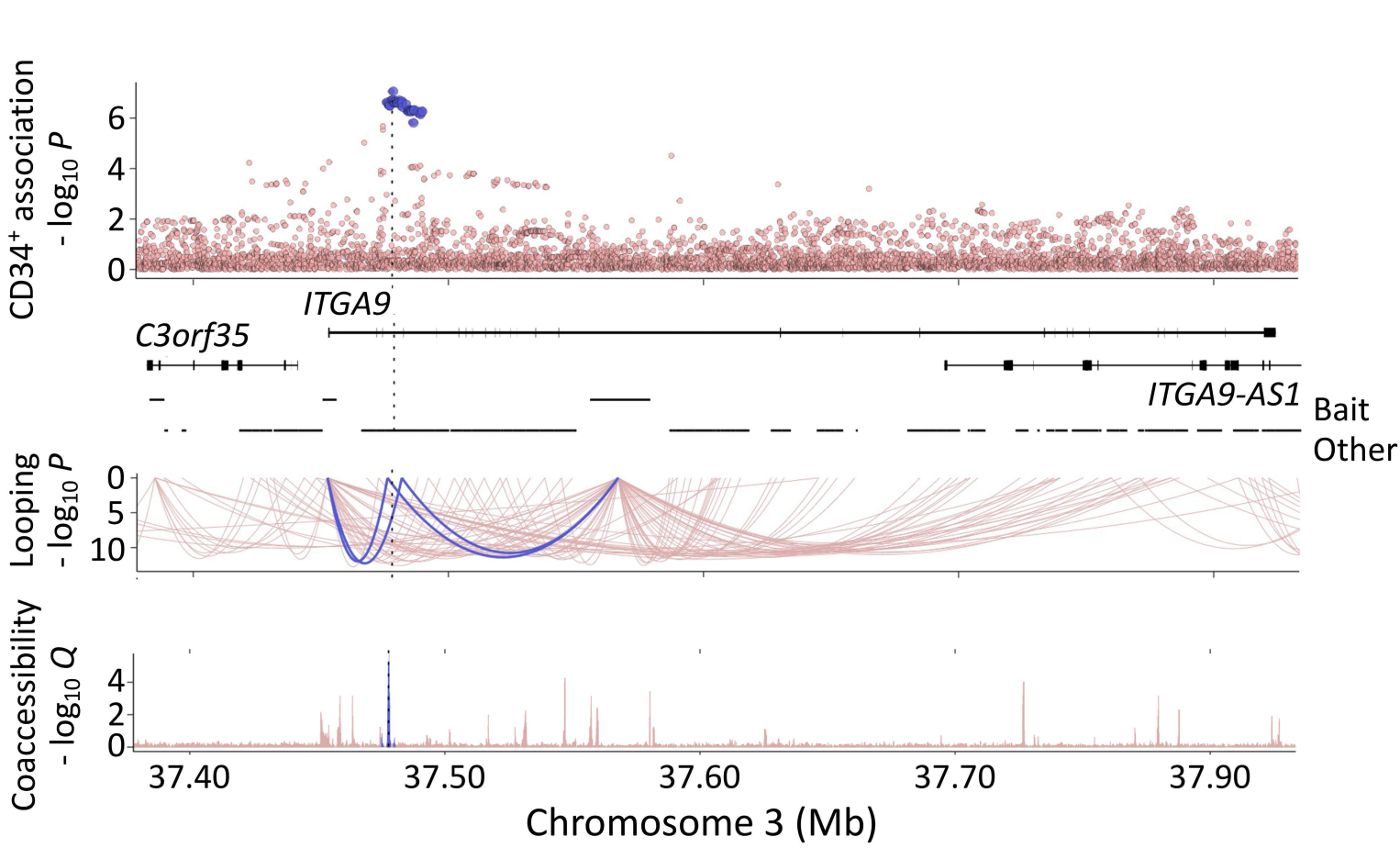
FIGURE LEGENDS

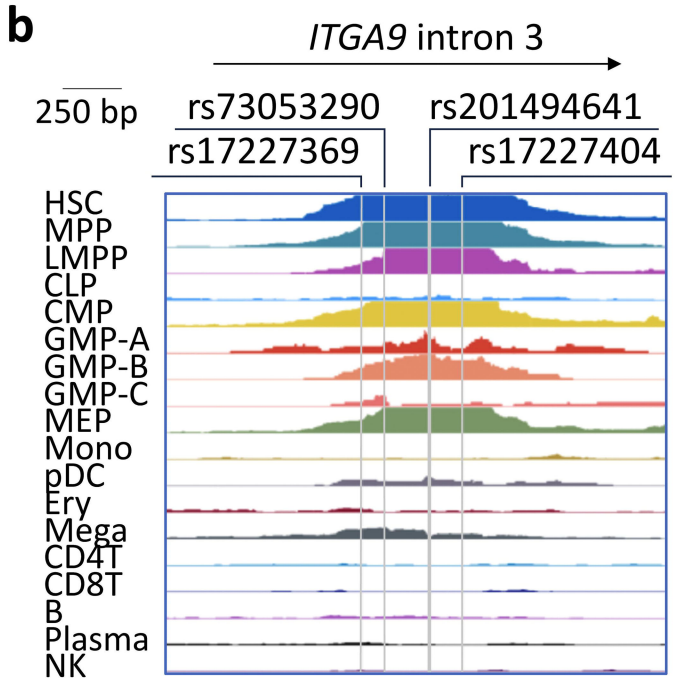
Figure 1: Association of the 3p22 genetic locus with *ITGA9*[1]. (a) *Top*: close-up of the 3p22 signal. *Middle*: chromatin looping interactions in $CD34^+$ cells with standard and internal promoter (blue arches; y-axis indicates PCHi-C P-score[7]). *Bottom*: coaccessibility plot showing positive correlation (blue peak) between *ITGA9* expression across sorted blood cell populations and ATAC-seq signal (100-bp sliding window) (*y*-axis indicates false discovery rate for Pearson correlation)[15]. **(b)** Four credible set variants map to the identified segment, which is accessible in HSCs, MPPs, LMPPs, CMPs, GMPs and MEPs (*y*-axis indicates ATAC-seq signal). **(c)** Correlation between rs17227369 genotype (credible set proxy for rs201494641 used for TaqMan genotyping) and *ITGA9* mRNA levels (residual FPKM)[1]. **(d)** Correlation between rs17227369 genotype and ITGA9 protein levels (median fluorescent intensity (MFI), quantified using phycoerythrin (PE)-conjugated monoclonal antibody towards ITGA9. Statistics are for Pearson correlation, *** $P \leq 0.001$.

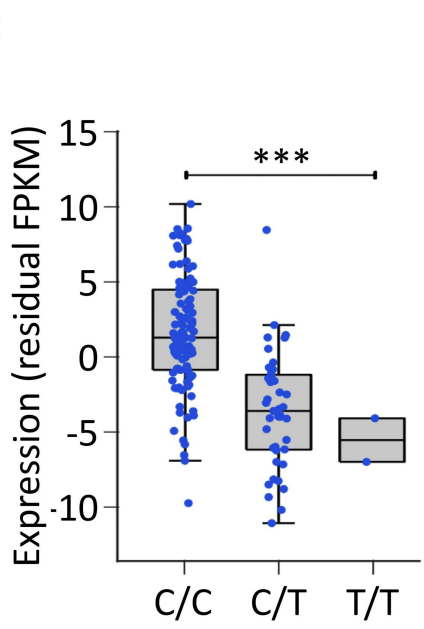
Figure 2: ZNF384 preferential binding to rs201494641-T repressing *ITGA9*. (a) *ITGA9* expression in HEL cells subjected to CRISPR/Cas9 editing with a non-targeting sgRNA pair control (Ctrl, 60-bp cut, eleven biological replicates), or an sgRNA pair designed to delete the region harboring the 4 putative causal variants at 3p22 (CRISPR, 486-bp cut) (**Supplementary Table 2**). a.u., arbitrary units; mRNA, messenger RNA; one-sided *t*-test, *** $P \leq 0.001$. **(b)** Luciferase activities of the four candidate causal variants in HEL cells (four biological replicates). Data normalized to empty vector control; one-sided *t*-test, *** $P \leq 0.001$, * $P \leq 0.05$, n.s. not significant. **(c)** *upper panel:* chromatin accessibility (ATAC-sequencing signal intensity)

across different blood cell types (colors as in **Fig. 1b**) in the approximately 1,300-bp wide region in *ITGA9* intron 3; *lower panel:* DNAse I footprint in primary $CD34^+$ cells and CHIP-seq signals in K562 cells from ReMap in the rs201494641-harboring region. **(d)** ZNF384 gain of binding to rs201494641-T from the Fabian-variant database. **(e)** JASPAR detailed Transcription Factor Flexible Model TFFM0157.1 showing the ZNF384 consensus motif. **(f)** *ZNF384* and *ITGA9* expression in HEL cells transfected with Ctrl or ZNF384 siRNA (four biological replicates). a.u., arbitrary units; mRNA, messenger RNA; one-sided *t*-test, *** $P \leq 0.001$, * $P \leq 0.05$. **(g)** ENCODE DNAse I consensus footprint data for the ZNF384 motif in primary $CD34^+$ cells, showing mean cleavage ratios in a 200-bp window centered on the core poly-T motif (dotted lines), for two different repeat lengths. In both cases the DNA binding footprint (solid lines) extends beyond the core poly-T stretch, suggesting that ZNF384 directly binds flanking sequences on either side of the repeat sequence.

Figure 1


a
CD34+ association
- log10 P
6
4
2
0
ITGA9
C3orf35
ITGA9-AS1
Bait
Other
Looping
- log10 P
0
5
10
Coaccessibility
- log10 Q
4
2
0
37.40
37.50
37.60
37.70
37.90
Chromosome 3 (Mb)


b
ITGA9 intron 3
250 bp
rs73053290
rs201494641
rs17227369
rs17227404
HSC
MPP
LMPP
CLP
CMP
GMP-A
GMP-B
GMP-C
MEP
Mono
pDC
Ery
Mega
CD4T
CD8T
B
Plasma
NK


c
Expression (residual FPKM)
15
10
5
0
-5
-10
***
C/C
C/T
T/T

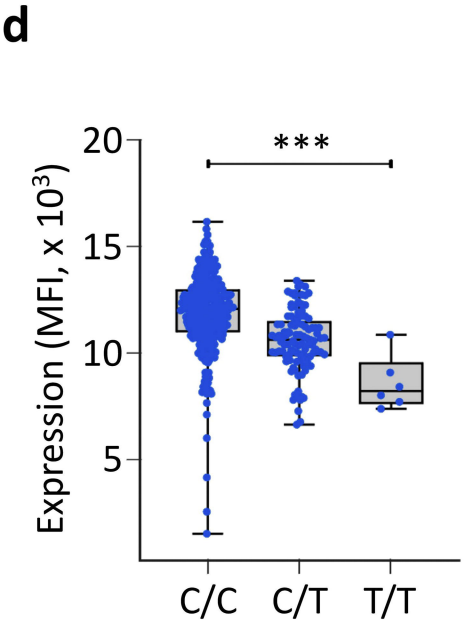
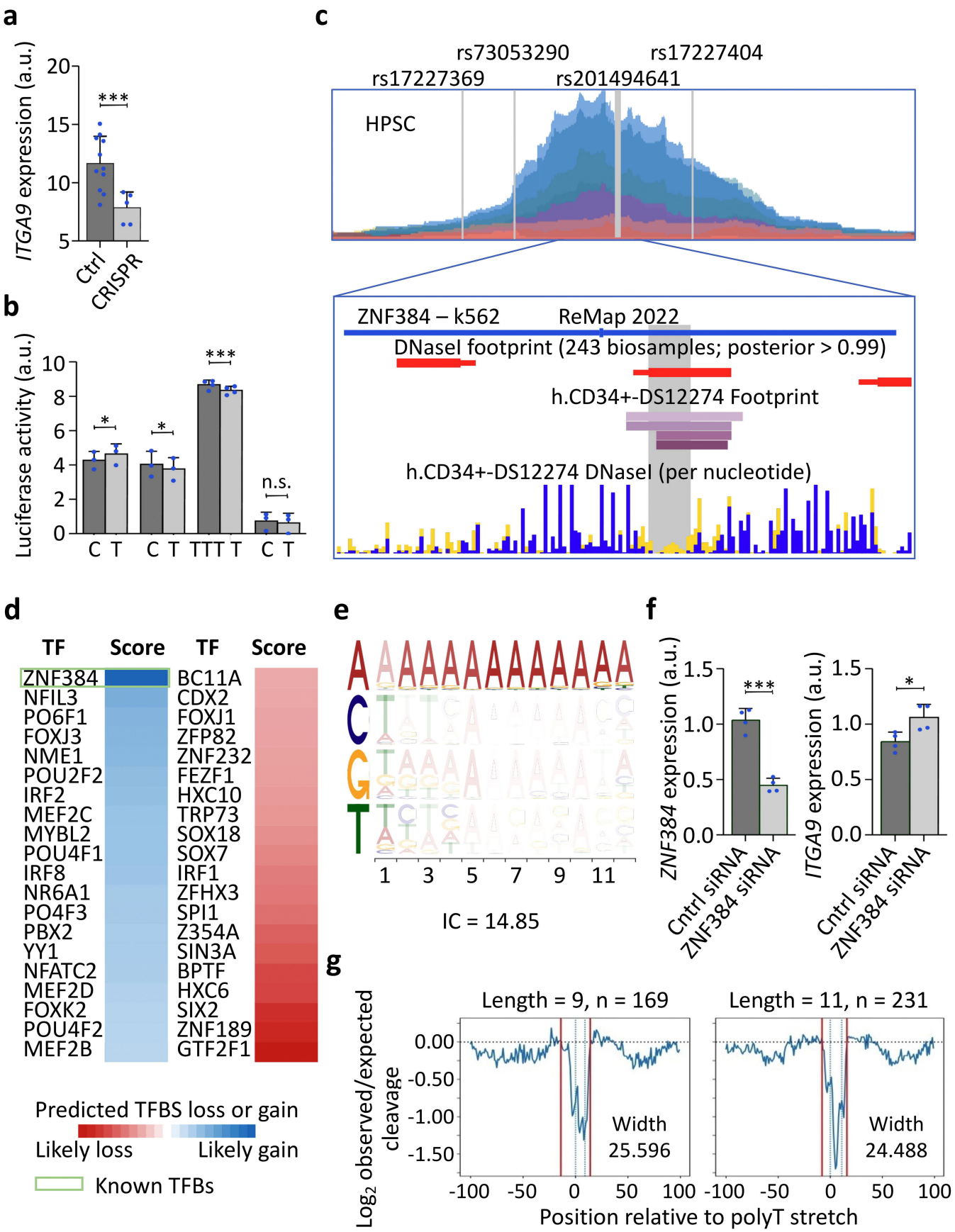
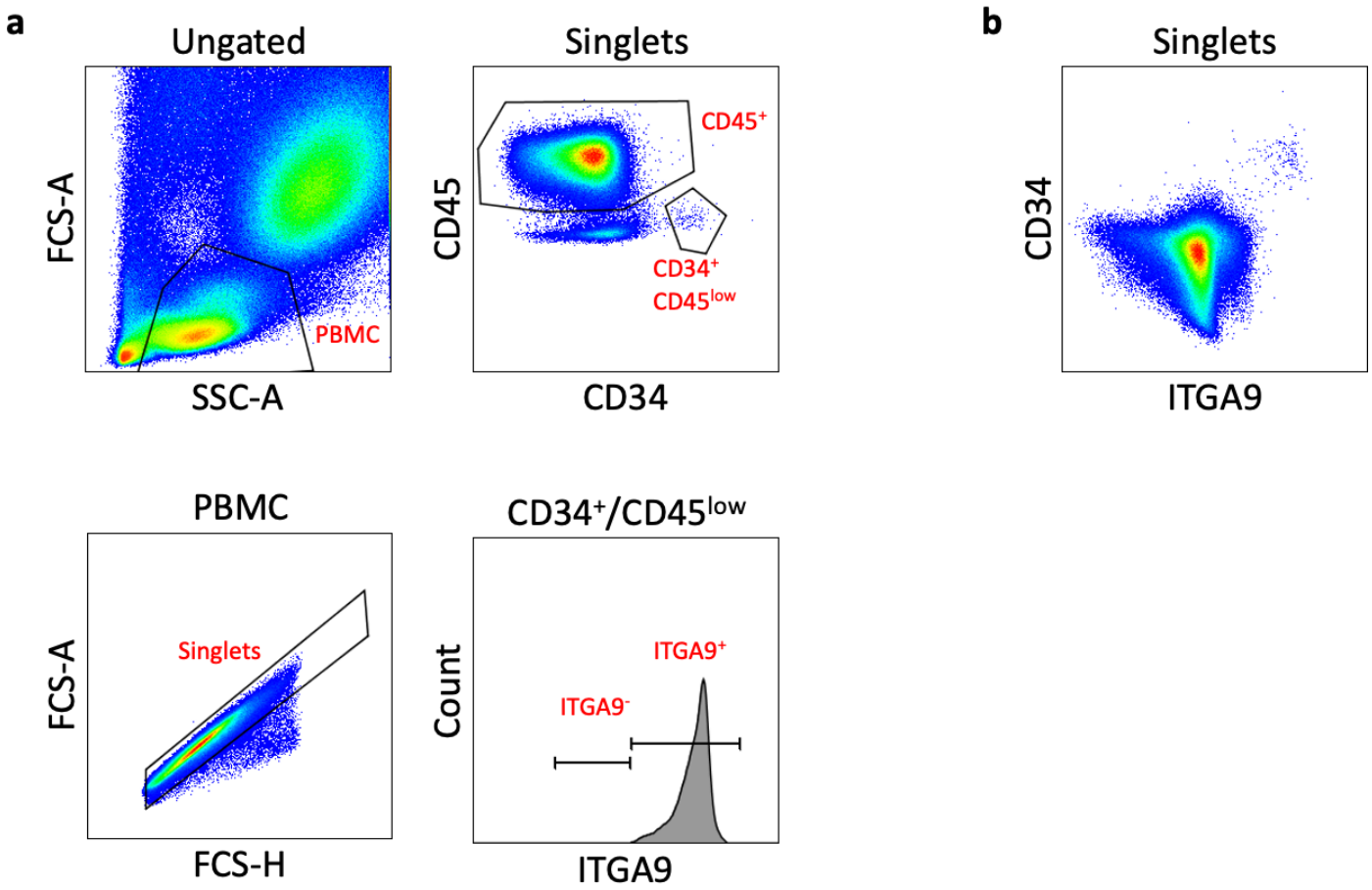

d
Expression (MFI, x 10³)
20
15
10
5
***
C/C
C/T
T/T

Figure 2


a
ITGA9 expression (a.u.)
20
15
10
5
***
Ctrl
CRISPR
b
Luciferase activity (a.u.)
10
8
6
4
2
0
*
*
***
n.s.
C T C T TTT T C T
c
rs73053290
rs17227404
rs17227369
rs201494641
HPSC
ZNF384 – k562
ReMap 2022
DNaseI footprint (243 biosamples; posterior > 0.99)
h.CD34+-DS12274 Footprint
h.CD34+-DS12274 DNaseI (per nucleotide)
d
TF
Score
TF
Score
ZNF384
NFIL3
PO6F1
FOXJ3
NME1
POU2F2
IRF2
MEF2C
MYBL2
POU4F1
IRF8
NR6A1
PO4F3
PBX2
YY1
NFATC2
MEF2D
FOXK2
POU4F2
MEF2B
BC11A
CDX2
FOXJ1
ZFP82
ZNF232
FEZF1
HXC10
TRP73
SOX18
SOX7
IRF1
ZFHX3
SPI1
Z354A
SIN3A
BPTF
HXC6
SIX2
ZNF189
GTF2F1
Predicted TFBS loss or gain
Likely loss
Likely gain
Known TFBs
e
1 3 5 7 9 11
IC = 14.85
f
ZNF384 expression (a.u.)
1.5
1.0
0.5
0.0
***
Cntrl siRNA
ZNF384 siRNA
ITGA9 expression (a.u.)
1.5
1.0
0.5
0.0
*
Cntrl siRNA
ZNF384 siRNA
g
Log2 observed/expected cleavage
Length = 9, n = 169
Length = 11, n = 231
0.00
-0.50
-1.00
-1.50
Width
25.596
Width
24.488
-100 -50 0 50 100
-100 -50 0 50 100
Position relative to polyT stretch

Supplementary Figure 1



(a) Gating procedure. **Top left:** peripheral blood mononuclear cells (PBMCs) were gated based on forward scatter area (FSC-A) and side scatter area (SSC-A) from 1 million cells. **Bottom left:** from PBMCs, singlets cells were gated based on forward scatter area (FSC-A) and forward scatter height (FSC-H). **Top right:** from singlets, $CD34^{+}45^{low}$ cells and $CD45^{+}$ cells were gated from the PBMCs cluster. **Bottom right:** $ITGA9^{+}$ cells were gated from the $CD34^{+}45^{low}$ cluster. **(b)** Co-expression of ITGA9 and CD34 in PBMCs (*i.e.*, cells that express ITGA9 also express CD34 and vice versa).

Supplementary Table 1

sgRNA and primer sequences used in CRISPR/Cas9 experiments.

sgRNA(s) (+ strand)		PAM		Strand		Primers for deletion confirmation	Deletion size (bp)
5'	3'	5'	3'	5'	3'	Forward/reverse	
sgRNA pair for deleting *ITGA9* whole region (chr3:37,477,598; "V1-V4")							
CACC**GATGAGAGAGTGAGAAGGGA**	CACC**GGTGTGAAAATGTTATTCTG**	AGG	AGG	-	-	CATGGACACTGTTGCACCTC	486
chr3:37,477,595-37,477,614	chr3:37,478,075-37,478,097					GACTCCCCTAAGAAGCTCCC	
sgRNA pair for deleting random control region							
CACCG**AAATTCCCGAAAAGCCTGGT**	CACCG**AAAGCACAGGCAGAAACTGT**	AGG	AGG	+	+	CCTGTTGCTGTGGCTACTGA	60
chr17:17,024,714-17,024,736	chr17:17024773-17024795					GCCAGGCTTGTTGGGAAGTA	

Supplementary Table 2

Reporter construct sequences used in luciferase experiments.

Variant	Sequence
rs73053290	ATAGGTACCTTCCCCAGAAACAAATGATACTGGATGTAAAGATAGCATCAGGATTAGAGGCAAGCCCCTTCC**[C/T]**TTCTCACTCTCTCATCCTGAGCTTCCTGTG TGCAGACGACATTTTAAGAAATGAGCGGGGAGGAGATCTATA
rs17227369	ATAGGTACCAAGAAATGAGCGGGGAGGGGAATCACTTCTTGAAATAACCACTGAACAAACTCTGTATAGGCA**[C/T]**TGACTGCTGCTCAGACTTTGGAGGACAAC TCAAAATGTTCAAGTCTAATGTCGCCCCTTTTCAAGATCTATA
rs201494641	ATAGGTACCTCTTTGTCATTTCCTTTCTAATTATCAGGGCTGTCCTTATCTGTTCAGTGCTTCCCATCTGTCCA**[TTTTTTTTTTT/TTTTTTTTT]**CAAAGCCACAGGG TACATTTTTGCATTTACTTCTGTTGCTCTGTTTCCTGATTTTTTTTAAAGTGAGATCTATA
rs17227404	ATAGGTACCACACACATGCAGAAAAAGGTATAAATCATCAATGTATACCTCAATAAATTTTCATTAGGTGAA**[C/T]**GTCCTCAGAATAACATTTTCACACCCCAGCA GCCCCTTCCCAATGCCCCTCCTGGCCACTTCCAGATCTATA

Major/minor allele in red within brackets