Identification of the transcription factor MAZ as a regulator of erythropoiesis

Darya Deen[1], Falk Butter[2], Deborah E. Daniels[3], Ivan Ferrer-Vicens[3], Daniel C. J. Ferguson[3], Michelle L. Holland[4], Vasiliki Samara[5], Jacqueline A. Sloane-Stanley[5], Helena Ayyub[5], Matthias Mann[6], Jan Frayne[3], David Garrick[5,7*] and Douglas Vernimmen[1*]

[1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, United Kingdom.

[2] Institute of Molecular Biology (IMB), 55128 Mainz, Germany

[3] School of Biochemistry, University of Bristol, Bristol, UK.

[4] Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London SE1 9RT, UK

[5] MRC Molecular Haematology Unit, Weatherall Institute for Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, United Kingdom.

[6] Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany.

[7] Current address : INSERM U976 Équipe 5, Institut de Recherche Saint Louis, Université de Paris, 75010 Paris, France.

[*] These authors contributed equally

Correspondence:
david.garrick@inserm.fr, douglas.vernimmen@roslin.ed.ac.uk

Supplemental Methods

ChIP-Seq analysis

Alignment. Reads were aligned using the hisat2 aligner [1] version 2.0.3 with the --no-spliced-alignment option, but otherwise default parameters, to a splicing-unaware index of the human GRCh38 genome.

The strength of enrichment in the IP was assessed plotting "fingerprints" using deeptools [2] and by calculating normalized strand cross-correlation coefficient (NSC) and relative strand cross-correlation coefficient (RSC) metrics using Phantompeakqualtools [3].

Normalisation of the ChIP-Seq signal. Reads of the ChIP-Seq samples were first normalised to the input and then scaled to 1x sequencing depth using deepTools v. 3.1.3 [4].

Peak calling. Peak calling was performed using MACS2 2.1.1 with the minimum FDR (q-value) cutoff of 0.01 [5]. The top 75% fold enriched peaks were selected for further analysis. For ENCODE datasets optimal IDR thresholded peaks provided by the ENCODE consortium have been used.

Peak annotation was made with UCSC RefSeq gene annotation (GRCh38 genome version) using HOMER suite 4.8 [6]. For these analyses we define promoters as regions located within a distance of -3000 bp upstream and 100 bp downstream of the TSS.

Coverage analysis and ChIP heatmap plots and profiles were performed using DeepTools suites. General genome arithmetics was performed using BEDTools v. 2.27.1 [7]. The set of erythroid-specific and housekeeping genes and erythroid enhancers was based on [8].

Correlation of the datasets. A clustered heatmap of correlation coefficients of bigwig signal was computed using the Pearson method; the bam signal combined from the replicates was plotted over 10 kb bins with DeepTools.

Defining erythroid-specific MAZ signal. The peaks called on MAZ datasets were overlapped with the peaks from five MAZ ENCODE datasets with the minimum overlap

of 1 bp using BEDTools. The MAZ peaks without the overlap in any of the regions were assigned to “erythroid-specific peakset”.

Motif Analysis. To find the sequence motifs enriched in MAZ peaks, genomic sequences (from -50 bp to +50 bp around the centres of the top 500 MAZ peaks ranked on their q-values) were extracted (hg38) and used as input for MEME *de novo* motif discovery [9] with *E*-value cutoff 0.01 and motif size defined as 6-30 bp.

To calculate the presence of the motifs in a given peak, we used FIMO with the above defined motifs at a *p*-value threshold of 10^{-4} [10]. Motif central enrichment analysis was performed using CentriMo [11] on regions from -250bp to +250 bp relative to the peak summits. The motif comparison was carried out using TomTom against JASPAR CORE 2018 database[12].

Predicted PWMs were calculated as previously described [13] by using a linear support vector machine method to infer statistical pairwise contact energies between amino acids in the ZF domains of MAZ and SP1 and the corresponding DNA nucleotides.

Functional analysis of genes with erythroid-specific MAZ enrichment. g:Profiler [14] was employed to conduct Gene Ontology (GO) Biological Function and HP (Human Phenotype) gene annotations. Fisher's exact test was used to retrieve significantly enriched GO terms for genes marked with erythroid-specific MAZ signal. Functional categories are defined as those containing at least five genes and a minimum enrichment score of 1.3 (*p*-value < 0.05). GeneATLAS database [15] was used for mining MAZ variants associated with erythroid-specific traits (p-value < 10^{-2}); the p-values were corrected for multiple comparison using Benjamini-Hochberg method [16].

Data sources, protocols and analysis.

Sources of ChIP-seq data are shown in Suppl. Table 3.

ADDITIONAL MATERIAL

Supplemental Fig 1. Chromatin accessible regions around the human HBA locus in erythroid cells. (A) Structure of the α-globin locus. Chromosomal position (hg38 genome build) is shown above the genes. The locus consists of the embryonic ζ gene (HBZ), the duplicated foetal/adult α genes (HBA2 and HBA1) together with two flanking pseudogenes (HBM and HBQ1). The upstream, widely-expressed gene, NPRL3 is transcribed from the opposite strand to that of the HBA genes and contains the remote regulatory elements of the α-globin locus (MCS -R1 to -R3) (indicated by red dots). Below is shown the ATAC-seq signal in human erythroblasts (taken from [17]) indicating chromatin accessible regions at the promoter and distal regulatory elements (shaded regions). **(B)** Fine-mapping of the promoter region of the HBA genes. Nuclei from EBV and K562 were exposed to the restriction enzymes *Bfa*I, *BsrF*I, *Fok*I, *Hinf*I, *Eag*I, *Nco*I, *Mse*I and *Hind*III as indicated. DNA was subsequently purified, limit-digested with *Pst*I and analysed by Southern blot using the 3' *Hind*III/*Pst*I segment of the HBA1 gene as a probe (blue rectangle). This blot does not distinguish between the HBA2 and HBA1 genes.

Supplemental Figure 2. Comparison of probes 11/12 and 13/14. (A) Aligned sequences of probes 11/12 and 13/14. The predicted core MAZ-binding motif is highlighted. **(B)** Gel-shift assays showing that both probes bind the same species. The protein responsible for species (d) has higher affinity for the 13/14 probe as demonstrated by using unlabelled oligonucleotides as competitor (100-fold molar excess over radiolabelled probe). Note that irrelevant lanes have been removed between lanes 1 and 2.

Supplemental Figure 3. Characterisation of DNA binding of unknown protein by EMSA. (A) Increased concentration of EDTA abolished the protein/DNA complex for bands (a), (b), (c), and (d) with the probe 13/14, but not for the probe 7/8, which is bound by a non-Zinc finger protein, NFY (see Figure 1B). **(B)** A single nucleotide mutation (G->T) abolishes binding of the (d) complex to probe 13/14. The sequence of the wild type and mutant (M3) probes are shown above, with the mutated nucleotide indicated in red. EMSA were carried out using K562 nuclear extracts and either wild type or mutant probes together with unlabelled competitor oligonucleotides (100-fold molar excess) as indicated.

Supplemental Figure 4. Dynamics of MAZ recruitment to the HBA genes during erythroid differentiation. (A) Diagram representing the expression of the transferrin receptor (CD71) and glycophorin A (GPA) during the second phase of *ex vivo* erythroblast differentiation cultures. **(B)** Flow cytometry analysis of CD71 and GPA during Phase II of a representative primary erythroblast differentiation culture. Cells initially acquire expression of CD71 alone (d6) before becoming double positive for both CD71 and GPA (d9-12). At later stages in culture, expression of CD71 decreases (d15). **(C)** Quantitation of flow cytometry staining as demonstrated in (B). Shown is the mean and standard deviation from two independent differentiation cultures. **(D)** Real-time RT-PCR analysis of expression of α- (HBA2) and β- (HBB) globin at the indicated time points during Phase II of primary erythroid differentiation cultures. Expression (relative to GAPDH) shows the mean and range from two independent differentiation cultures. **(E)** ChIP-qPCR analysis of MAZ binding at the α-globin genes at the indicated time points during Phase II of primary erythroid differentiation cultures. The y axis represents enrichment over the input DNA, normalised to a control sequence in the human 18S gene. The x axis indicates the Taqman probes used. The positions of

probes at the HBA Promoter/Exon1 and HBA Ex2, and probe 82 (used as a negative control) are indicated in Figure 2D. The promoter of the MYC gene, which has been previously shown to be bound by MAZ [18] was used as a positive control for the ChIP.

Supplemental Figure 5 (**A**) Photomicrographs of *ex vivo* erythroid differentiation culture of untransduced cells at the timepoints indicated. (**B**) *Ex vivo* differentiation cultures of cells transduced with the scrambled shRNA or the MAZ-targeting shRNA 699 or 703 at day 15 of culture. The MAZ shRNA cultures displayed an increase in the proportion of early stage progenitor BFU-E cells (small cells with very high nuclear:cytoplastic ratio) (black arrows). Scale bars (20 μm). Images were taken using an Olympus CX43 microscope mounted with Olympus SC50 camera. Cells stained with Leishman's Eosin-Methylene blue solution (Leishman's stain; VWR) (**C**) Effect of MAZ knockdown using shRNA 699 and 703 in K562 cells measured by real-time RT-PCR. Expression of MAZ (left panel) relative to the PABPC1 gene is normalized to the mean value observed with scramble shRNA. The expression of HBA relative to HBG (middle panel) or HBE (right panel) is normalised to the ratio observed with scrambled shRNA (SCR). Data shows average of two independent experiments with 3 technical repeats of each.

Supplemental Figure 6. Expression of MAZ in human and mouse haematopoietic populations. (A) Expression of MAZ across the normal mouse (left) and human (right) haematopoietic system. In both mouse and human haematopoietic systems, MAZ is highest expressed in multipotent progenitors, and in the erythroid lineage. The immunophenotype of the populations shown are as follows : MOUSE (1) Long Term Haematopoietic Stem Cell (Lin-, ckit+, Sca1+, Flk2-, CD34-); (2) Haematopoietic Stem

Cell (Lin-, ckit+, Sca1+, Flk2-, CD34+); (3) Multipotent Progenitor (Lin-, ckit+, Sca1+, Flk2+, CD34+); (4) Common Lymphoid Progenitor (Lin-, ckit+, Flk2+, Il7R+); (5) Common Myeloid Progenitor (Lin-, ckit+, Sca1-, CD34+, FcgRIII^{int}); (6) Megakaryocytic erythroid progenitor (Lin-, ckit+, Sca1-, Flk2-, CD34-); (7), Granulocyte Monocyte Progenitor (Lin-, ckit+, Sca1-, CD34+, FcgRIII^{high}); and corresponding HUMAN populations (1) Haematopoietic stem cell (CD133+, CD34dim); (5) Common myeloid progenitor (CD34+, CD38+, IL-3R^{lo}, CD45RA-); (6) Megakaryocyte/ erythroid progenitor (CD34+, CD38+, IL-3R-, CD45RA-); (7) Granulocyte/monocyte progenitor (CD34+, CD38+, IL-3R^{lo}, CD45RA+). (**B**) Normalised expression of MAZ in mature haematopoietic populations (* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$; one-way ANOVA with Bonferroni's correction for multiple testing. For (A) and (B), data for MAZ expression (microarray probe 212064_x_at on Affymetrix GeneChip HT-HG_U133A array) in human haematopoietic populations was extracted from the DMAP dataset (GSE24759 [19]), by Bloodspot [20,21]. Data for Maz expression in mouse haematopoietic populations was extracted from RNAseq datasets GSE14833 [22] and GSE6506 [23] by Bloodspot. Population trees in (A) were generated using Bloodspot.

Supplemental Figure 7. (A) The binding of MAZ to its own promoter overlaps with a SNP which is associated with increased haematocrit (see Figure 3G). A second SNP is located in the last intron and is associated with small RBC volume, and a third SNP is located in the 3'UTR and is associated with familial erythrocytosis. **(B)** MAZ occupies loci coding for erythroid transcription factors GATA1 and KLF1.

Supplemental Figure 8. MAZ ChIP-Seq quality control metrics. (A) Fingerprint of MAZ ChIP-Seq dataset compared to input showing an enrichment of DNA fragments for MAZ. **(B)** Cross-correlation plot profile for MAZ ChIP-Seq dataset indicating the normalized strand coefficient (NSC) and the ratio between the fragment-length peak and the read-length peak (relative strand correlation, RSC) for assessing signal-to-noise ratios. **(C)** Heat map representation of MAZ ChIP-seq signal at +/- 3 kb relative to the centre of called MAZ peaks. **(D)** The average relative MAZ signal around the MAZ peaks located at the promoters, intergenic and genic regions.

Supplemental Figure 9. Mutation analysis of binding of MAZ to the 13/14 probe. EMSA assays were carried out with the indicated probes and K562 nuclear extract. The consensus G_3CG_4 MAZ-binding motif is shown in bold. Variation from the wild type 13/14 probe in mutant probes is indicated in red. Variation from the G_3CG_4 consensus in probe 11/12 is indicated in blue. The retarded species caused by MAZ (species (d)) is indicated by an arrow.

Supplemental Figure 10. MAZ-like position weight matrices. (A). The alignment of the MAZ position weight matrix derived from our ChIP-seq dataset, to significantly related (E_value <0.01) binding profiles of transcription factors present in the JASPAR database. Also shown is the alignment with the PWM of KLF3. **(B).** Expression of the

corresponding TFs and MAZ during erythroid differentiation as measured by RNAseq. The y axis represents FPKM and the x-axis indicates days of differentiation. Note KLF5 is not expressed in erythroid cells. **(C).** Comparison of experimentally-determined PWMs (left panel), with matrices predicted based on inferred contact energies for the C2H2-ZF domains of each protein (right panel).

Supplemental Figure 11. Expression of MAZ, SP1, Sp3 and KLF3 in some of the cell lines used in this study. RNAseq expression data were extracted from proteinatlas.org. The y axis shows the normalized expression values.

Supplemental Table 1. Results of affinity-purification screen. Proteins detected in both WT replicates and showing enriched binding (nom significance<0.05) to the 13/14 WT probe relative to the 13/14 M3 mutant probe. Shown is the gene name, description and unique identifiers (Uniprot, VEGA, Ensembl and Refseq). Pfam annotations (protein domains) were mapped to the protein entries. MS/MS count indicates the number of spectra recorded for all peptides of the protein group binding to the WT and mutant probes (from two independent replicate experiments). Ratio for enrichment between the two oligonucleotide baits is calculated based on protein intensity and significance from MaxQuant analysis [24].

Supplemental Table 2. Erythroid-related traits used to search in the GeneATLAS database for MAZ variants.

Supplemental Table 3: Datasets used in the study

Supplemental Table 4. Tissue specific MAZ peaks in ENCODE human cell lines and human primary erythroid cells.

Supplemental Table 5. Location of MAZ peaks within genomic loci associated with erythroid traits and erythropoiesis.

Supplemental Table 6. Sequences of oligonucleotides used in this study

REFERENCES

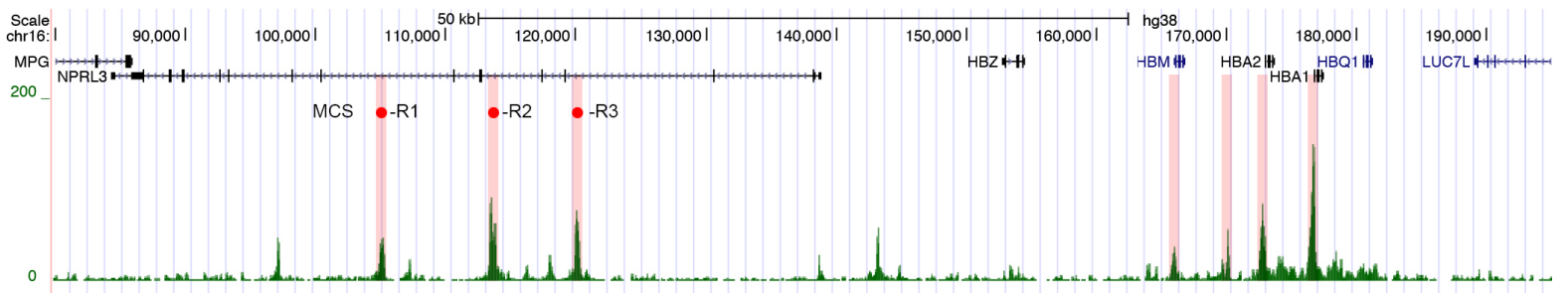
1. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360.
2. Diaz A, Park K, Lim DA, Song JS. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol*. 2012;11(3):Article 9.
3. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012;22(9):1813-1831.
4. Ramirez F, Ryan DP, Gruning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-165.
5. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
6. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-589.
7. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11 12 11-34.
8. van de Lagemaat LN, Flenley M, Lynch MD, et al. CpG binding protein (CFP1) occupies open chromatin regions of active genes, including enhancers and non-CpG islands. *Epigenetics Chromatin*. 2018;11(1):59.
9. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res*. 2015;43(W1):W39-49.
10. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017-1018.
11. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*. 2012;40(17):e128.
12. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):R24.
13. Persikov AV, Singh M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res*. 2014;42(1):97-108.
14. Raudvere U, Kolberg L, Kuzmin I, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019.
15. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50(11):1593-1599.
16. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *ournal of the Royal Statistical Society Series B* 1995;57(1):12.

17. 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.
18. Bossone SA, Asselin C, Patel AJ, Marcu KB. MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A*. 1992;89(16):7452-7456.
19. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296-309.
20. Bagger FO, Sasivarevic D, Sohi SH, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res*. 2016;44(D1):D917-924.
21. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res*. 2019;47(D1):D881-D885.
22. Di Tullio A, Vu Manh TP, Schubert A, Castellano G, Mansson R, Graf T. CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc Natl Acad Sci U S A*. 2011;108(41):17016-17021.
23. Chambers SM, Boles NC, Lin KY, et al. Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*. 2007;1(5):578-591.
24. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372.
25. Xu J, Shao Z, Glass K, et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev Cell*. 2012;23(4):796-811.
26. Huang J, Liu X, Li D, et al. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell*. 2016;36(1):9-23.
27. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

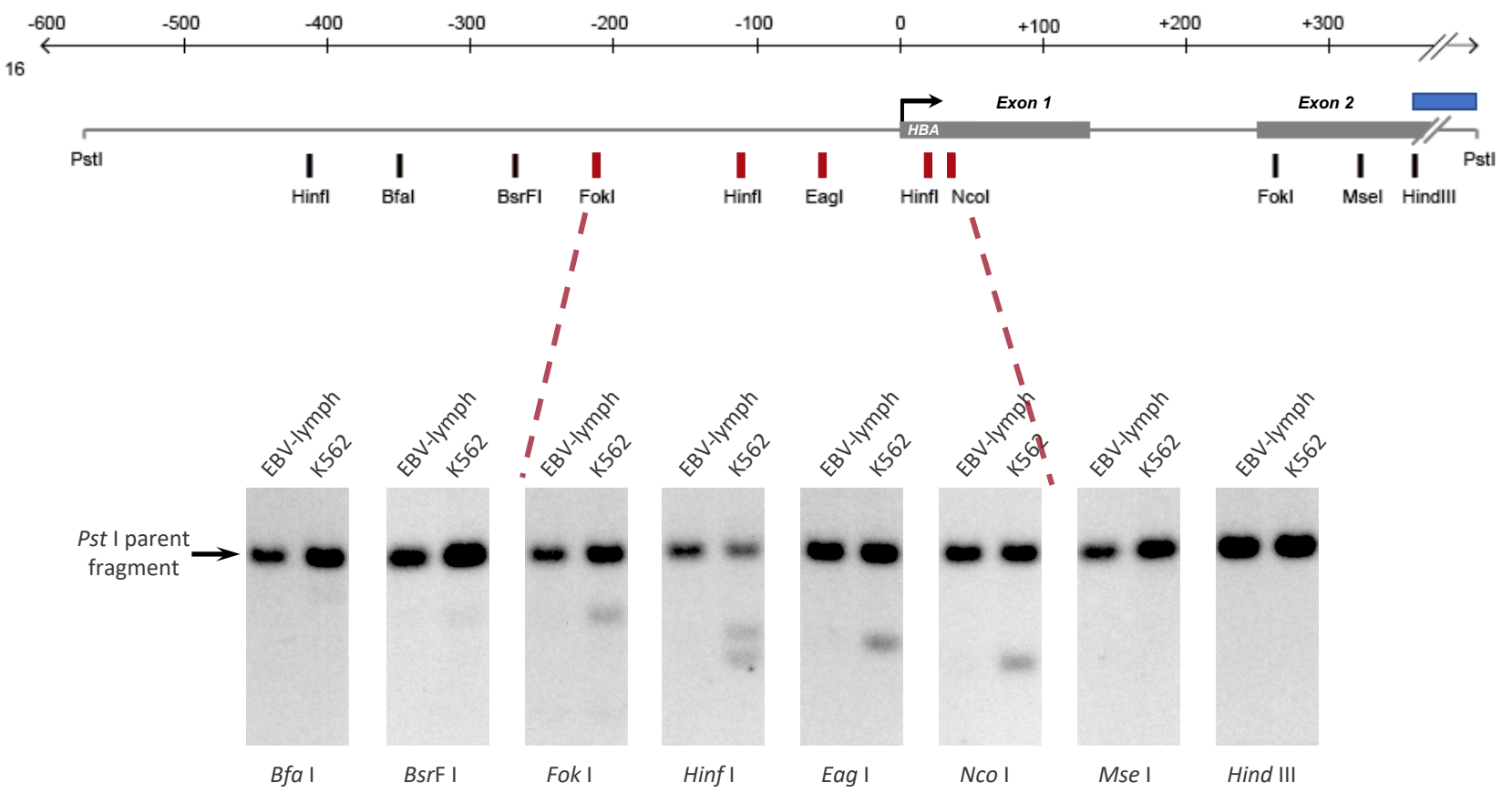
Deen et al.,

Suppl Figure 1

A



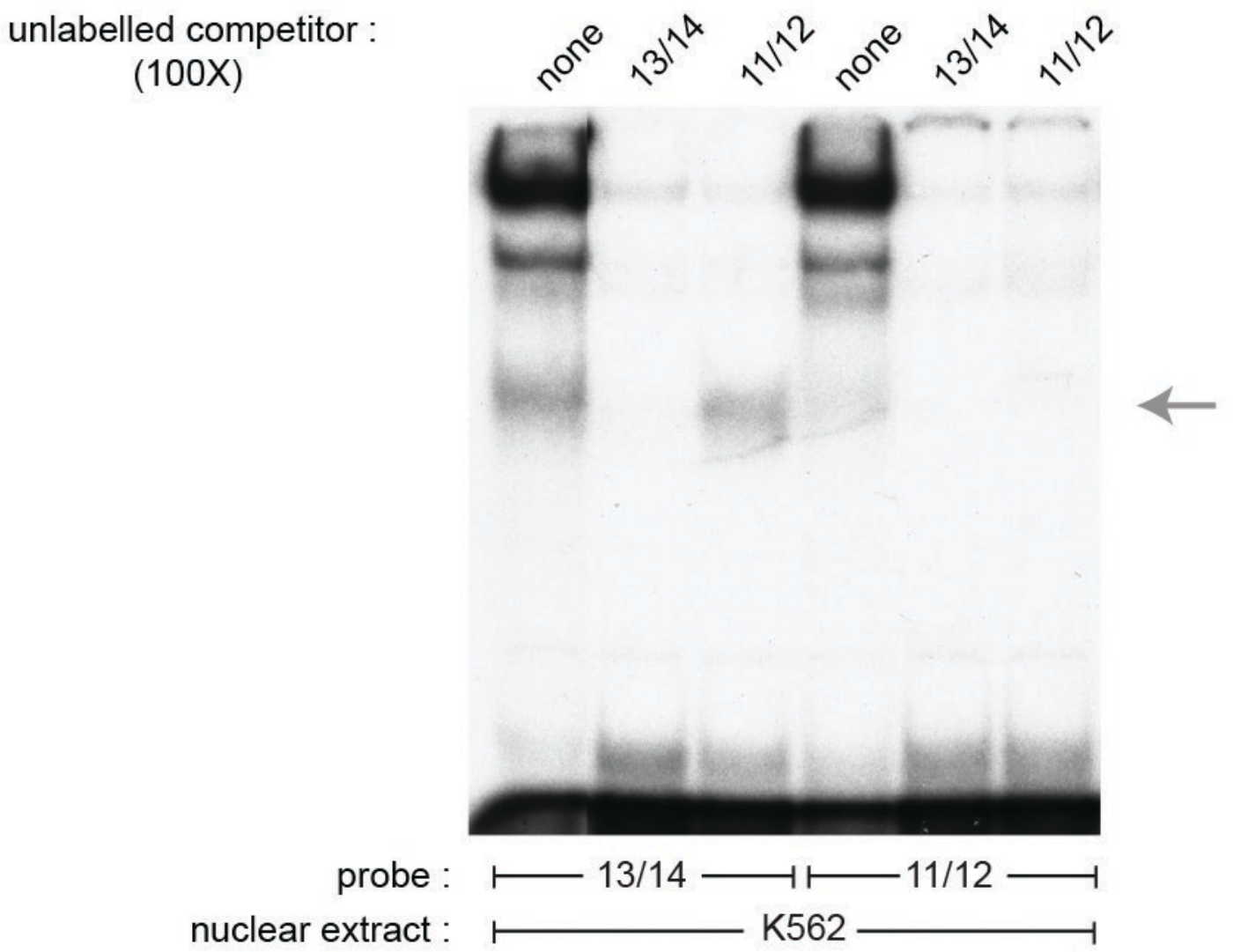
B

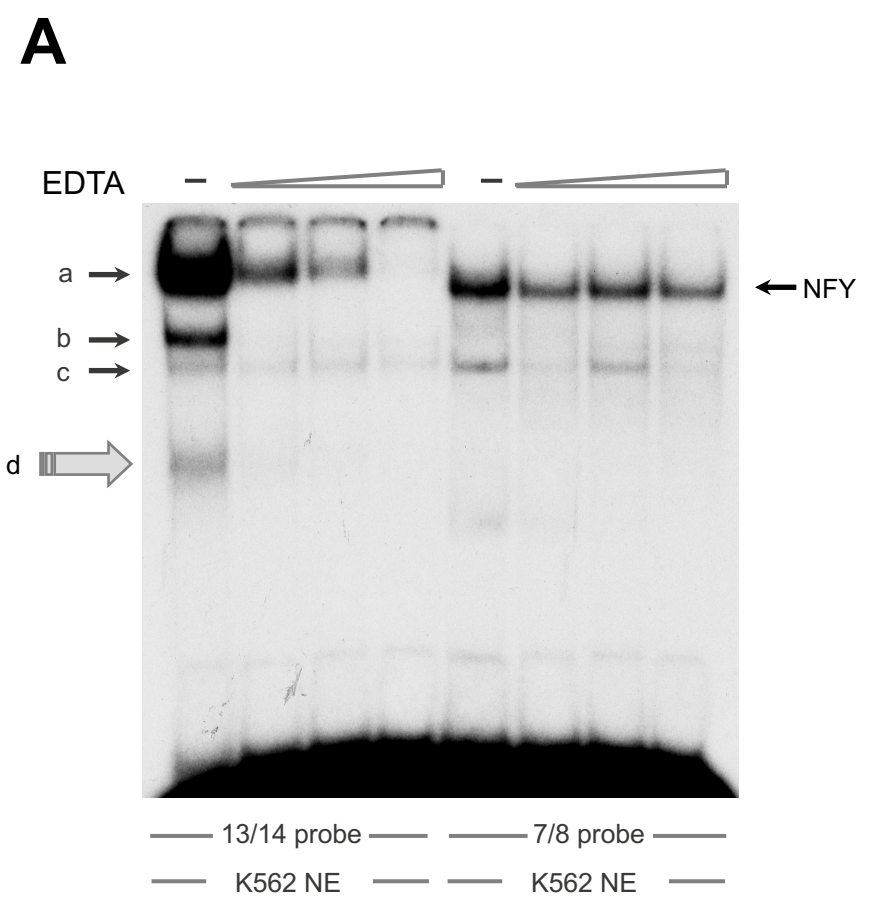


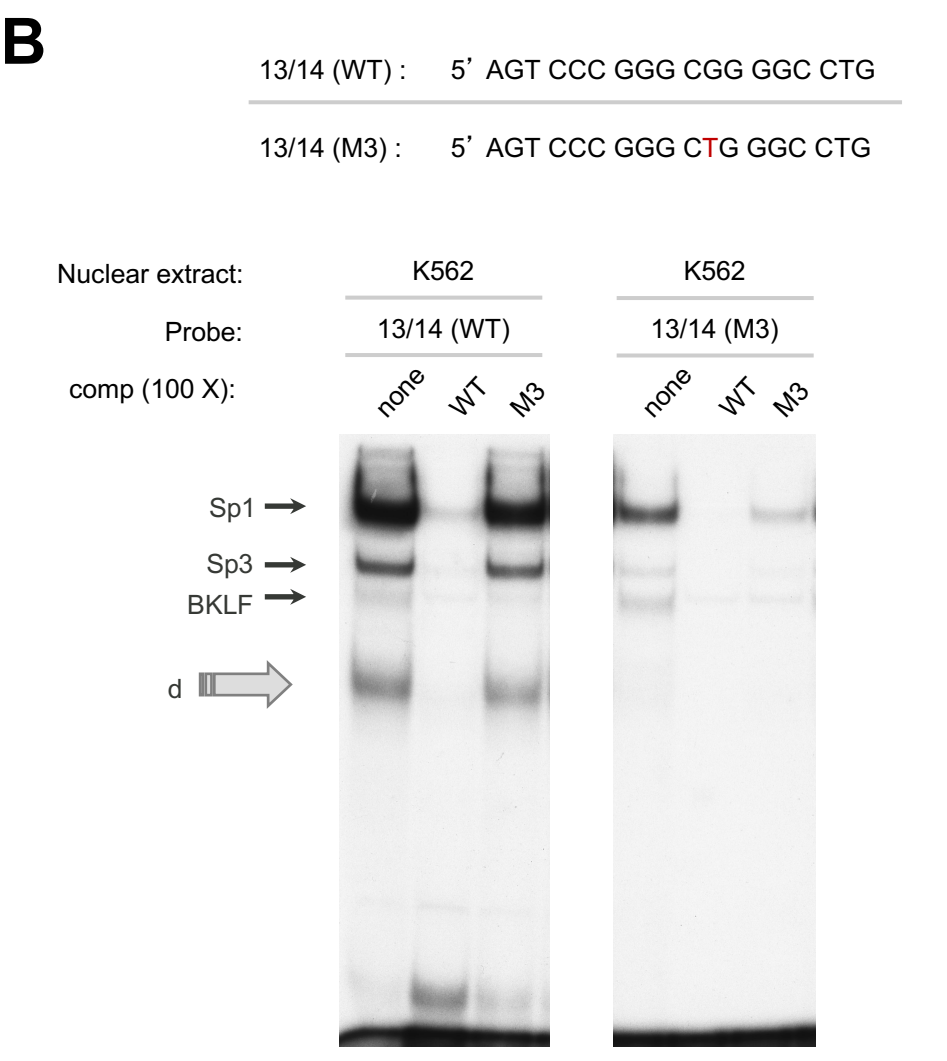
Suppl Figure 2

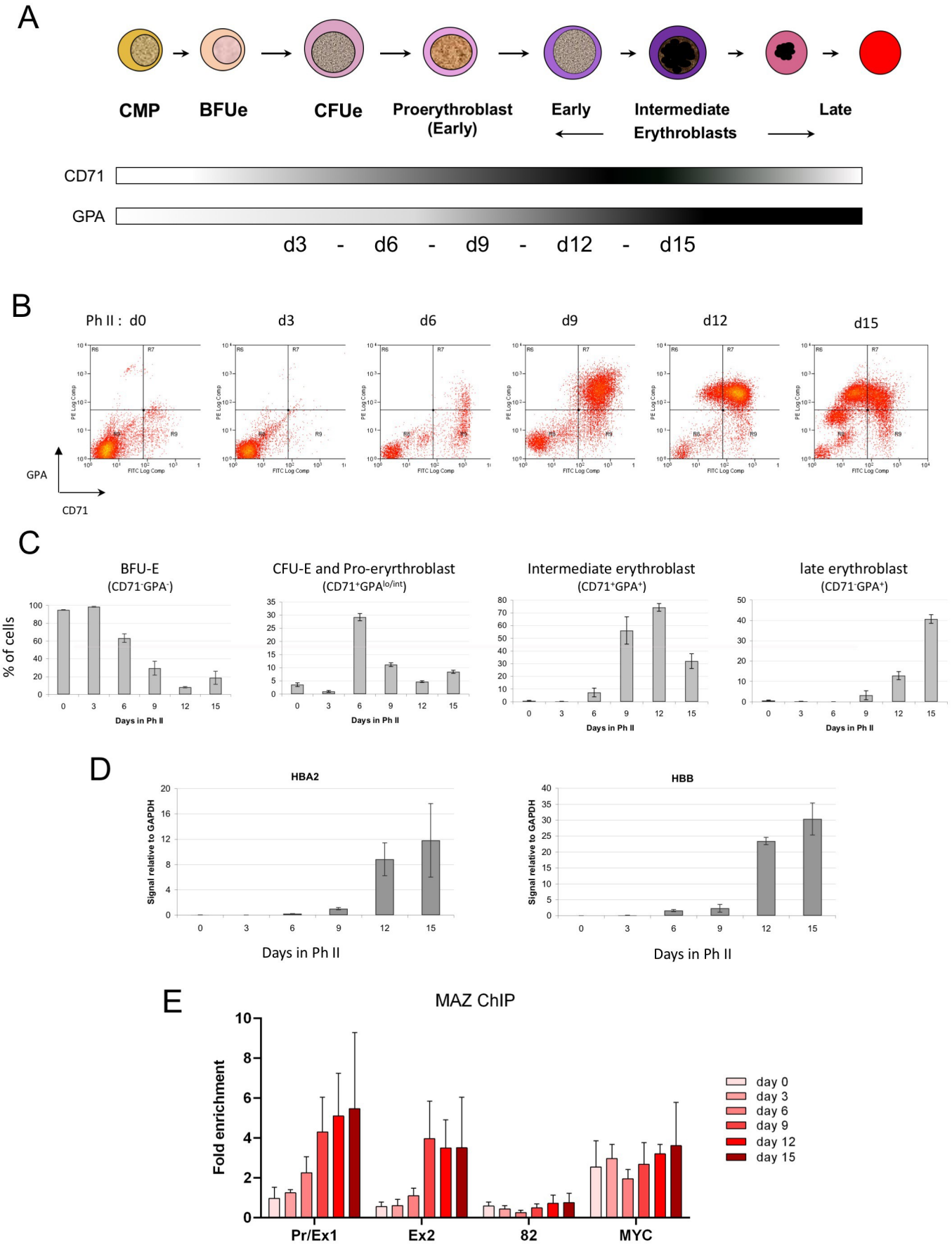
A

B

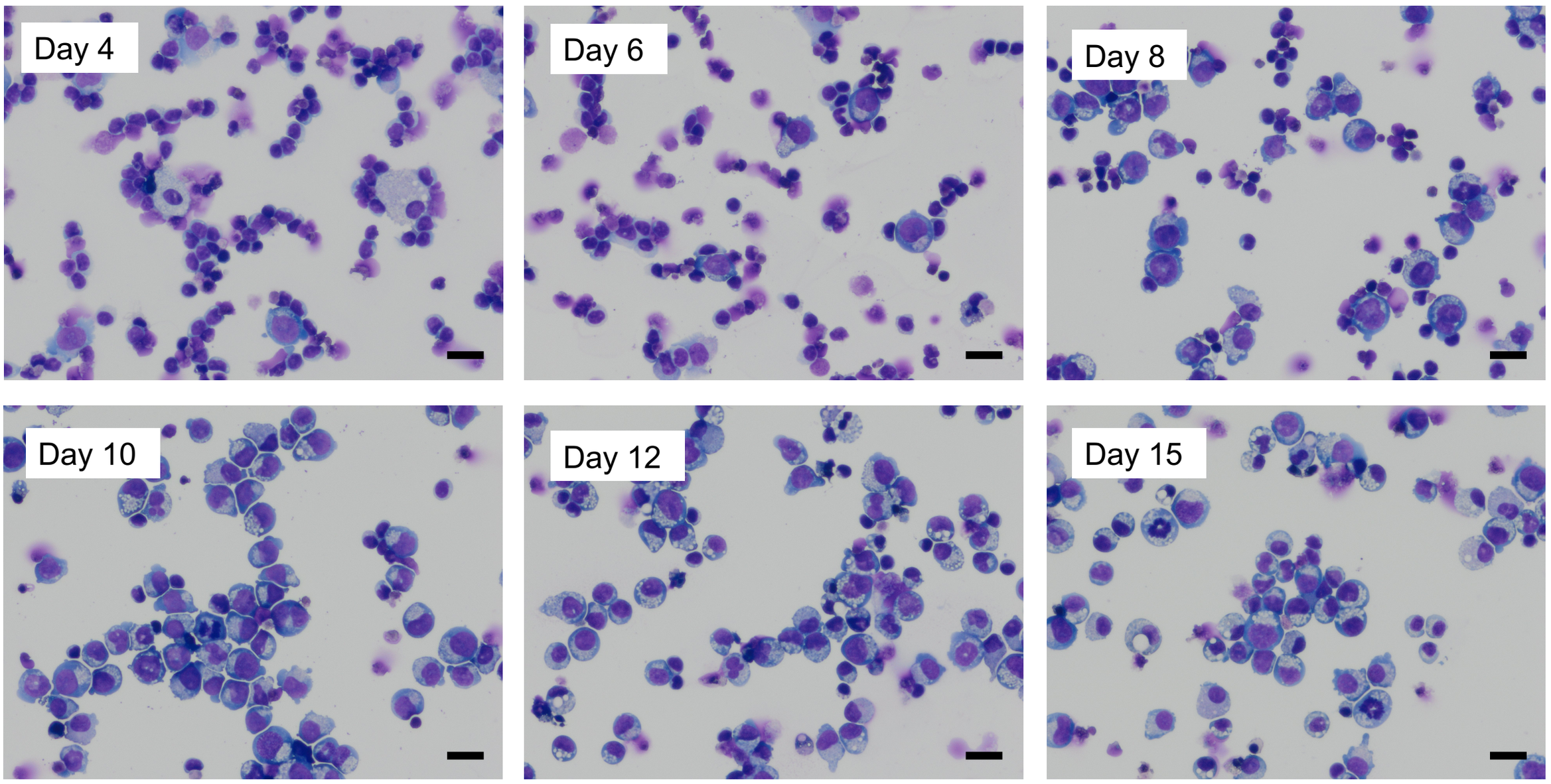



A
EDTA
–
–
a
b
c
d
NFY
13/14 probe
7/8 probe
K562 NE
K562 NE

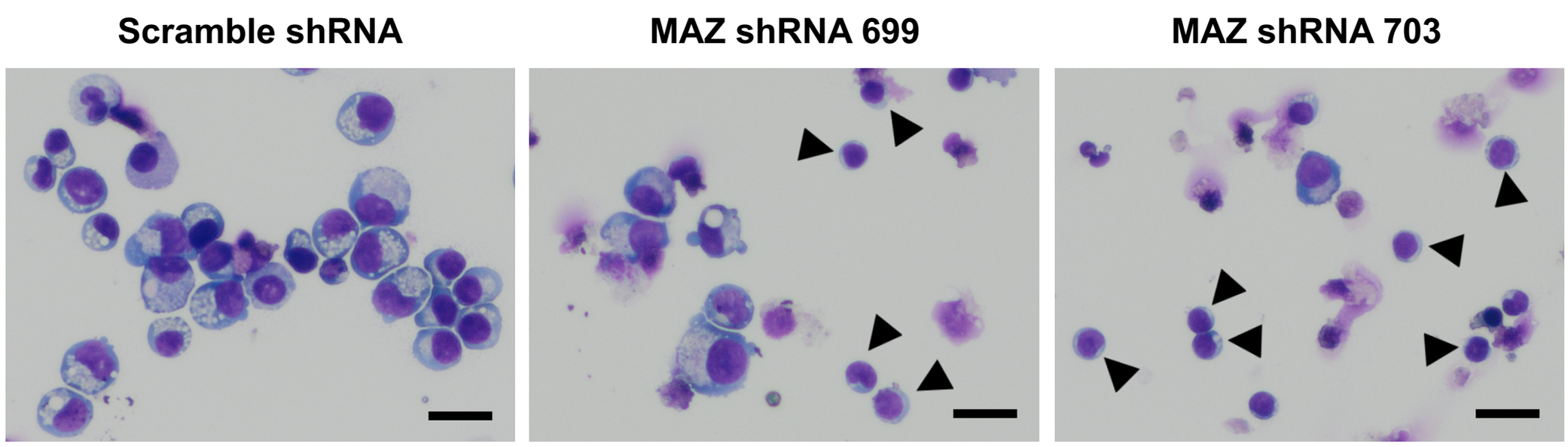

B
13/14 (WT) : 5’ AGT CCC GGG CGG GGC CTG
13/14 (M3) : 5’ AGT CCC GGG CTG GGC CTG
Nuclear extract:
K562
K562
Probe:
13/14 (WT)
13/14 (M3)
comp (100 X):
none
WT
M3
none
WT
M3
Sp1
Sp3
BKLF
d


A
CMP
BFUe
CFUe
Proerythroblast (Early)
Early
Intermediate Erythroblasts
Late
CD71
GPA
d3 - d6 - d9 - d12 - d15
B
Ph II : d0
d3
d6
d9
d12
d15
GPA
CD71
C
BFU-E (CD71-GPA-)
CFU-E and Pro-eryrthroblast (CD71+GPAlo/int)
Intermediate erythroblast (CD71+GPA+)
late erythroblast (CD71-GPA+)
% of cells
Days in Ph II
D
HBA2
HBB
Signal relative to GAPDH
Days in Ph II
E
MAZ ChIP
Fold enrichment
Pr/Ex1
Ex2
82
MYC
day 0
day 3
day 6
day 9
day 12
day 15

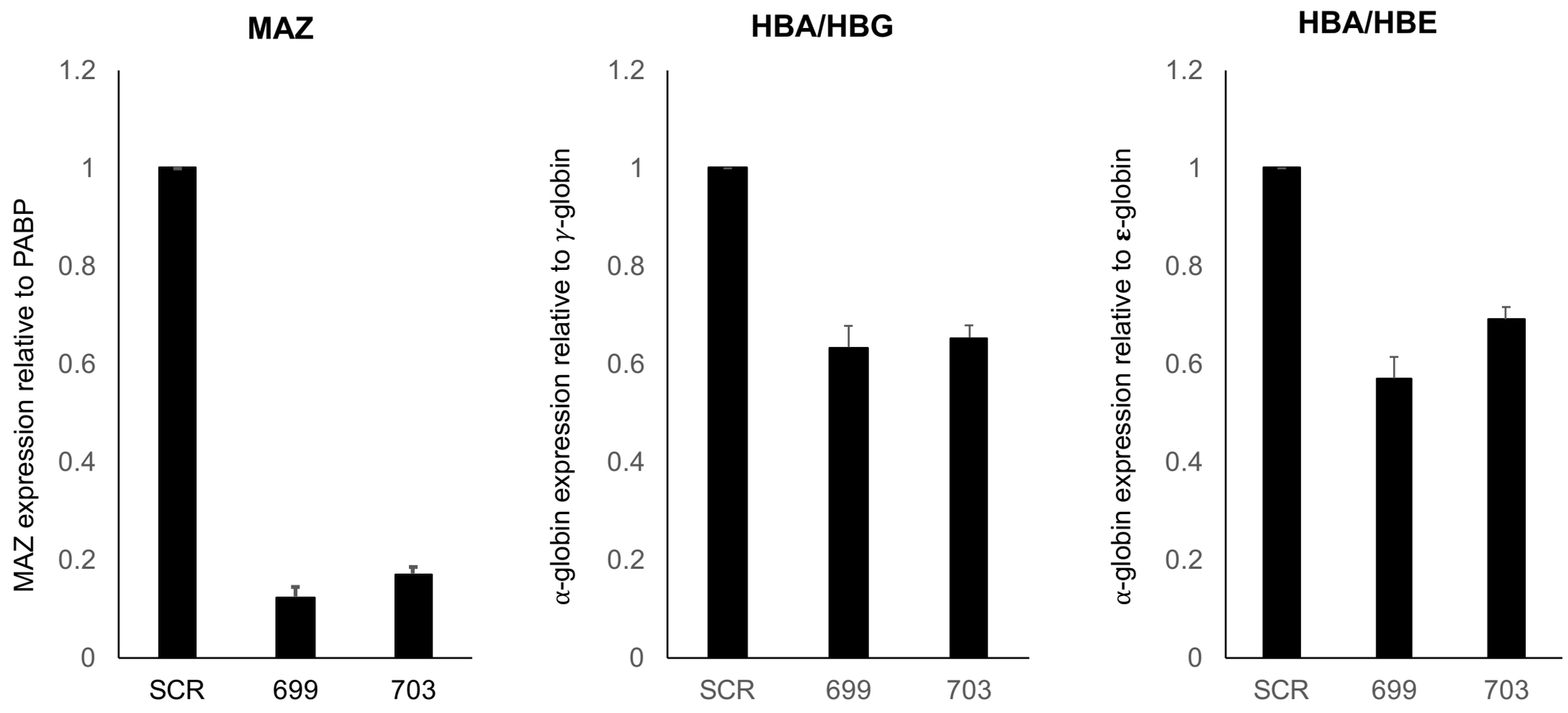
A

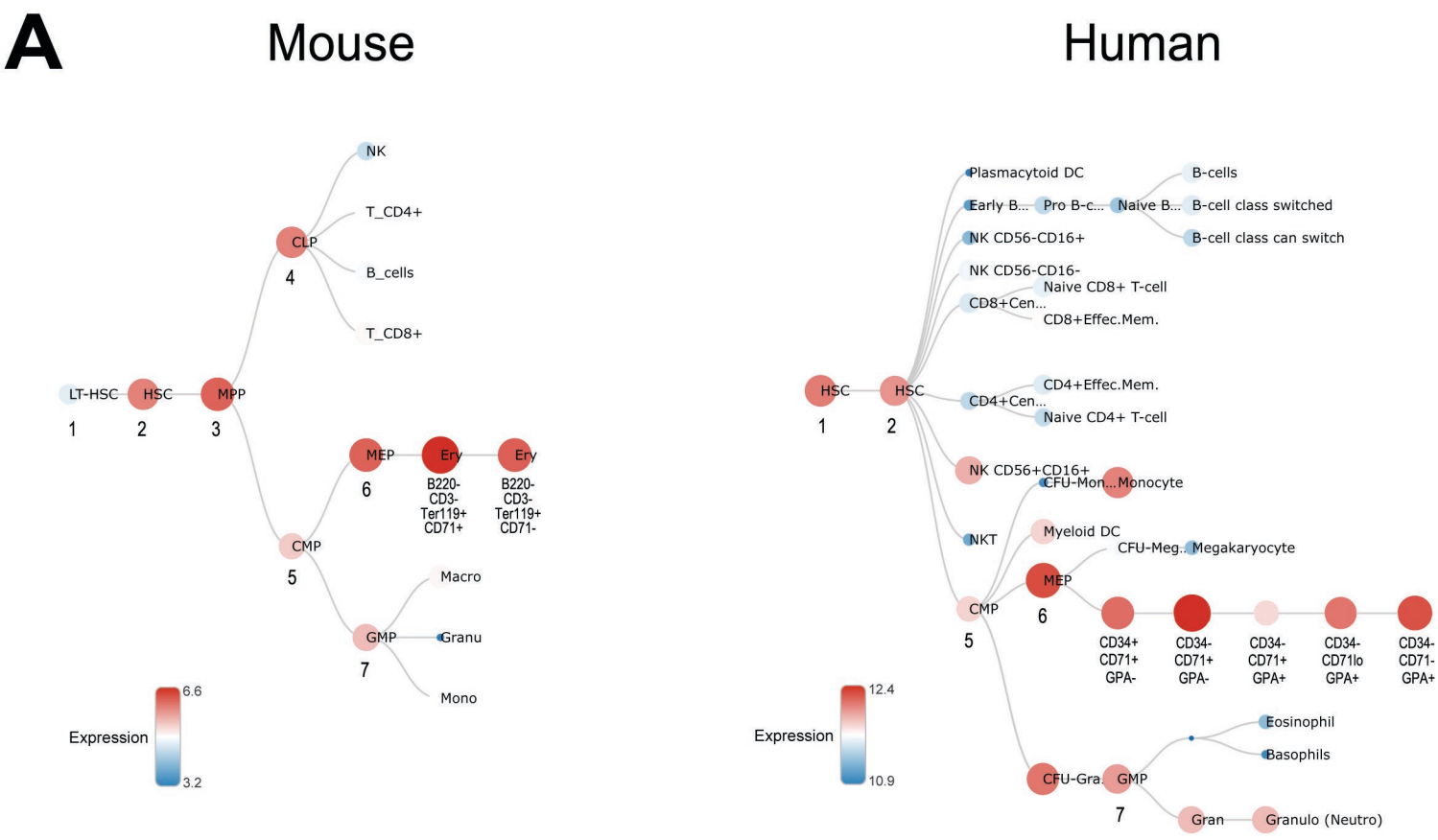

Day 4
Day 6
Day 8
Day 10
Day 12
Day 15

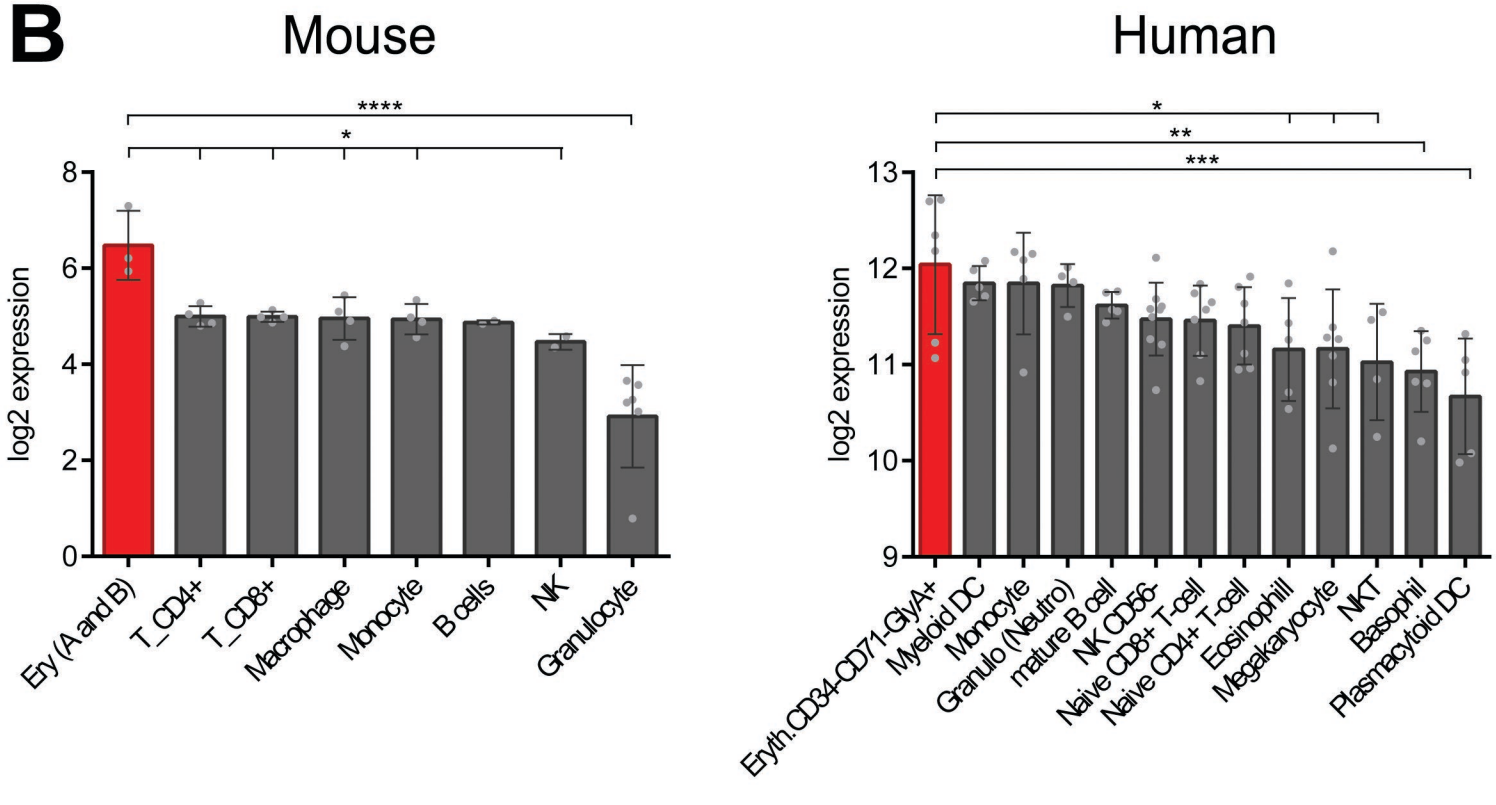
B


Scramble shRNA
MAZ shRNA 699
MAZ shRNA 703

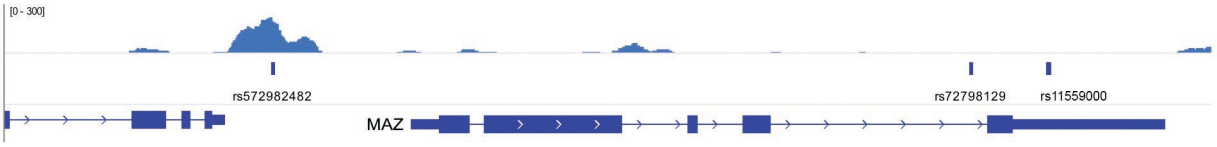
C


MAZ
MAZ expression relative to PABP
1.2
1
0.8
0.6
0.4
0.2
0
SCR
699
703
HBA/HBG
α-globin expression relative to γ-globin
1.2
1
0.8
0.6
0.4
0.2
0
SCR
699
703
HBA/HBE
α-globin expression relative to ε-globin
1.2
1
0.8
0.6
0.4
0.2
0
SCR
699
703

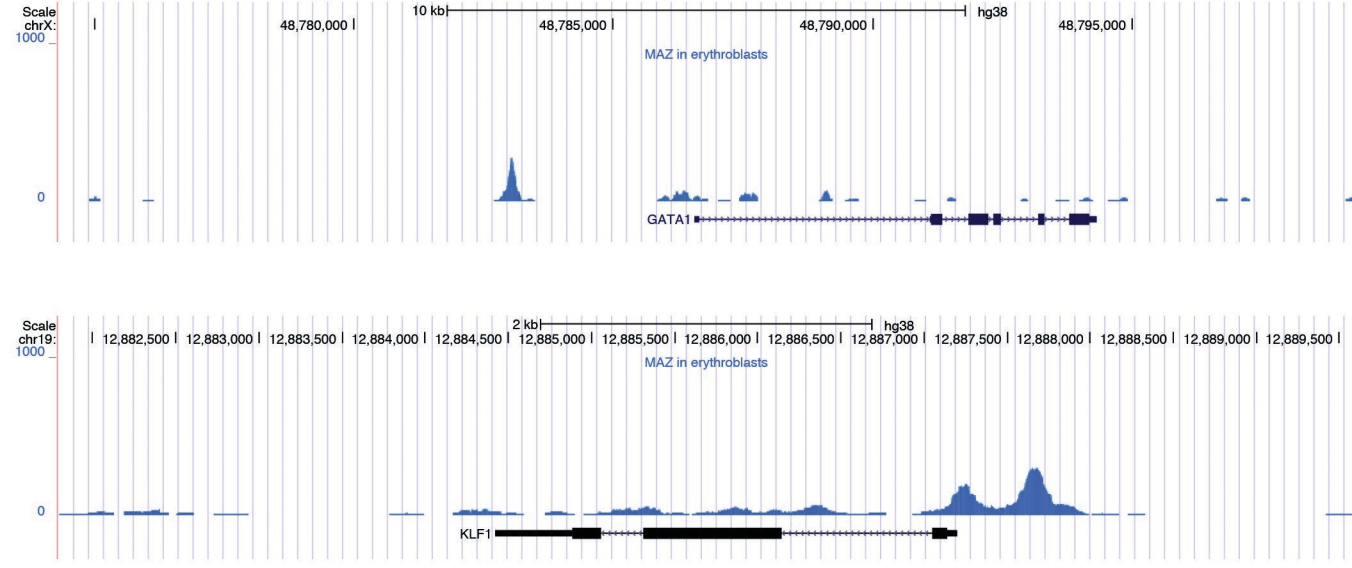

A
Mouse
Human
NK
T_CD4+
CLP
4
B_cells
T_CD8+
LT-HSC
HSC
MPP
1
2
3
MEP
Ery
Ery
6
B220-
CD3-
Ter119+
CD71+
B220-
CD3-
Ter119+
CD71-
CMP
5
Macro
GMP
Granu
7
Mono
Expression
6.6
3.2
Plasmacytoid DC
B-cells
Early B...
Pro B-c...
Naive B...
B-cell class switched
NK CD56-CD16+
B-cell class can switch
NK CD56-CD16-
Naive CD8+ T-cell
CD8+Cen...
CD8+Effec.Mem.
HSC
HSC
CD4+Effec.Mem.
CD4+Cen...
Naive CD4+ T-cell
NK CD56+CD16+
CFU-Mon...
Monocyte
NKT
Myeloid DC
CFU-Meg...
Megakaryocyte
MEP
CMP
6
5
CD34+
CD71+
GPA-
CD34-
CD71+
GPA-
CD34-
CD71+
GPA+
CD34-
CD71lo
GPA+
CD34-
CD71-
GPA+
Expression
12.4
10.9
Eosinophil
Basophils
CFU-Gra...
GMP
7
Gran
Granulo (Neutro)

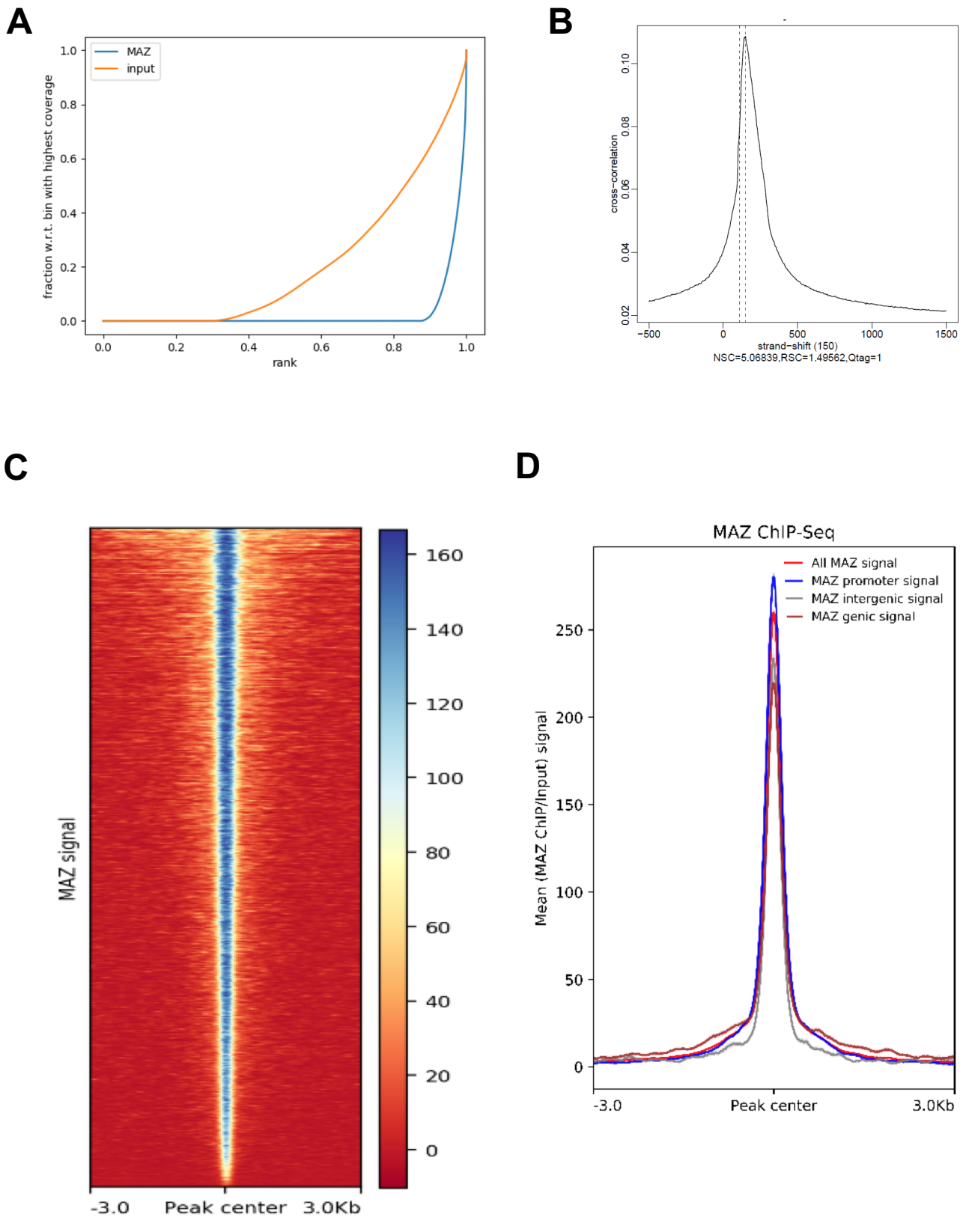

B
Mouse
Human
****
*
log2 expression
8
6
4
2
0
Ery (A and B)
T_CD4+
T_CD8+
Macrophage
Monocyte
B cells
NK
Granulocyte
*
**
***
log2 expression
13
12
11
10
9
Eryth.CD34-CD71-GlyA+
Myeloid DC
Monocyte
Granulo (Neutro)
mature B cell
NK CD56-
Naive CD8+ T-cell
Naive CD4+ T-cell
Eosinophill
Megakaryocyte
NKT
Basophil
Plasmacytoid DC

A


[0 - 300]
rs572982482
rs72798129
rs11559000
MAZ

B

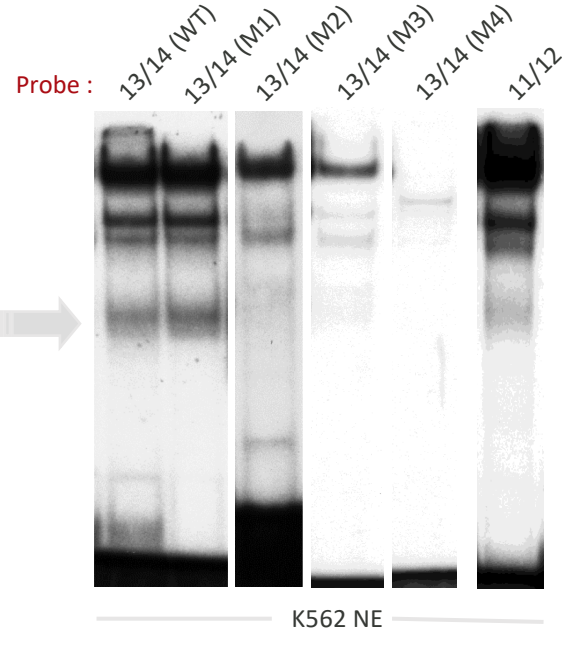

Scale
chrX:
1000
0
48,780,000
10 kb
48,785,000
48,790,000
hg38
48,795,000
MAZ in erythroblasts
GATA1
Scale
chr19:
1000
0
12,882,500
12,883,000
12,883,500
12,884,000
12,884,500
2 kb
12,885,000
12,885,500
12,886,000
12,886,500
12,887,000
hg38
12,887,500
12,888,000
12,888,500
12,889,000
12,889,500
MAZ in erythroblasts
KLF1

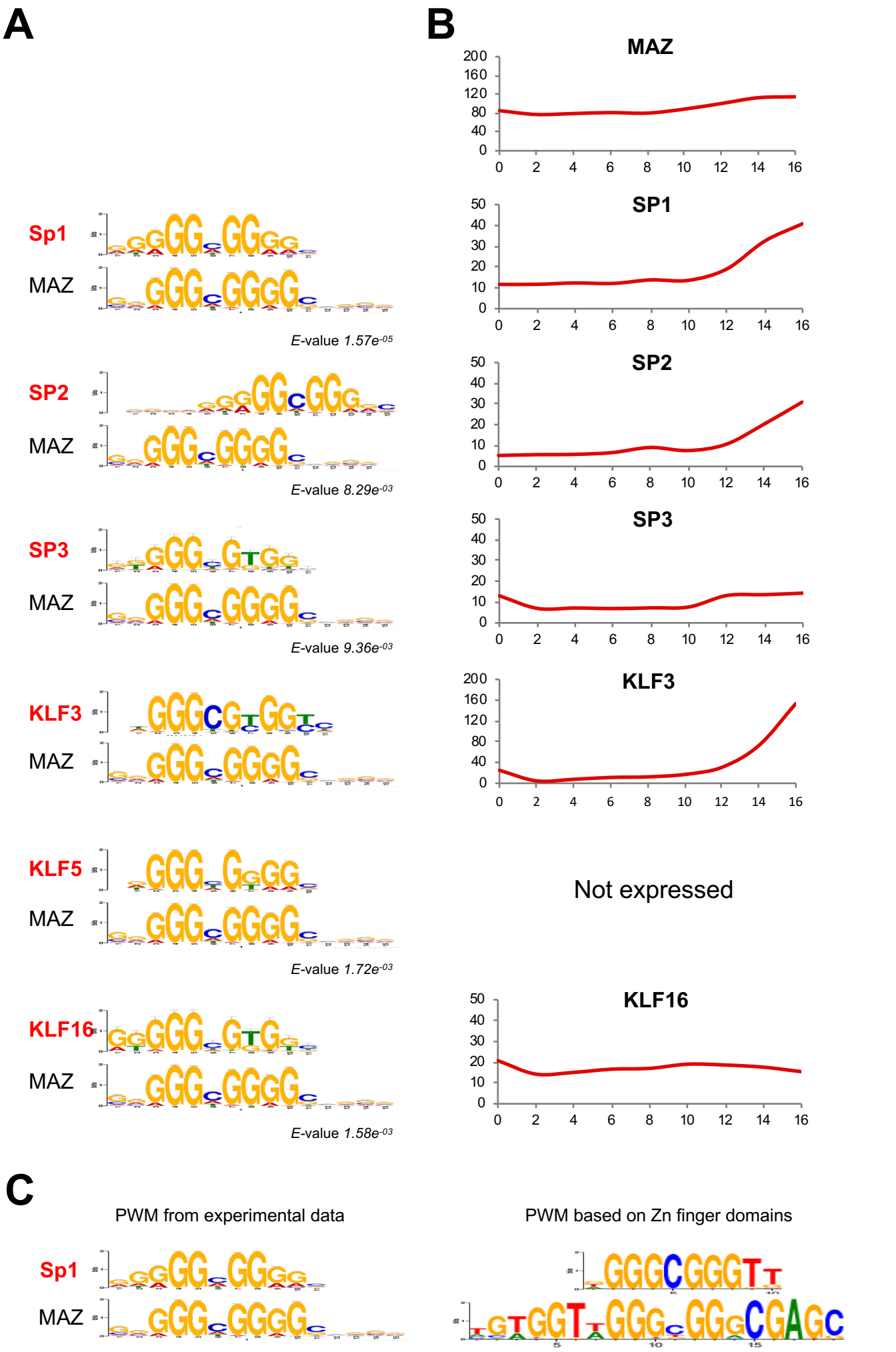

A
MAZ
input
fraction w.r.t. bin with highest coverage
rank
B
cross−correlation
strand−shift (150)
NSC=5.06839,RSC=1.49562,Qtag=1
C
MAZ signal
-3.0
Peak center
3.0Kb
D
MAZ ChIP-Seq
All MAZ signal
MAZ promoter signal
MAZ intergenic signal
MAZ genic signal
Mean (MAZ ChIP/Input) signal
-3.0
Peak center
3.0Kb

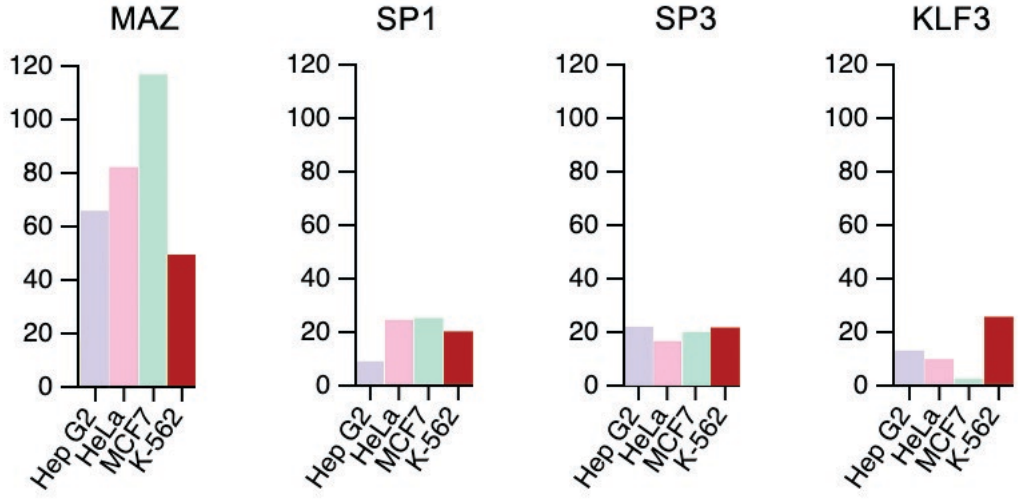
Deen et al.,

Suppl Figure 9

Oligo	Sequence	Binding of MAZ
13/14 (WT) :	5' AGT CCC **GGG CGG GGC** CTG	+
13/14 (M1) :	5' **CAC A**CC **GGG CGG GGC** CTG	+
13/14 (M2) :	5' AGT CCC **GGG CGT ATA** CTG	-
13/14 (M3) :	5' AGT CCC **GGG CTG GGC** CTG	- (see also Sup. Fig.3B)
13/14 (M4) :	5' AGT CCC **GTG CGG GGC** CTG	-
11/12 (WT):	5' GC CCG **GGG CGC GGC** CTG	<




A
Sp1
MAZ
E-value 1.57e-05
SP2
MAZ
E-value 8.29e-03
SP3
MAZ
E-value 9.36e-03
KLF3
MAZ
E-value 1.72e-03
KLF5
MAZ
E-value 1.58e-03
KLF16
MAZ
B
MAZ
SP1
SP2
SP3
KLF3
Not expressed
KLF16
C
PWM from experimental data
Sp1
MAZ
PWM based on Zn finger domains


MAZ
SP1
SP3
KLF3
120
100
80
60
40
20
0
Hep G2
HeLa
MCF7
K-562

Suppl. Table 1: Results from mass spectrometry screen

MS/MS Counts — The number of MS/MS spectra recorded for all peptides of the protein group
Enrichment — Calculated enrichment using the summed protein ion intensity
Significance Enrichment — Statistical assessment of significance of enrichment

	Gene Names	Protein Names	Mol. Weight [kDa]	Uniprot	Pfam	Pfam Descriptions	ENSEMBL	Chrom	Strand	Position	MS/MS Count (MUT13/14) replicate 1	MS/MS Count (MUT13/14) replicate 2	MS/MS Count wild-type replicate 1	MS/MS Count wild-type replicate 2	EnrichmentWT/MUT	Significance Enrichment
1	MAZ	MAZ protein;Myc-associated zinc finger protein;MAZI;Purine-binding transcription factor;Pur	51.1	Q8IUI2;Q8NFN7;P56270;Q59GP8;Q8N	PF00096	Zinc finger, C2H2 type	ENSG00000103495	16	+	29725356	0	1	10	19	14.3	5.2E-06
2	HIST1H1D;H1F3	Histone H1.3;Histone H1c	22.4	P16402	PF00538	linker histone H1 and H5 family	ENSG00000124575	6	-	26342475	3	0	1	3	10.9	3.7E-05
3	HIST1H1E;H1F4	Histone H1.4;Histone H1b	21.9	P10412;A3R0T7;A3R0T8;Q4VB24	PF00538	linker histone H1 and H5 family	ENSG00000168298	6	+	26264566	2	1	1	3	9.2	1.1E-04
4	H1FX	Histone H1x	22.5	Q92522	PF00538	linker histone H1 and H5 family	ENSG00000184897	3	-	130516794	2	3	4	21	9.0	1.3E-04
5	RSL1D1;CATX11;CSIG;PBK1;L12	Ribosomal L1 domain-containing protein 1;Cellular senescence-inhibited gene protein;Protei	55.0	O76021;A0PJ87;Q32Q62;A0PJ61			ENSG00000171490	16	-	11835554	1	2	5	18	8.8	1.5E-04
6	EBNA1BP2;EBP2	Probable rRNA-processing protein EBP2;EBNA1-binding protein 2;Nucleolar protein p40	34.9	Q99848;Q6IB29	PF05890	Eukaryotic rRNA processing protein EBP2	ENSG00000117395	1	-	43402440	0	1	5	19	8.6	1.8E-04
7	HIST1H1B;H1F5	Histone H1.5;Histone H1a	22.6	P16401	PF00538	linker histone H1 and H5 family	ENSG00000184357	6	-	27942606	6	5	5	8	8.2	2.4E-04
8	RBM34;KIAA0117;RP11-739C15.1-008	RNA-binding protein 34;RNA-binding motif protein 34;Novel protein	48.6	P42696;A8K8J7;A2A2V2	PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	ENSG00000188739	1	-	233361121	0	4	5	33	8.0	2.8E-04
9	DHX36;DDX36;KIAA1488;MLE1;RHAU	Probable ATP-dependent RNA helicase DHX36;DEAH box protein 36;MLE-like protein 1;RNA h	114.8	Q9H2U1	PF00270;PF07717;PF04408;PF00271	DEAD/DEAH box helicase;Domain of unknown function (I	ENSG00000174953	3	-	155476152	6	6	42	55	7.4	4.4E-04
10	DDX21	Nucleolar RNA helicase 2;Nucleolar RNA helicase II;Nucleolar RNA helicase Gu;RH II/Gu;Gu-alp	87.3	Q9NR30;Q3SWU7;Q8NI92	PF00270;PF08152;PF00271	DEAD/DEAH box helicase;GUCT (NUC152) domain;Helicas	ENSG00000165732	10	+	70385898	2	1	16	50	7.3	4.8E-04
11	101F10.1	Putative uncharacterized protein 101F10.1	51.6	O43328;Q1ED39;Q5FWF3			ENSG00000103550	16	-	19625224	2	8	8	44	6.3	1.1E-03
12	HLTF;HIP116A;RNF80;SMARCA3;SNF2L3;ZBU1	cDNA FLJ76830, highly similar to Homo sapiens SWI/SNF related, matrix associated, actin dep	113.9	A8K5B6;Q14527;Q05BZ6;Q05DF7;Q5	PF00271;PF08797;PF00176;PF00097	Helicase conserved C-terminal domain;HIRAN domain;SN	ENSG00000071794	3	-	150230594	0	0	2	17	5.9	1.7E-03
13	RFC1;RFC140	Replication factor C subunit 1;Replication factor C large subunit;RF-C 140 kDa subunit;Activa	128.3	P35251;A8K6E7;Q14756;A6NH54	PF00004;PF00533;PF08519	ATPase family associated with various cellular activities (	ENSG00000035928	4	-	38965471	0	2	2	35	5.8	1.8E-03
14	HIST1H1C;H1F2	Histone H1.2;Histone H1d	21.4	P16403;A8K4I2	PF00538	linker histone H1 and H5 family	ENSG00000187837	6	-	26163994	4	11	13	28	5.6	2.2E-03
15	NOL5A;NOP56;RP4-686C3.1-010;RP4-686C3.1-016	Nucleolar protein 5A;Nucleolar protein Nop56;56kDa with KKE/D repeat	66.0	O00567;A0PJ92;A8K9K6;Q9BSN3;Q5J	PF01798;PF08156;PF08060	Putative snoRNA binding domain;NOP5 NT (NUC127) dom	ENSG00000101361	20	+	2581178	0	0	3	13	5.5	2.4E-03
16	RPS17;LOC392101;hCG_1808625;tcag7.955	Uncharacterized protein ENSP00000352682;40S ribosomal protein S17;Similar to dJ753D5.	15.9	A8MVC5;A6NH77;P08708;A4D1Q6	PF00833	Ribosomal S17	ENSG00000197575;ENSG0	5	+	116079834	1	0	3	2	5.4	2.7E-03
17	NKRF;ITBA4;NRF	NF-kappa-B-repressing factor;NFkB-repressing factor;Transcription factor NRF;ITBA4 protein	77.7	O15226;A1L1A7;A2I827;A2I828;A2I82	PF00035;PF01585;PF01424	Double-stranded RNA binding motif;G-patch domain;R3H	ENSG00000184416	X	-	118606328	2	1	1	26	5.3	2.9E-03
18	RBM28	RNA-binding protein 28;RNA-binding motif protein 28	85.7	Q9NW13;A4D100;Q53H65	PF00076	RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	ENSG00000106344	7	-	127737673	0	1	1	13	5.2	3.1E-03
19	RPL31	60S ribosomal protein L31;RPL31 protein	14.5	P62899;Q76N53;Q6IR20;A6NF91	PF01198	Ribosomal protein L31e	ENSG00000071082	2	+	100985183	0	3	4	3	5.2	3.2E-03
20	SP3	Transcription factor Sp3;SPR-2	81.9	Q02447;Q59FX5;Q6BDP2;Q86TP0;Q8	PF00096	Zinc finger, C2H2 type	ENSG00000172845	2	-	174481505	0	0	2	2	#DIV/0!	3.5E-03
21	YBX1;NSEP1;YB1	Nuclease sensitive element-binding protein 1;Y-box-binding protein 1;Y-box transcription fac	35.9	P67809;A0JLU4;Q05D43;Q6PKI6;A6N	PF00313	'Cold-shock' DNA-binding domain	ENSG00000065978	1	+	42920659	19	26	59	56	4.9	4.1E-03
22	HP1BP3;RP5-930I4.3-001;hCG_38870;RP5-930I4.3-005;RP5-930I4.3-002;DKFZp434I0612	Heterochromatin protein 1, binding protein 3;Heterochromatin protein 1, binding protein 3	61.4	A6NI71;A8K5D7;Q05BI0;Q5SJI5;Q9UH	PF00538	linker histone H1 and H5 family	ENSG00000127483	1	-	20941758	1	3	4	15	4.9	4.2E-03
23	ZNF512;KIAA1805;DKFZp666L156	Zinc finger protein 512;Putative uncharacterized protein DKFZp666L156	64.6	Q96ME7;Q53RZ7;Q8S8M0;Q86XK6	PF00096	Zinc finger, C2H2 type	ENSG00000198522	2	+	27659397	0	0	1	5	4.8	4.5E-03
24	ZNF579	Zinc finger protein 579	60.5	Q8NAF0	PF00096	Zinc finger, C2H2 type	ENSG00000179943	19	-	60780705	1	0	1	1	4.6	5.6E-03
25	ZNF646;hCG_18384;KIAA0296	Zinc finger protein 646;Zinc finger protein 646, isoform CRA_a	200.8	Q8IVD8;O15015	PF00096	Zinc finger, C2H2 type	ENSG00000167395	16	+	30993263	0	0	1	7	4.5	6.0E-03
26	TOP1	DNA topoisomerase 1;DNA topoisomerase I	90.7	P11387;A8K4Z8;Q6NW25;Q6PK93;Q5	PF01028;PF02919	Eukaryotic DNA topoisomerase 1, catalytic core;Eukaryot	ENSG00000198900	20	+	39090876	61	64	124	174	4.5	6.5E-03
27	SPTAN1;SPTA2	Uncharacterized protein SPTAN1;Spectrin alpha chain, brain;Spectrin, non-erythroid alpha ch	286.0	A6NGS1;A6NNB8;Q13813;Q9UG16;Q	PF00036;PF08726;PF00018;PF00435	EF hand;Ca2+ insensitive EF hand;SH3 domain;Spectrin re	ENSG00000197694	9	+	130354690	1	0	1	3	4.1	9.2E-03
28	RPS2S	40S ribosomal protein S2S	13.7	P62851;A6NL90;A6NP42;A8MVS9	PF03297	S2S ribosomal protein	ENSG00000118181	11	-	118391635	2	1	3	5	4.1	9.7E-03
29	TCOF1	Treacle protein;Treacher Collins syndrome protein;TCOF1 protein	155.9	Q13428;Q59FZ2;Q9UFD4;A0JLU0	PF03546	Treacher Collins syndrome protein Treacle	ENSG00000070814	5	+	149717428	0	2	4	18	4.0	1.0E-02
30	RPL11;RP11-223J15.3-005;RP11-223J15.3-004	60S ribosomal protein L11;CLL-associated antigen KW-12;Ribosomal protein L11	20.3	P62913;Q3B5S8;Q5VVO0;Q5VVC8;Q5	PF00281;PF00673	Ribosomal protein L5;ribosomal L5P family C-terminus	ENSG00000142676	1	+	23890881	1	4	1	2	4.0	1.1E-02
31	BXDC2;BRIX	Brix domain-containing protein 2;Ribosome biogenesis protein Brix	41.4	Q8TDN6;A0JLQ5;A8K0P5;Q9NUW4	PF04427	Brix domain	ENSG00000113460	5	+	34951238	1	1	6	16	4.0	1.1E-02
32	PDCD11;KIAA0185	Protein RRP5 homolog;Programmed cell death protein 11	208.7	Q14690;Q2TA92;Q5W093;Q6PJ13	PF00575	S1 RNA binding domain	ENSG00000148843	10	+	105146402	1	2	8	22	3.8	1.4E-02
33	MMTAG2;C1orf35	Multiple myeloma tumor-associated protein 2;hMMTAG2	29.4	Q9BU76			ENSG00000143793	1	-	226355052	0	1	1	3	3.6	1.6E-02
34	RPS4X;CCG2;RPS4;SCAR;RPS4Y2;RPS4Y2P	40S ribosomal protein S4, X isoform;Single copy abundant mRNA protein;SCR10;40S ribosom	29.6	P62701;Q53HV1;Q96IR1;Q8TD47;A6N	PF00467;PF00900;PF08071;PF01479	KOW motif;Ribosomal family S4e;RS4NT (NUC023) dom	ENSG00000198034;ENSG0	X	-	71408650	0	1	1	3	3.6	1.7E-02
35	LYAR;PNAS-5	Cell growth-regulating nucleolar protein	43.6	Q9NXS8;A8K3Y5	PF08790	LYAR-type C2HC zinc finger	ENSG00000145220	4	-	4320329	25	25	49	83	3.6	1.7E-02
36	CCDC137	Coiled-coil domain-containing protein 137	33.2	Q6PKX4			ENSG00000185298	17	+	77244191	3	7	8	20	3.5	1.8E-02
37	ISG20L2;HSD38;HSD-38	Interferon-stimulated 20 kDa exonuclease-like 2	39.2	Q9H9L3	PF00929	Exonuclease	ENSG00000143319	1	-	154959019	0	0	5	10	3.5	1.8E-02
38	RRP1B;KIAA0179	RRP1-like protein B	84.4	Q14684;Q6PJM8	PF05997	Nucleolar protein,Nop52	ENSG00000160208	21	+	43903860	2	6	9	40	3.4	2.1E-02
39	BAG2	BAG family molecular chaperone regulator 2;Bcl-2-associated athanogene 2;BAG-2	23.8	O95816	PF02179	BAG domain	ENSG00000112208	6	+	57145083	0	0	1	1	3.4	2.2E-02
40	PURB	Transcriptional activator protein Pur-beta;Purine-rich element-binding protein B	33.2	Q96QR8;A4D2L7	PF04845	PurA ssDNA and RNA-binding protein	ENSG00000146676	7	-	44889059	3	5	5	9	3.4	2.2E-02
41	ZBTB48;HKR3;RP11-58A11.4-004;RP11-58A11.4-005	Zinc finger and BTB domain-containing protein 48;Krueppel-related zinc finger protein 3;Prot	77.1	P10074;A8K291;Q6LCP1;Q5SY20;Q5S	PF00651;PF00096	BTB/POZ domain;Zinc finger, C2H2 type	ENSG00000204859	1	+	6562698	0	0	3	3	3.4	2.2E-02
42	MECP2	Methyl-CpG-binding protein 2;MeCP-2 protein;MeCP2;Uncharacterized protein MECP2	53.3	P51608;A8K079;A8MW75	PF02178;PF01429	AT hook motif;Methyl-CpG binding domain	ENSG00000169057	X	-	152940218	1	5	15	22	3.3	2.3E-02
43	RBMX;HNRPG;RBMXP1;CCBL2;DKFZp547N1117;RBMXL1;RP4-531M19.2-001;hCG_23341;R	Heterogeneous nuclear ribonucleoprotein G;hnRNP G;RNA-binding motif protein, X chromos	42.3	P38159;Q8N8Y7;Q7VIN3;Q96E39;Q5T	PF08081;PF00076	RBM1CTR (NUC064) family;RNA recognition motif. (a.k.a	ENSG00000147274;ENSG0	X	-	135783288	0	0	2	1	3.3	2.3E-02
44	WRN;RECQ3;RECQL2	Werner syndrome ATP-dependent helicase	162.5	Q14191;A1KYY9;Q59F09	PF01612;PF00270;PF00271;PF00570	3'-5' exonuclease;DEAD/DEAH box helicase;Helicase cons	ENSG00000165392	8	+	31010320	6	2	15	16	3.3	2.4E-02
45	CSDA;DBPA	DNA-binding protein A;Cold shock domain-containing protein A;Single-strand DNA-binding p	40.1	P16989;Q59EB5	PF00313	'Cold-shock' DNA-binding domain	ENSG00000060138	12	-	10742956	1	0	7	7	3.1	3.0E-02
46	MLF1IP;CENPU;ICEN24;KLIP1;PBIP1	Centromere protein U;CENP-U;CENP-U(50);CENP-50;Interphase centromere complex protein	47.5	Q71F23;A2RRD9;A5D8X7;A8K8D2;Q9	PF04641;Q9GN2		ENSG00000151725	4	-	185852232	0	1	1	1	3.1	3.1E-02
47	RPL10A;NEDD6	60S ribosomal protein L10a;CSA-19;Protein NEDD6;Neural precursor cell expressed developm	25.0	P62906;Q1JQ76;A6NGU2	PF00687	Ribosomal protein L1p/L10e family	ENSG00000198755	6	+	35544156	10	4	3	5	3.0	3.5E-02
48	CCDC59;BR22;TAP26;HSPC128	Thyroid transcription factor 1-associated protein 26;TTF-1-associated protein 26;TTF-1-assoc	28.7	Q9P031			ENSG00000133773	12	-	81270761	1	1	1	2	3.0	3.5E-02
49	UBTF;UBF;UBF1	Nuclear transcription factor 1;Upstream-binding factor 1;UBF-1;Autoantigen NOR-90;Ribos	89.4	P17480;A8K6R8;A8K962;Q05BZ1;Q9B	PF00505	HMG (high mobility group) box	ENSG00000108312	17	-	39637928	1	0	2	11	3.0	3.6E-02
50	BLM;RECQ2;RECQL3	Bloom syndrome protein;RecQ protein-like 3;DNA helicase, RecQ-like type 2	159.0	P54132;Q3B7X0	PF00270;PF00271;PF00570	BDHCT (NUC031) domain;DEAD/DEAH box helicase;Helic	ENSG00000197299	15	+	89061583	12	2	12	24	2.9	4.1E-02
51	RPS3A;MFTL;hCG_33299	40S ribosomal protein S3a;Uncharacterized protein ENSP00000343748;HCG33299, isoform	29.9	P61247;A8K4W0;Q6NXR8;A6NCR2	PF01015	Ribosomal S3Ae family	ENSG00000145425;ENSG0	4	+	152240204	6	7	5	6	2.8	4.2E-02
52	PARP1;ADPRT;PPOL	Poly [ADP-ribose] polymerase 1;PARP-1;ADPRT;NAD(+) ADP-ribosyltransferase 1;Poly(ADP-rib	113.1	P09874;Q05D33;Q5VX86;Q6PJ10;Q9	PF00533;PF08063;PF00644;PF02877;PF0	BRCA1 C Terminus (BRCT) domain;PADR1 (NUC008) dom	ENSG00000143799	1	-	224615016	123	119	200	304	2.8	4.2E-02
53	DDX18	ATP-dependent RNA helicase DDX18;DEAD box protein 18;Myc-regulated DEAD box protein;N	75.4	Q9NVP1;Q4ZG72;Q53TI6;Q8N254	PF00270;PF00271	DEAD/DEAH box helicase;Helicase conserved C-terminal d	ENSG00000088205	2	+	118288725	13	16	27	34	2.7	4.7E-02

Table S2. Erythroid-related traits used to search for MAZ variants in the GeneATLAS database

ID of the trait	Description of the trait	Category
selfReported_n_1504	anaemia	Binary
clinical_c_D50	D50 Iron deficiency anaemia	Binary
clinical_c_Block_D50-D53	D50-D53 Nutritional anaemias	Binary
clinical_c_D51	D51 Vitamin B12 deficiency anaemia	Binary
clinical_c_Block_D60-D64	D60-D64 Aplastic and other anaemias	Binary
clinical_c_D63	D63 Anaemia in chronic diseases classified elsewhere	Binary
clinical_c_D64	D64 Other anaemias	Binary
clinical_c_Block_D70-D77	D70-D77 Other diseases of blood and blood-forming organs	Binary
clinical_c_D75	D75 Other diseases of blood and blood-forming organs	Binary
30030-0.0	Haematocrit percentage	Non-Binary
selfReported_n_1502	haematology	Binary
30020-0.0	Haemoglobin concentration	Non-Binary
30300-0.0	High light scatter reticulocyte count	Non-Binary
30290-0.0	High light scatter reticulocyte percentage	Non-Binary
30280-0.0	Immature reticulocyte fraction	Non-Binary
30050-0.0	Mean corpuscular haemoglobin	Non-Binary
30060-0.0	Mean corpuscular haemoglobin concentration	Non-Binary
30260-0.0	Mean reticulocyte volume	Non-Binary
30270-0.0	Mean sphered cell volume	Non-Binary

30170-0.0	Nucleated red blood cell count	Non-Binary
30230-0.0	Nucleated red blood cell percentage	Non-Binary
30010-0.0	Red blood cell (erythrocyte) count	Non-Binary
30070-0.0	Red blood cell (erythrocyte) distribution width	Non-Binary
30250-0.0	Reticulocyte count	Non-Binary
30240-0.0	Reticulocyte percentage	Non-Binary

SupplementalTable 3: Datasets used in the study

Cell type	Data set	Crosslinking/ Reference	Mapped reads	GEO accession
Erythroblasts	MAZ	EGS + CH_2O	12 058 863	GSE139281
Erythroblasts, default input data set	Input 1	8	13 158 630	GSE114084
Erythroblasts	ATAC-Seq	17	70 014 192	GSE74912
Erythroblasts	H3K4me3	18	5 349 306	GSE36985
Erythroblasts	H3K27ac	19	56 439 919	GSE70660
Erythroblasts	Pol2	18	97 237 763	GSE36985
Erythroblasts	H3K27me3	18	4 990 983	GSE36985
Erythroblasts	H3K4me1	19	27 531 447	GSE70660
Erythroblasts	GATA1	19	118 938 490	GSE36985
Erythroblasts	CTCF	20	59 309 750	GSE102184
HepG2 (hepatocellular carcinoma)	MAZ	21	51 465 720	GSE31477
A549 (lung carcinoma)	MAZ	21	39 848 427	GSE91939
GM12878 (B lymphoblastoid)	MAZ	21	40 269 745	GSE106046
MCF-7 (mammary adenocarcinoma)	MAZ	21	128 288 726	GSE91633
IMR90 (fetal lung fibroblast)	MAZ	21	34 117 266	GSE31477

Suppl Table 4. Tissue specific MAZ peaks in ENCODE human cell lines and human primary erythroid cells.

Cell line	Unique peaks	Total number of peaks	% unique peaks
HepG2	3636	15481	23.5%
Gm12878	9480	23951	39.6%
A549	335	4323	7.7%
MCF-7	6512	20419	31.9%
IMR90	11383	28208	40.4%
Ery	1780	10088	17.6%

Supplemental Table 5. Genomic loci associated with erythroid traits and erythropoiesis which are bound by MAZ

Genomic locus	*Erythroid traits*	*Location of the MAZ peak*	*Position of the MAZ peak*	*Fold change relative to the input*	*log_{10} q-value of the MAZ peak*
ITFG3(FAM234A)	Hb, MCH, MCHC, MCV, RBC	ITFG3 promoter	chr16:234379-235136	36.8	81.46
HBS1L-MYB	Hct, MCH, MCHC, MCV, RBC	HBSL1 promoter, MYB first intron	chr6:135054682-135055128 chr6:135183043-135183281	26.95 7.15	57.41 8.84
G6PD	Hct, Hb, MCV, RBC, RDW	G6PD promoter, G6PD second intron	chrX:154547246-154547697 chrX:154541913-154542409	13.49 16.38	21.39 29.99
CCND3	MCH, MCV, RBC	CCND3 promoter, CCND3 first intron, CCND3 second intron	chr6:42045709-42046059 chr6:42030970-42031193 chr6:42017103-42017413 chr6:41941500-41942068 chr6:41938239-41938465	20.64 9.47 18.51 35.9 6.48	38.02 12.15 32.58 78.1 6.64
TFR2	Hct, MCV, RBC	TFR2 promoter	chr7:100641422-100642132	78.49	205.83
TFRC	MCH, MCV	TFRC promoter	chr8:20197105-20197542	13.49	21.39
PDGFRA-KIT	MCV, RBC	KIT promoter	chr4:54657841-54658018	6.68	7.18
CITED2	MCH, MCV	CITED2 promoter	chr6:139374652-139374972 chr6:139375180-139375707	6.16 8.01	5.72 9.22
HMOX2	MCH, MCV	HMOX2 promoter	chr16:4476227-4476485	9.47	12.15

PRKCE	Hct	PRKCE first intron, PRKCE second intron	chr2:45650249-45650426 chr2:45778683-45778969	6.01 6.55	6.17 6.46
CD164	MCH	CD164 promoter	chr6:109381892-109382875	7.91	9.22
SH2B3	Hb	SH2B3 promoter, SH2B3 first intron, SH2B3 second intron, SH2B3 third intron	chr12:111405191-111405978 chr12:111406092-111406300 chr12:111406596-111406773 chr12:111429666-111429889 chr12:111437838-111438136	13.18 8.68 7.91 6.17 6.55	21.95 10.7 9.22 5.88 6.46
CDT1	MCHC	CDT1 promoter	chr16:88803519-88804082	22.76	43.54
FBXO7	MCV, MCH	FBXO7 promoter	chr22:32474726-32475349	6.33	6.25
RPS19	Hb, RBC, MCV	RPS19 promoter	chr19:41859549-41860329	10.86	18.19
EPOR	RBC, Hb	EPOR promoter, EPOR first intron	chr19:11384246-11384607 chr19:11383132-11383772	21.92 19.35	41.42 35.01
KLF1	MCH	KLF1 promoter, KLF1 first intron, KLF1 second intron	chr19:12887097-12887963 chr19:12886160-12886497 chr19:12885126-12885353	17.2 14.2 6.01	29.7 22.61 5.55
GATA1	RBC	GATA1 promoter	chrX:48786184-48786580	11.57	16.63
SBDS	Hb	SBDS promoter	chr7:66995478-66995773	12.9	19.59
CDAN1	Hb	CDAN1 promoter	chr15:42736973-42737392	14.33	23.29
PTPLAD1(HACD3)	MCH	PTPLAD1 promoter	chr15:65530326-65530535	7.7	8.62

Table S6 Sequences of oligonucleotides used in this study

EMSA Probe	Sequence
1/2	CCCAAGCATAAACCCTGG
3/4	GGCCGGGCGTGCCCCCGC
5/6	GAGCGCCGCCCGGCCGGG
7/8	CGCCAGCCAATGAGCGCC
9/10	CCGGGCTCCGCGCCAGC
11/12	CCAGGCCGCGCCCCGGGC
13/14	CAGGCCCCGCCCGGGACT
PCR Primers	**Sequence**
HBA Fw	GAGGCCCTGGAGAGGATGTTCC
HBA Rev	ACAGCGCGTTGGGCATGTCGTC
HBB Fw	GCTCACCTGGACAACCTCA
HBB Rev	CGTTGCCCAGGAGCCTGAA
HBE1 Fw	GGGCAGACTCCTCGTTGTT
HBE1 Rev	GCCTTGACCTTGGGGTTG
HBG Fw	TGGGTCATTTCACAGAGGAG
HBG Rev	TAGACAACCAGGAGCCTTCC
MAZ Fw	TGTGAGAAATGTGAGGCAGC
MAZ Rev	GCCGAGCTCAGCATCTTG
PABPC1 Fw	AGCTGTTCCCAACCCTGTAATC
PABPC1 Rev	GGATAGTATGCAGCACGGTTCTG
shRNAs	**Sequence**
TRCN0000235699	CCGGGATGCTGAGCTCGGCTTATATCTCGAGATATAAGCCGAGCTCAGCATCTTTTTG
TRCN0000235703	CCGGTCTGTGAGCTCTGCAACAAAGCTCGAGCTTTGTTGCAGAGCTCACAGATTTTTG