Identification of the transcription factor MAZ as a regulator of erythropoiesis

blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

Identification of the transcription factor MAZ as a regulator of erythropoiesis

Tracking no: ADV-2021-004609R2

Douglas VERNIMMEN (University of Edinburgh, United Kingdom) David Garrick (Institut de Recherche Saint Louis, France) Darya Deen (University of Edinburgh, United Kingdom) Falk Butter (Institute of Molecular Biology, Germany) Michelle Holland (Department of Medical and Molecular Genetics, United Kingdom) Vasiliki Samara (MRC Molecular Haematology Unit, United Kingdom) Jacqueline Sloane-Stanley (MRC Molecular Haematology Unit, United Kingdom) Helena Ayyub (MRC Molecular Haematology Unit, United Kingdom) Matthias Mann (Max-Planck Institute for Biochemistry, Germany) Jan Frayne (University of Bristol, United Kingdom) Deborah Daniels (University of Bristol, United Kingdom) Ivan Ferrer-Vicens (University of Bristol, United Kingdom) Daniel Ferguson (University of Bristol, United Kingdom)

Abstract:
Erythropoiesis requires a combination of ubiquitous and tissue-specific transcription factors. Here, through DNA affinity purification followed by mass spectrometry, we have identified the widely expressed protein MAZ (Myc-associated zinc finger) as a transcription factor that binds to the promoter of the erythroid-specific human α-globin gene. Genome-wide mapping in primary human erythroid cells revealed that MAZ also occupies active promoters as well as GATA1-bound enhancer elements of key erythroid genes. Consistent with an important role during erythropoiesis, knockdown of MAZ reduces α-globin expression in K562 cells and impairs erythroid differentiation in primary human erythroid cells. Genetic variants in the MAZ locus are associated with changes in clinically important human erythroid traits. Taken together, these findings reveal the Zinc-finger transcription factor MAZ to be a previously unrecognised regulator of the erythroid differentiation program.

Conflict of interest: No COI declared

COI notes:

Preprint server: Yes; BioRXiv https://doi.org/10.1101/2020.05.10.087254

Author contributions and disclosures: D.G. and D.V. developed the hypothesis; F.B., D.E.D., I.F-V. D.C.J.F., M.L.H., V.S., J.A.S.S., H.A., and D.G. performed experiments and/or collected data; D.D., F.B., D.E.D., I.F-V., D.C.J.F., M.M., J.F., D.G. and D.V. analyzed data; interpreted data; and D.D., J.F., D.G., and D.V. wrote the manuscript, which was revised and approved by all authors.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: GEO Submission GSE139281

Clinical trial registration information (if any):

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

Abstract word count: 131

Characters (including space): 35,848

Words count: 5,371

Figures: 6

Table: 1

Suppl Figures: 11

Suppl Tables: 6

References: 58

Running title: MAZ regulates erythropoiesis

Key words: Erythropoiesis, Chromatin, Gene regulation, MAZ, Proteomics

KEY POINTS

A mass-spectrometry screen identified Myc-associated zinc finger (MAZ) as a transcription factor binding to the human α-globin promoter

In erythroid cells, MAZ targets many erythroid-specific regulatory elements and knockdown of MAZ compromises erythropoiesis

ABSTRACT

Erythropoiesis requires a combination of ubiquitous and tissue-specific transcription factors. Here, through DNA affinity purification followed by mass spectrometry, we have identified the widely expressed protein MAZ (Myc-associated zinc finger) as a transcription factor that binds to the promoter of the erythroid-specific human α-globin gene. Genome-wide mapping in primary human erythroid cells revealed that MAZ also occupies active promoters as well as GATA1-bound enhancer elements of key erythroid genes. Consistent with an important role during erythropoiesis, knockdown of MAZ reduces α-globin expression in K562 cells and impairs erythroid differentiation in primary human erythroid cells. Genetic variants in the *MAZ* locus are associated with changes in clinically important human erythroid traits. Taken together, these findings reveal the Zinc-finger transcription factor MAZ to be a previously unrecognised regulator of the erythroid differentiation program.

INTRODUCTION

For over four decades, studies of the α- and β-globin gene clusters have contributed to our understanding of some of the fundamental principles of mammalian gene regulation, including RNA stability, termination of transcription, and the identification of remote regulatory elements[1]. A complex network of DNA sequence elements, chromatin accessibility, histone modifications, and transcription factor (TF) occupancy orchestrates expression of globin genes, which are exclusively expressed in erythroid cells [2]. Both proximal regulatory regions (promoters) and distal regulatory elements (enhancers) are required for the initiation of α-globin (*HBA*) and β-globin (*HBB*) gene expression in erythroid cells. When active, these regulatory elements are characterised by the presence of open chromatin regions and active histone modifications. A small group of lineage-restricted TFs including GATA binding protein 1 (GATA1), T cell acute lymphocytic leukemia 1 protein (TAL1), and Erythroid Krüppel-like factor (EKLF; henceforth referred to as KLF1) act as erythroid 'master regulators [3], by binding to both promoters and enhancers of the globin genes as well as to other genes important for erythropoiesis, [reviewed in [2,4,5] where they usually act together with other widely-expressed TFs [6-14]. However, despite the enormous advances that have been achieved in our understanding of the molecular mechanisms controlling globin gene expression, this research area keeps progressing.

Here, we have carried out an unbiased screen for proteins that bind to adjacent GC-rich motifs in the promoter of the duplicated α-globin genes (HBA2 and HBA1) in human erythroid cells, combining electrophoretic mobility shift assays (EMSAs), DNA affinity purification, and mass spectrometry. This screen identified Myc-Associated Zinc-finger protein (MAZ) as a direct binder of the promoter region of the human α-globin gene *in vitro* and *in vivo.* Knockdown of MAZ in primary human erythroid cells

led to impaired erythroid differentiation and a reduction in globin expression. Moreover, variants in the MAZ gene and its promoter region are associated with changes in red blood cell parameters and erythroid traits. ChIP-seq experiments in primary human erythroblasts revealed that MAZ recognises a canonical $G_3(C/A)G_4$ binding motif and is enriched at transcription start sites (TSSs) of transcriptionally active genes and distal regulatory elements. Erythroid-specific MAZ signal is enriched at promoters of genes associated with erythropoietic disorders. We found that MAZ erythroid-specific binding sites frequently colocalize with GATA1 particularly at enhancer elements, suggesting functional synergy between these two transcription factors. Together our findings have identified MAZ as an important regulator of the erythroid differentiation program.

MATERIALS AND METHODS

Cell lines and primary erythroid cells

Primary erythroid cells and EBV-infected B-lymphoblasts were obtained as previously described [15]. Cell lines (K562, HT29, HeLa, SW13 and COS7) were cultured in RPMI 1640 supplemented with 10% FBS. For siRNA-mediated knockdown of MAZ, K562 cells were transfected with MAZ SMARTPool siRNA or Non-targeting control pool siRNA (Dharmacon) using Lipofectamine according to the manufacturer's instructions, and cells were harvested after 4 days. For overexpression of MAZ, COS7 cells were transfected with the plasmid pcDSAF-1, which expresses full length MAZ (SAF-1) cDNA under the control of the CMV promoter [16]. Cells were transfected using Lipofectamine according to the manufacturer's instructions and were harvested after 30 hours. Adult PBMCs were isolated from LRS cones, with informed consent from all donors, and used in accordance with the Declaration of Helsinki and approved by the

National Health Service National Research Ethics Committee (reference number 08/H0102/26) and the Bristol Research Ethics Committee (reference 12/SW/0199).

Electrophoretic mobility-shift assay (EMSA)

Nuclear extracts were performed as described [17] except that nuclei were incubated in buffer C for 30 min. Protein concentration was determined using the Qubit Protein Assay Kit (ThermoFisher). Oligonucleotide probes (Suppl Table 6) were designed to have an additional 5'GG overhang on each end after annealing and were labelled by filling in with the large Klenow fragment of DNA Polymerase I in the presence of [α-^{32}P]-dCTP as described by the manufacturer (New England Biolabs). For gel shift reactions, 5 µg of nuclear extract was incubated with 1 ng radiolabelled probe (>10,000 cpm) in buffer [10 mM HEPES (pH 7.8), 50 mM potassium glutamate, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM dithiothreitol, 0.5 µg of poly(dI-dC), 10 µg of bovine serum albumin, 1 mM $ZnSO_4$, 5% glycerol] for 30 min on ice. For competition experiments, unlabelled competitor oligonucleotides were added to the binding reactions at 100 X molar excess. For gel supershift experiments, antibodies were incubated with nuclear extract for 20 min on ice prior to the addition of the radiolabelled probe. Antibodies used in supershift experiments were as described in previous ChIP studies [18]. Following binding, samples were subjected to electrophoresis at 4°C for 2.5 h at 12 V/cm on a native polyacrylamide gel (6% [19:1] *bis*:acrylamide in 0.5 X Tris-borate-EDTA). The gels were dried and analysed using a Storm 860 Molecular Imager (Molecular Dynamics Ltd).

shRNA knockdown in primary erythroid cultures

Lentiviral constructs expressing shRNAs targeting non-overlapping sites in Exon 4 of MAZ (clones TRCN0000235699 and TRCN0000235703) or non-targeting scrambled shRNAs were obtained from Merck. These constructs express the shRNA from the

human U6 promoter and the Puro^r gene from the hPGK promoter. Lentiviral preparations were prepared using standard techniques by co-transfection of 293T cells with the lentiviral construct together with psPax2 and pMD2.G using calcium phosphate mediated transfection. Culture supernatants were harvested at 48 h after transfection and concentrated by ultracentrifugation using standard techniques. For lentiviral-mediated knockdown experiments, primary erythroid differentiation cultures were carried out essentially as described in Leberbauer et al., [19] with the following modifications: peripheral blood mononuclear cells were cultured for 8 days in erythroblast expansion medium (StemSpan™ SFEM medium (Stem Cell Technologies) supplemented with Epo (2 U/mL, Roche), IGF-1 (40 ng/mL, R&D systems), SCF (100 ng/mL, R&D Systems), dexamethasone (1 µM, Sigma) and cholesterol-rich lipids (40 µg/mL, Sigma). Cells were infected with concentrated lentivirus for 24 h in expansion medium supplemented with hexadimethrine bromide (8 µg/ml). The cells were seeded in fresh expansion medium for 24 h prior to selection with puromycin (2 µg/ml) for 48 h. After removing Puromycin, cells were cultured for a further 3 days in expansion medium prior to harvesting for flow cytometry, RNA and protein (total 15 days in culture).

Mass Spectrometry

For proteomics analysis, the DNA baits corresponding to oligonucleotide 13/14 WT and 13/14 M3 with a TT/AA-overhang were annealed, phosphorylated, ligated and purified as previously described[20]. The desthiobiotinylated oligonucleotides were coupled to Streptavidin Dynabeads MyOne C1 (Life Technologies) for 60 min at room temperature and excess oligonucleotides removed by washing. The oligonucleotide coupled streptavidin beads were incubated with 250 µg of K562 nuclear extract diluted

in PBB buffer (150 mM NaCl, 50 mM Tris–HCl pH 8.0, 10 mM $MgCl_2$, 0.5 % Igepal CA630, Complete Protease Inhibitor without EDTA (Roche)) for 2 hours at 4°C under slight agitation. Non-bound proteins were removed by washing three times with PBB buffer and the bound DNA-protein complexes liberated from the streptavidin beads with 200 µl of a 16 mM biotin/50 mM ABC (pH 8.0) solution. The complexes were precipitated with pure ethanol overnight at room temperature. The pellet was resuspended in 50 µl 8M urea and first digested with LysC (Wako) for 3 hours and subsequently diluted to 250 µl with 50mM ABC buffer and digested with 200 ng trypsin overnight, both at room temperature. The tryptic peptides were loaded onto an SCX StageTip[21] and stored until the MS measurement.

The tryptic peptides were fractionated on an in-house packed 25 cm microcapillary column (Reprosil 3.0 µm; Dr. Maisch GmbH) in a 125 minute gradient from 5 to 60 % acetonitrile. The MS measurement was performed on an LTQ Orbitrap XL (Thermo) with a top5 DDA method using CID fragmentation. The MS data was processed with MaxQuant[22] version 1.0.12.5 using the human Uniprot database.

Expression analysis

RNA was extracted with Tri reagent (Sigma) according to the manufacturer's instructions. RNAs were DNaseI treated (Ambion) and cDNA was generated with SuperScript III (Invitrogen) as previously described [23]. Primers used are listed in Suppl Table 6.

Antibodies

Anti-MAZ antibody was from Bethyl Laboratories (A301-652A), Alpha globin (sc-514378) (Santa Cruz) and Beta globin (sc-21757) from Santa Cruz Biotechnology, and Beta actin (A1978) from Sigma. The following antibodies were used for flow cytometry: CD71 (sc-7327) from Santa Cruz Biotechnology (primary) and IgG1-APC (406610)

from Biolegend (secondary), and GPA (130-120-473) PE-conjugated antibody from Miltenyi Biotec.

ChIP-qPCR and ChIP-Seq assay

ChIP was performed as previously described [24]. Briefly, chromatin was first cross-linked with ethylene glycol bis(succinimidyl succinate) (EGS) in PBS at a final concentration of 2 mM for 60 min at RT. Formaldehyde (CH_2O) was then added at a final concentration of 1 % for 15 min at RT and samples were sonicated over 20 min (10 x 30 s pulses) at 4°C to cleave genomic DNA (Bioruptor, Diagenode). Anti-MAZ antibody was from Bethyl Laboratories (A301-652A). Each ChIP was performed as two independent experiments and quality was assessed by qPCR using primers and probes (5'FAM-3'TAMRA) described in [18]. The ChIP-seq libraries were prepared using the New England Biolabs NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina according to the manufacturer's protocol and starting with between 6 and 40 ng of captured DNA. All libraries received 12 cycles of PCR amplification in the final step before PCR clean-up using Ampure beads. Sequencing was carried out on the HiSeq 2500 using Illumina HiSeq Rapid Cluster Kit v2 - Paired-End for 100 cycles. RTA software version used was Illumina RTA 1.17.20 and the Pipeline software version was bcl2fastq-1.8.3. Analysis of ChIP-seq is described in Supplemental material.

RESULTS

A novel DNA-protein complex occupies the distal promoter region of the human HBA genes.

Transcriptional activation of the HBA genes during erythroid differentiation is associated with localised relaxation of chromatin structure, which is observed experimentally as a chromatin accessible region immediately upstream (promoter region) of the adult HBA genes (chr16: 172k and 176k, Suppl. Figure 1A) specifically in erythroid cells[25]. In order to map chromatin accessibility within this region at a higher resolution than provided by DNase- and ATAC-seq approaches, intact nuclei isolated from cells expressing α-globin (the erythroid cell line K562) and not expressing α-globin (EBV-transformed B lymphoblast cell line) were digested with low concentrations of selected restriction enzymes. The extent of digestion reflects accessibility of the specific recognition sites and was assessed by Southern blotting (Suppl. Figure 1B). This assay revealed that the region of erythroid-specific sensitivity extends from -220 bp (*Fok*I site) to +35 bp (*Nco*I site) relative to the TSS of the HBA genes (Suppl. Figure 1B).

In order to characterise protein binding across this hypersensitive region, we carried out electrophoretic mobility shift assays (EMSAs) using a series of overlapping oligonucleotide probes to compare *in vitro* binding proteins in nuclear extracts prepared from the same erythroid (K562) and non-erythroid (EBV) cells. Oligonucleotide probes were specifically placed at positions suggestive of protein binding by *in vivo* dimethyl-sulfate (DMS) footprinting assays reported previously (Figure 1A, Table 1) [26].

We observed band shifts for five out of seven probes used (Figure 1B). Previous motif analysis had suggested potential binding motifs for Krüppel-like (Sp/X-KLF) family proteins (-121/-116, -61/-56), NFY (-72/-65), Nuclear Factor I (NF1) (-85/-72), and α-inverted repeat protein (α-IRP) (-51/-42) [26] (Table 1). Gel supershift assays, in which nuclear extracts were pre-incubated with anti-NFY or anti-NF1 antibodies, show retarded bands with probes 7/8 and 9/10 respectively, confirming that these sites are indeed bound by NFY and NF1 (Figure 1B, lanes 13 and 17, respectively). Interestingly, complex binding patterns comprised of four shifted bands were observed at two neighbouring GC boxes situated at -100/-83 (probe 11/12) and -128/-111 (probe 13/14). These GC boxes have a highly similar sequence (Suppl. Figure 2A) and were able to cross-compete with each other in competition assays (Suppl. Figure 2B), suggesting that they are likely to be bound by the same proteins. Interestingly, while the most heavily retarded bands (labelled a, b and c in Figure 1B) were detected in both K562 and EBV nuclear extracts, the faster migrating species (d) was observed only in K562 cells. This factor bound more strongly to probe 13/14 than to 11/12 (Figure 1B, compare lanes 20 and 23), a finding that was confirmed in competition assays (Suppl. Figure 2B). Gel shifts carried out with nuclear extracts from other cell types revealed that species (d) was the most abundant complex in primary human erythroblasts but was either not detected or weak in a range of non-erythroid cells types (HT-29, SW13, HepG2, and HeLa) (Figure 1C).

The probe 13/14 (covering -128/-111 region) contains a predicted binding motif for the Krüppel-like zinc-finger transcription factor SP1 [26] (Table 1). We found that the binding of all four proteins to probe 13/14 was sensitive to the presence of the zinc-chelating agent EDTA (Suppl. Figure 3A), consistent with Zn-finger-dependent binding. In contrast, binding of the non-Zn-finger protein NFY to probe 7/8 was not strongly

affected by EDTA (Suppl. Figure 3A). To identify the protein responsible for these bands on probe 13/14, we carried out gel supershift reactions using antibodies against candidate Zn-finger proteins (SP1, SP3, KLF1, and Krüppel-like factor 3 (KLF3, also known as BKLF)). These experiments revealed species (a), (b) and (c) to be SP1, SP3 and KLF3, respectively (Figure 1D). These findings are consistent with species (a), (b) and (c) being observed in both erythroid and non-erythroid nuclear extracts (Figure 1C), since SP1, SP3 and KLF3 are all widely-expressed TFs and have previously been shown to regulate α-globin expression [10,18]. In contrast, the binding of species (d) to probe 13/14 was not strongly affected by any of the antibodies tested. Taken together, these experiments have revealed an erythroid-enriched complex at neighbouring sites in the -128/-90 region upstream of HBA genes.

A mass spectrometry-based screen identifies MYC-associated Zn-finger protein MAZ as a factor binding at the α-globin promoter

In order to identify the factor responsible for species (d), we carried out a protein affinity purification screen, using probe 13/14 as the bait (Figure 2A). As a negative control, we designed a mutant version of probe 13/14 (M3) in which a central guanidine was changed to thymidine, preventing formation of species (d) while also depleting binding of SP1, SP3 and KLF3 (Suppl. Figure 3B). To carry out this screen, K562 nuclear extracts were incubated with desthiobiotin-modified concatenated DNA probes, protein-DNA complexes were purified using streptavidin beads, and eluted proteins were subjected to mass spectrometry (Figure 2A)[20]. Of the proteins most significantly enriched for binding to the wild type relative to the mutant probe (Supplemental Table 1), the Myc-associated Zn-finger protein (MAZ) was the strongest candidate, exhibiting strong enrichment at the wild-type compared to the mutant probe and high tryptic

peptide counts in both wild-type replicates (Supplemental Table 1). MAZ is a C2H2-type Zinc finger protein which has been previously shown to bind *in vitro* to a G-rich consensus motif (G_3AG_3) in non-erythroid cells [27-29]. In order to investigate whether MAZ is indeed the protein responsible for species (d), we performed siRNA-mediated knockdown (KD) of MAZ in K562 cells and found that depletion of endogenous MAZ protein (Figure 2B, left panel) was associated with reduction of species (d) in the gel shift assays (Figure 2B, right panel). Conversely, expression of exogenous MAZ protein in COS7 cells was associated with strong enrichment of species (d) (Figure 2C). Binding of the over-expressed MAZ protein resulted in a depletion of the SP3 gel shift, indicating competition for binding under these *in vitro* conditions. Altogether, these experiments confirm that MAZ is indeed the protein responsible for species (d) and indicate that MAZ binds to neighbouring GC boxes at -100/-83 and -128/-111 within the human α-globin promoter *in vitro*.

In order to confirm binding by MAZ at the active HBA promoter *in vivo*, we carried out ChIP-qPCR experiments using a previously described series of qPCR amplicons throughout the HBA locus[18]. Whereas binding was absent or low in non-expressing cells (EBV lymphoblast), strong enrichment of MAZ was observed at the HBA promoter and gene body in cells expressing HBA, with the highest enrichment observed in primary erythroid cells (Figure 2D). In these primary cells, MAZ was also enriched at the important distal regulatory element of the α-globin locus (MCS–R2). We further investigated the dynamics of MAZ recruitment to the α-globin genes during erythroid maturation in primary erythroblast differentiation cultures (Suppl. Figure 4A-C). During differentiation, expression of the adult globin genes HBB and HBA is strongly upregulated, peaking in intermediate and late stage erythroblasts (Suppl. Figure 4D, see also [30]). Concomitantly with the strong upregulation of α-globin

expression, MAZ was dynamically recruited to the HBA promoter and genes in these cultures (Suppl. Figure 4E). Altogether, our results indicate that MAZ binds to the active human HBA locus both *in vitro* and *in vivo*.

MAZ is required for erythroid differentiation and is associated with changes in erythroid-related traits

To investigate the functional significance of MAZ during erythroid differentiation, we used two different shRNAs to deplete MAZ expression in differentiating primary human erythroid cultures (Figure 3). Flow cytometry analysis revealed an impaired differentiation of primary erythroid cultures infected with the MAZ shRNA lentivirus, with cells accumulating at the BFU-E stage ($CD71^{lo}GPA^{lo}$), and fewer cells reaching the stage of intermediate erythroblasts ($CD71^{+}GPA^{+}$) (Figure 3A, B, Suppl. Figure 5). This impaired differentiation could also be observed as a deficiency of haemoglobinisation (Figure 3C). In cells harvested at the end of the culture, knockdown of MAZ expression was confirmed by RT-qPCR and western blots and was associated with downregulation of expression of both α- and β-globin (Figure 3D, E, F) consistent with the impeded differentiation. In K562 cells (where differentiation is not a factor), shRNA-mediated knockdown of MAZ reduced expression of HBA relative to the HBG and HBE genes of the β-globin cluster (Suppl Figure 5C). Taken together with the EMSA and ChIP-~~seq~~ qPCR results, this finding is consistent with MAZ exerting a direct effect on α-globin expression which is independent of an effect on erythroid differentiation.

Though widely expressed across the haematopoietic system in both humans and mice, in both organisms MAZ is particularly elevated in the erythroid lineage, c~~C~~onsistent with an important role for MAZ in erythropoietic differentiation~~, MAZ is strongly expressed in the erythroid branch of the haematopoietic system in both~~

~~humans and mice~~ (Suppl. Figure 6). We further investigated whether human genetic variants in the *MAZ* gene are associated with clinically important erythroid traits. We analysed the GeneATLAS database that reports associations between 778 traits and millions of DNA variants [31]. Out of 25 erythroid-related traits in the GeneATLAS (Suppl. Table 2), changes in eight traits (32%) were associated with three variants in the MAZ gene or promoter region (rs11559000, rs572982482, rs72798129) (adjusted *p*-value <0.01, Figure 3G, see also Suppl. Figure 7A). Taken together, these findings indicate that MAZ is an important factor in the erythroid differentiation program and suggest that this locus contributes to important erythroid traits.

MAZ occupies TSS and binds directly to DNA through a $G_3(C/A)G_4$ motif

Following from our findings at the HBA locus, we expanded our analysis by investigating binding of MAZ genome-wide in primary human erythroid cells by carrying out ChIP-seq (Suppl. Figure 8A-B). As previously observed by ChIP-qPCR (Figure 2D), MAZ was strongly enriched at both HBA2/1 promoters as well as, to a weaker extent, at the MCS-2 remote regulatory element (Figure 4A, top). In contrast, very low binding of MAZ was detected at the promoter and regulatory elements (locus control region; LCR) of the β-globin gene cluster (Figure 4A, bottom). Genome-wide, we identified 10,088 MAZ binding sites in primary erythroid cells (Figure 4B, Suppl. Figure 8C). While the majority of MAZ binding sites (65%) were located at promoter regions (Figure 4B), the average MAZ enrichment was comparable for peaks in the promoter regions, intergenic, and genic regions (Suppl. Figure 8D). Overall, MAZ was present at least at one TSS of 28% (5466/19646) of protein-coding genes. Comparison with published ChIP- and ATAC-seq datasets from erythroid cells[25,32-34] (Suppl. Table 3) revealed that MAZ binding sites are located in regions of high chromatin accessibility (as detected by ATAC-seq) that are enriched for activating histone

modifications (H3K4me3, H3K27ac) and Pol II, but depleted of the repressive histone mark H3K27me3 (Figure 4C). ~~Consistent with a recent report that MAZ interacts physically with the CCCTC-binding factor (CTCF), an important regulator of genomic architecture [35], we also observed that some MAZ peaks (24%; 2448/10090) overlap with CTCF peaks mapped in primary erythroid cells [36]. The heatmap representation (Figure 4C) indicates clear enrichment of CTCF signal over MAZ peaks, although not directly correlating with the intensity of MAZ signal.~~

In order to identify motifs contributing to MAZ binding, we used the MEME software suite [35] to discover *de novo* enriched sequence motifs in a training set consisting of the 500 highest ranked MAZ peaks [36]. Overall, six enriched motifs were detected in the MAZ erythroblast training dataset (E-value <0.01) and were then investigated for their occurrence in all detected MAZ peaks (Figure 4D). The most significantly-enriched motif, $G_3(C/A)G_4$, is contained within the 13/14 probe derived from the HBA promoter used to identify MAZ in the MS screen, and was similar to the previously published MAZ motif G_3AG_3 [37, 38]. The canonical motif observed this study $G_3(C/A)G_4$ was present in 92% of all MAZ peaks (Figure 4D), and was the only enriched motif with a narrow unimodal central enrichment in the peaks and a large maximum site probability (Figure 4E), suggesting that the vast majority of MAZ genomic binding in erythroblasts is due to direct DNA binding of MAZ to this DNA sequence motif. In agreement with these observations, mutations within this core G_3CG_4 motif present in the 13/14 probe prevented MAZ binding *in vitro,* while mutations outside this core motif had little effect (Suppl. Figure 9). Furthermore, the 11/12 probe, which exhibits lower affinity for MAZ (Figure 1B and Sup Fig. 2B), contains a less-perfect version of this core motif than the 13/14 probe (Suppl. Figure 9). The second most abundant motif detected in our MAZ training set was a GGAGGA-containing motif. This motif was

present in 31% of MAZ peaks, but did not exhibit central localisation, suggesting it may contribute to MAZ binding in a cooperative manner. The enrichment of this motif is consistent with the previous reports demonstrating binding by MAZ to GGA repeats with a high propensity to form G4-quadruplexes [39-42].

Our gel shift results indicated that MAZ can bind the same probes within the α-globin promoter as other C2H2 zinc finger TFs. To investigate the potential for regulatory crosstalk between MAZ and other TFs more generally, we quantified the similarities between the identified MAZ position weight matrix (PWM) and known binding profiles of other TFs curated in the JASPAR database [43]. This analysis revealed that the binding motifs of five C2H2 zinc finger TFs (SP1, KLF16, KLF5, SP2, SP3) significantly match the derived MAZ motif (E-value $<1e^{-3}$) (Suppl. Figure 10A). KLF3, which binds to the same GC boxes within the α-globin promoter as MAZ (Figure 1) and has been previously described to bind the α-globin promoter[10], also shows a similar PWM. Mining of a published RNAseq dataset[44] revealed that, with the exception of KLF5, all of these factors are expressed in erythroblasts (Suppl Figure 10B). These findings suggest a potential for widespread regulatory crosstalk between MAZ and multiple other ZF TF in erythroid cells.

For MAZ and SP1, whose experimentally-derived PWM were most related (Suppl Figure 10C), we also compared these experimentally-derived matrices with PWM that are predicted based on inferred contact energies for the C2H2-ZF domains of these two proteins[45]. Interestingly, while the predicted PWM of SP1 was very similar to the experimentally derived matrix, the predicted PWM of MAZ was much more extended and variable than that observed experimentally in our ChIP-seq dataset, consistent with the presence of six Zn fingers in MAZ compared to three in SP1 (Suppl Figure

10C). These observations suggest that outside of an obligatory core subsequence G_3(C/A)G_4, detected in 92% of all MAZ peaks), MAZ binding in erythroid cells is less determined by primary DNA sequence than predicted, and are consistent with co-associated TFs or secondary DNA structure playing an important role in determining MAZ binding.

Erythroid-specific MAZ binding sites are associated with the promoters of key erythropoiesis genes

To gain further insight into the specific role of MAZ during erythroid differentiation, we compared the genome-wide profile of MAZ binding in primary erythroblasts with MAZ ChIP-seq profiles from five human non-erythroid cell lines (HepG2, GM12848, MCF-7, IMR90, A549) generated by the ENCODE consortium [46,47] (Suppl. Table 3). Comparison of MAZ peaks identified in these six cell types revealed that, as expected for a housekeeping TF, a large proportion of MAZ peaks are shared between at least two different cell types (so called “common” peaks, (n=8310) (Figure 5A). However, between 8 and 40% of the MAZ peaks observed in a given cell line were not observed in any other cell type (Figure 5A, Suppl Table 4), suggesting that as well as regulating housekeeping functions, MAZ also plays an important role in the control of cell-type restricted gene expression programs. Consistent with this, Pearson correlation coefficients for these datasets indicated that MAZ exhibited globally distinct binding profiles in the different cell types (correlation coefficient < 0.85 for all pairwise comparisons) (Figure 5B). In particular, 18% of MAZ peaks observed in erythroid cells were not shared with other cell types (1778 peaks, termed “erythroid-specific peaks”). Comparison with gene expression profiles revealed that MAZ binds 41% (218/528) of promoters of genes with erythroid-specific expression, as defined in van de Lagemaat

et al., 2018 [34]. Gene ontology analysis revealed that genes with a TSS bound by MAZ in an erythroid-specific manner were significantly associated with erythroid differentiation and blood haemostasis functions (Figure 5C) as well as phenotypes associated with haematological diseases (Figure 5D). We also investigated MAZ binding at genomic loci which have previously been linked to clinical erythroid phenotypes in GWAS studies. Out of these 31 genes[48], MAZ peaks were present at the promoters of 21 of them (Suppl. Table 5). Interestingly, among these genes with promoters bound by MAZ in an erythroid-specific manner were the TFs GATA1 and KLF1, master regulators of erythropoiesis (Suppl. Table 5 and Suppl. Figure 7B). Taken together, these findings indicate that MAZ binds to the promoters of genes with key roles in erythroid differentiation and homeostasis.

MAZ signal is enriched at GATA1-bound enhancers in erythroid cells

Interestingly, the common and erythroid-specific peaks of MAZ displayed distinct localisation relative to genomic features (Figure 6A). While the majority (71%) of “common” peaks were localised at promoter regions, only 24% of the MAZ erythroid-specific peaks were located at promoters. In contrast to common MAZ peaks, erythroid-specific MAZ peaks were enriched at intergenic (24%) and intronic (40%) sites, suggesting that MAZ may play an important role at erythroid-specific distal regulatory elements. Comparison with our previously published catalogue of erythroid enhancers[34] confirmed that 27% of the erythroid-specific MAZ peaks coincided with known enhancer elements, compared with only 7% of the “common” peaks. Consistent with this observation, we observed enrichment of the enhancer-associated histone modification H3K4me1 on erythroid-specific non-TSS MAZ peaks, but not on common non-TSS MAZ peaks, with H3K4me3 following the reverse trend (Figure 6B).

Ubiquitous and lineage specific factors often work together at proximal and distal regulatory sequences. As such, we explored the association between MAZ and the erythroid master regulator GATA1 in primary erythroid cells. Overall, 49% (864/1778) of erythroid-specific MAZ binding sites overlapped with GATA1 binding sites, consistent with a functional cooperative interaction between these two factors. Interestingly, this co-association between MAZ and GATA was particularly prominent within erythroid enhancer regions, with GATA1 signal being particularly elevated at erythroid MAZ peaks within enhancers (Figure 6C, left panel) and conversely, MAZ signal being stronger at GATA1 enhancer peaks (Figure 6C, right panel). Taken together, these findings suggest a particular functional cooperativity between MAZ and GATA1 at erythroid enhancer elements.

DISCUSSION

In this study, a combination of biochemical characterisation and unbiased proteomic screening led to the identification of the ubiquitously expressed Zn-finger protein MAZ as a factor binding to neighbouring GC-rich sites within the human α-globin promoters. ChIP and ChIP-seq experiments subsequently confirmed *in vivo* binding of MAZ to the active HBA genes as well as to promoters and enhancers of other key genes within the erythroid differentiation pathway and to genomic loci linked to clinical erythroid phenotypes. Moreover, erythroid-specific binding of MAZ was enriched at distal regulatory enhancer elements, where it frequently co-associated with the erythroid master-regulator GATA1. We further showed that genetic variants within the MAZ locus were associated with phenotypic erythroid traits including haematocrit and familial erythrocytosis. Of interest, MAZ is recruited to its own promoter (Suppl. Figure

7A) and one of the variants, rs572982482, is located within this MAZ peak just between two MAZ canonical binding sequences, suggesting that this SNP could interfere with autoregulation by MAZ of its own gene. In addition, familial erythrocytosis, (associated with the rs11559000 SNP in the MAZ locus), involves the EPOR gene encoding the receptor for erythropoietin[49], which we identify as a target of MAZ binding in erythroid cells (Suppl. Table 5). Taken together, these findings reveal a previously unrecognised role for MAZ in the erythroid differentiation program.

Germline deletion of MAZ in mice results in perinatal lethality [50], as might be expected for a ubiquitously expressed TF. Recently, MAZ has been shown to be important for developmental haematopoiesis in Zebrafish [14]. Here, we showed that MAZ plays an important role during human erythropoiesis, with shRNA-mediated knockdown of MAZ impairing differentiation in primary erythroid cultures. Through integration of MAZ binding profiles from different cell types, we observed both common and erythroid-specific MAZ binding sites. Erythroid-specific MAZ binding is particularly enriched at distal regulatory elements, which are considered to be primary determinants of tissue-specific gene expression programs [32,51]. Importantly, it has been shown that the activity of erythroid enhancer elements cannot be predicted based on the binding of the master regulators of erythroid differentiation, GATA1 and TAL1, alone [32]. In contrast, combinatorial co-occupancy of enhancers by both lineage-specific and ubiquitously expressed TFs is a more reliable indicator of enhancer activity and cell-specific expression [32,51,52]. Our findings suggest that binding of MAZ together with lineage-restricted factors such as GATA1 might be an important step in the activation or maintenance of many erythroid enhancer elements. ~~We also observed that a subset of MAZ peaks in erythroid cells are co-bound by CTCF, suggesting that, as has been~~

~~recently observed in mouse ES and motorneuron cells [35], MAZ may contribute to the regulation of genome architecture by CTCF at some sites also in erythroid cells.~~

At present, it is unclear how binding of MAZ to its target sites is regulated. In particular, while MAZ is a ubiquitously expressed protein whose levels do not change dramatically during erythroid differentiation (Suppl Fig 6 and Suppl Fig 10), we observed differential *in vitro* binding by MAZ to the 13/14 probe in EMSA, as well as a highly dynamic recruitment of MAZ to the HBA locus in primary erythroid cultures, suggesting that its binding is not primarily regulated at the level of expression. Further, the differential binding of MAZ in EMSA is also unlikely to be due to relative abundance of the other competing Sp/X-KLFs in the different cell lines (Suppl Fig 11).

Western blot with α-MAZ antibody detected several bands in K562 cells, which were sensitive to MAZ shRNA (Figure 2C), suggesting the possibility that MAZ is subject to post-translational modifications in these cells. Indeed, Ray and Ray showed that phosphorylation of MAZ exerted strong effects on DNA binding in the absence of changes in total protein levels. Moreover, it is clear that while some phosphorylation events enhance MAZ binding [53,54] [55], phosphorylation at other sites actually decreases binding activity [56], suggesting that the regulation of MAZ by post-translational modifications during erythroid differentiation is likely to be complex. As well as post-translational modifications, our data also support the potential for regulatory crosstalk between MAZ and other KLFs, including SP1, SP3 and KLF3, which can compete for the same binding sites *in vitro*, and which exhibit related binding motifs in global ChIP-seq profiles. Indeed, complex cross-regulatory interactions involving MAZ and SP1 have also been reported at other promoters [57,58].

Overall, this study serves as an example of how a precise molecular characterisation of protein binding at a single regulatory element can be combined with an unbiased generic proteomics method to identify new trans-acting regulators. This approach has revealed the Zn-finger transcription factor MAZ as a previously unrecognised regulator of the erythroid differentiation program, which may be important for human erythroid phenotypic traits.

DECLARATIONS

Authors' contributions

D.G. and D.V. developed the hypothesis; F.B., D.E.D., I.F-V. D.C.J.F., M.L.H., V.S., J.A.S.S., H.A., and D.G. performed experiments and/or collected data; D.D., F.B., D.E.D., I.F-V., D.C.J.F., M.M., J.F., D.G. and D.V. analyzed data; interpreted data; and D.D., J.F., D.G., and D.V. wrote the manuscript, which was revised and approved by all authors.

Consent for publication

All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Doug Higgs, Robert Beagrie, Michele Goodhardt and Philipp Voigt for critically reading the manuscript, Dolores Lamb and Olga Medina-Martinez for helpful discussions. High-throughput sequencing was provided by Edinburgh Genomics (http://genomics.ed.ac.uk).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and protocols received ethical approval from Oxford University Ethical Review Panel (Reference T648). The data were analysed anonymously.

Data Availability

ChIP-seq datasets have been deposited with the NCBI Gene Expression Omnibus data bank under accession number GSE139281.

Funding

This work was supported by a University of Edinburgh Chancellor's Fellowship to Douglas Vernimmen and by Institute Strategic Grant funding to the Roslin Institute from the BBSRC [BB/J004235/1] and [BB/P013732/1]. Darya Deen is supported by Roslin Institute core funding to Douglas Vernimmen. David Garrick was supported by the Medical Research Council (UK) and INSERM (France).

FIGURE CAPTIONS

Figure 1. An erythroid-enriched complex binds a distal element of the HBA promoter. (A) Structure of the HBA core and distal promoter elements. The TSS is depicted as an angled arrow, and the locations of the EMSA probes are shown as grey bars. *In vivo* DMS footprints detected in K562 and erythroblasts [26] are represented by the black boxes. The nucleotide located 18 bp downstream of the TATA box (*) differs between the HBA1 (C) and HBA2 (G) promoters. **(B)** EMSA and supershift assays using K562 and EBV nuclear extracts with the indicated probes. **(C)** The fast-migrating species with probe 13/14 (species d) is enriched in erythroid cells **(D)** Sp/X-KLF-family antibodies cause retardation or disruption of the bands (a), (b) and (c), but not (d).

Figure 2. Identification of MAZ by mass spectrometry. (A) Schematic representation of the affinity purification screen. **(B)** and **(C)** Western blots (left) and EMSA (right) showing that the intensity of species (d) is dependent on MAZ expression using knockdown (B) or overexpression (C) experiments. **(D)** MAZ is recruited to the active α-globin promoter *in vivo*. Analysis of MAZ binding at the α-globin locus in EBV-lymphoblasts, K562 cells and human primary erythroid cells by ChIP-qPCR. 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 position of probes within the α-globin cluster are indicated on the heading map. The α-globin genes themselves are covered by three probes (Pr/Ex1, Ex2, Ex3). Error bars correspond to one SEM from two independent ChIPs.

Figure 3. MAZ knockdown impairs erythropoiesis. (A) Flow cytometry analysis of CD71 and GPA expression at day 15 of representative primary erythroblast differentiation cultures either untransduced (UT), or infected with lentivirus expressing

scramble shRNA (Scr), MAZ shRNA 699 or MAZ shRNA 703. **(B)** Quantitation of flow cytometry in (A). Shown is the mean and standard deviation from two independent differentiation cultures. BFU-E are $CD71^{-}/GPA^{-}$, CFU-E/Pro-E are $CD71^{+}/GPA^{lo/int}$ and Int. Ery are $CD71^{+}/GPA^{+}$. **(C)** Cell pellets from primary cultures at d15 of differentiation. Cultures with the MAZ shRNA 699 and 703 are pale compared to Scr control, indicating less extensive hemoglobinisation. **(D)** Real-time RT-PCR analysis of expression of MAZ, HBA and HBB after 15 d of primary erythroid differentiation cultures. Shown is the mean and standard deviation from two independent differentiation cultures. For each gene, expression (relative to the PABPC1 gene) is normalized to the mean value observed with scramble shRNA. **(E)** Western blot analysis of the expression of MAZ, HBA, and HBB after 15 d in two independent primary erythroid differentiation cultures. β-actin was used as loading control. **(F)** Densitometry analysis of the western blots shown in (E). For each protein, expression (relative to the β actin) is normalized to the mean value observed with scramble shRNA. **(G)** Variants around the MAZ locus significantly associated with changes in clinical erythroid traits (adjusted p-value <0.01).

Figure 4. MAZ is enriched at active TSS and binds to DNA through a $(G)_3C/A(G)_4$ consensus site. (A) MAZ ChIP-seq enrichment profiles surrounding the α- (top) and β- (bottom) globin loci. **(B)** Genomic distribution of MAZ binding sites showing the number of peaks overlapping the category of genomic element indicated. **(C)** Heatmap plots of ATAC-seq, H3K4me3, H3K27ac, Pol II and H3K27me3 signal centred on MAZ peaks in erythroblast cells (sorted according to decreasing MAZ ChIP signal). **(**D) Position weight matrices of the enriched motifs (E-value <0.01) in the training set of MAZ peaks (500 highest ranked MAZ peaks) and their occurrence in all MAZ peaks.

(E) Motif localisation curves relative to the MAZ peak centres. The numbers of the enriched motifs correspond to the logos in (D).

Figure 5. Analysis of erythroid-specific MAZ signal. (A) Heatmap plots of MAZ ChIP-seq datasets from six cell types separated into MAZ common and cell-line-specific peaksets. Numbers on left indicate the common peaks (1), and peaks specific to erythroblasts (2), HepG2 (3), A549 (4), GM12878 (5), MCF-7 (6) and IMR90 (7) cells. (**B)** Pearson correlation analysis of MAZ ChIP-seq datasets from the six cell types. **(C)** Gene ontology (by biological function) of genes containing erythroid-specific MAZ promoter signal. **(D)** Human Phenotype Ontology of genes containing erythroid-specific MAZ promoter signal.

Figure 6. Erythroid-specific MAZ signal is enriched on GATA1-bound enhancers. (A) Genome distribution of MAZ common and erythroid-specific binding sites. **(B)** Average H3K4me1, H3K4me3, MAZ and ATAC-seq signal in primary erythroid cells plotted against MAZ erythroid-specific peakset (black) and MAZ common peakset (red) where signal from promoters have been excluded (non-TSS set). **(C)** Left panel: Heatmap plot of GATA1 ChIP-seq centred on MAZ erythroid peaks overlapping enhancers (top) and promoters (bottom). Right panel: Heatmap plot of MAZ ChIP-seq centred on GATA1 peaks overlapping enhancers (top) and promoters (bottom).

REFERENCES

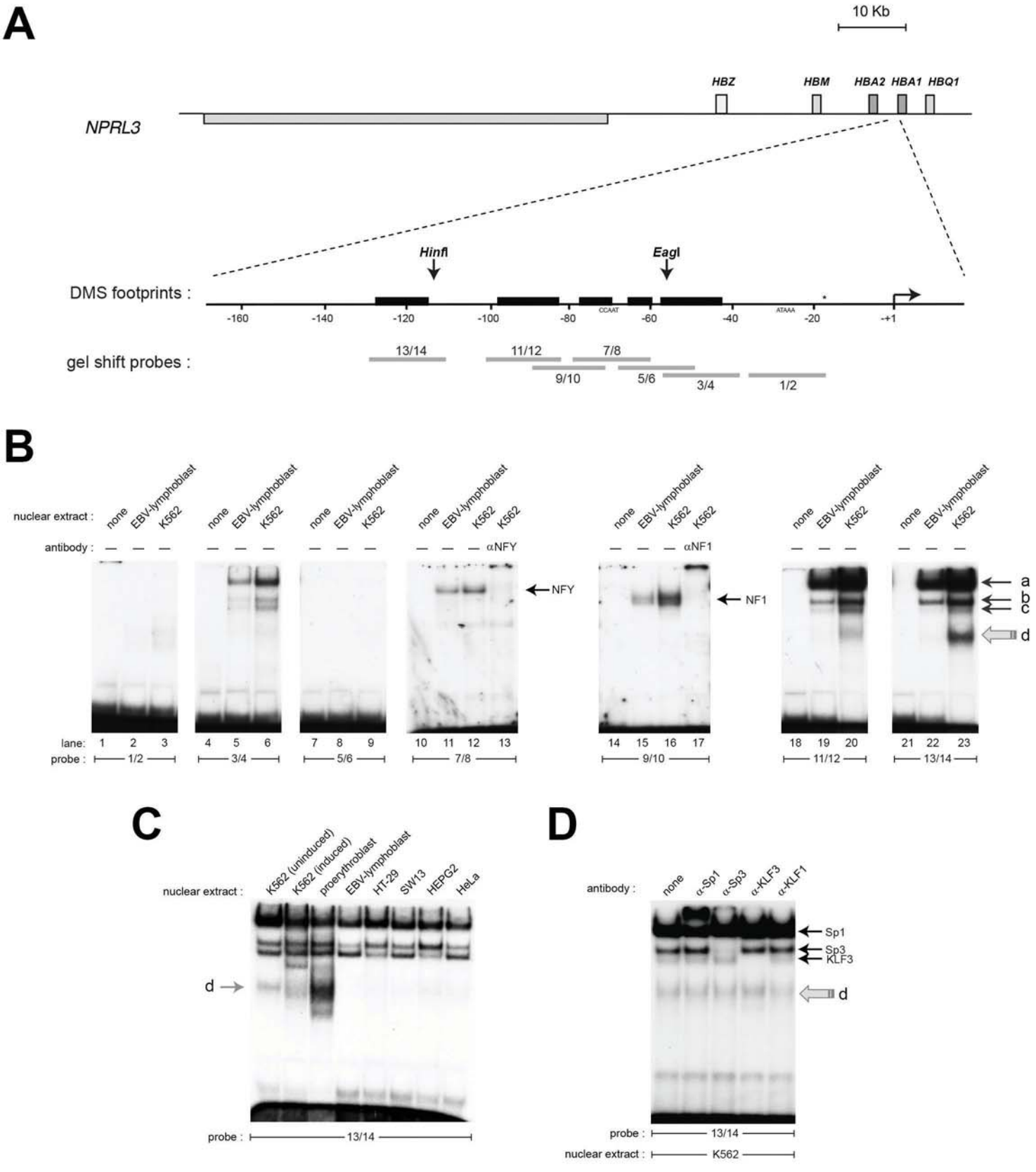
1. Vernimmen D. Globins, from Genes to Physiology and Diseases. *Blood Cells Mol Dis*. 2018;70:1.
2. Philipsen S, Hardison RC. Evolution of hemoglobin loci and their regulatory elements. *Blood Cells Mol Dis*. 2018;70:2-12.
3. Wontakal SN, Guo X, Smith C, et al. A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation. *Proc Natl Acad Sci U S A*. 2012;109(10):3832-3837.
4. Perkins A, Xu X, Higgs DR, et al. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood*. 2016;127(15):1856-1862.
5. Katsumura KR, DeVilbiss AW, Pope NJ, Johnson KD, Bresnick EH. Transcriptional mechanisms underlying hemoglobin synthesis. *Cold Spring Harb Perspect Med*. 2013;3(9):a015412.
6. Merika M, Orkin SH. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol Cell Biol*. 1995;15(5):2437-2447.
7. Gregory RC, Taxman DJ, Seshasayee D, Kensinger MH, Bieker JJ, Wojchowski DM. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood*. 1996;87(5):1793-1801.
8. Van Loo PF, Bouwman P, Ling KW, et al. Impaired hematopoiesis in mice lacking the transcription factor Sp3. *Blood*. 2003;102(3):858-866.
9. Woo AJ, Kim J, Xu J, Huang H, Cantor AB. Role of ZBP-89 in human globin gene regulation and erythroid differentiation. *Blood*. 2011;118(13):3684-3693.
10. Funnell AP, Vernimmen D, Lim WF, et al. Differential regulation of the alpha-globin locus by Kruppel-like factor 3 in erythroid and non-erythroid cells. *BMC Mol Biol*. 2014;15(1):8.
11. Meinders M, Kulu DI, van de Werken HJ, et al. Sp1/Sp3 transcription factors regulate hallmarks of megakaryocyte maturation and platelet formation and function. *Blood*. 2015;125(12):1957-1967.
12. Woo AJ, Patry CA, Ghamari A, et al. Zfp281 (ZBP-99) plays a functionally redundant role with Zfp148 (ZBP-89) during erythroid development. *Blood Adv*. 2019;3(16):2499-2511.
13. Gilmour J, O'Connor L, Middleton CP, et al. Robust hematopoietic specification requires the ubiquitous Sp1 and Sp3 transcription factors. *Epigenetics Chromatin*. 2019;12(1):33.
14. Gao P, Chen C, Howell ED, et al. Transcriptional regulatory network controlling the ontogeny of hematopoietic stem cells. *Genes Dev*. 2020.
15. Pope SH, Fibach E, Sun J, Chin K, Rodgers GP. Two-phase liquid culture system models normal human adult erythropoiesis at the molecular level. *Eur J Haematol*. 2000;64(5):292-303.
16. Ray A, Shakya A, Kumar D, Ray BK. Overexpression of serum amyloid A-activating factor 1 inhibits cell proliferation by the induction of cyclin-dependent protein kinase inhibitor p21WAF-1/Cip-1/Sdi-1 expression. *J Immunol*. 2004;172(8):5006-5015.

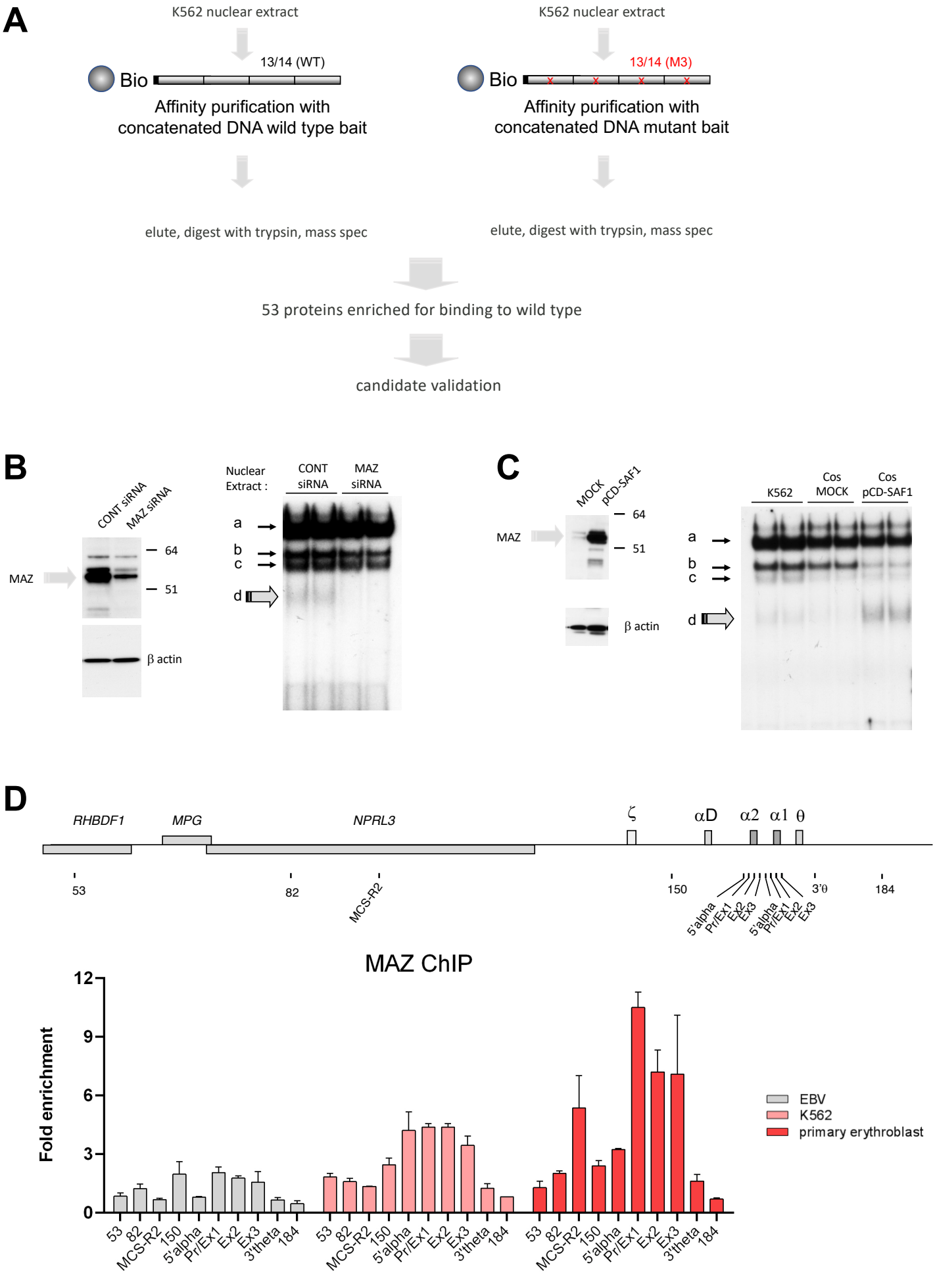
17. Andrews NC, Faller DV. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res*. 1991;19(9):2499.
18. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J*. 2007;26(8):2041-2051.
19. Leberbauer C, Boulme F, Unfried G, Huber J, Beug H, Mullner EW. Different steroids co-regulate long-term expansion versus terminal differentiation in primary human erythroid progenitors. *Blood*. 2005;105(1):85-94.
20. Butter F, Davison L, Viturawong T, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet*. 2012;8(9):e1002982.
21. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2007;2(8):1896-1906.
22. 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.
23. Lower KM, Hughes JR, De Gobbi M, et al. Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc Natl Acad Sci U S A*. 2009;106(51):21771-21776.
24. Vernimmen D, Lynch MD, De Gobbi M, et al. Polycomb eviction as a new distant enhancer function. *Genes Dev*. 2011;25(15):1583-1588.
25. 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.
26. Rombel I, Hu KY, Zhang Q, Papayannopoulou T, Stamatoyannopoulos G, Shen CK. Transcriptional activation of human adult alpha-globin genes by hypersensitive site-40 enhancer: function of nuclear factor-binding motifs occupied in erythroid cells. *Proc Natl Acad Sci U S A*. 1995;92(14):6454-6458.
27. Himeda CL, Ranish JA, Hauschka SD. Quantitative proteomic identification of MAZ as a transcriptional regulator of muscle-specific genes in skeletal and cardiac myocytes. *Mol Cell Biol*. 2008;28(20):6521-6535.
28. Parks CL, Shenk T. Activation of the adenovirus major late promoter by transcription factors MAZ and Sp1. *J Virol*. 1997;71(12):9600-9607.
29. Ashfield R, Patel AJ, Bossone SA, et al. MAZ-dependent termination between closely spaced human complement genes. *EMBO J*. 1994;13(23):5656-5667.
30. Brown JM, Leach J, Reittie JE, et al. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*. 2006;172(2):177-187.
31. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50(11):1593-1599.
32. 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.
33. 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.
34. 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.

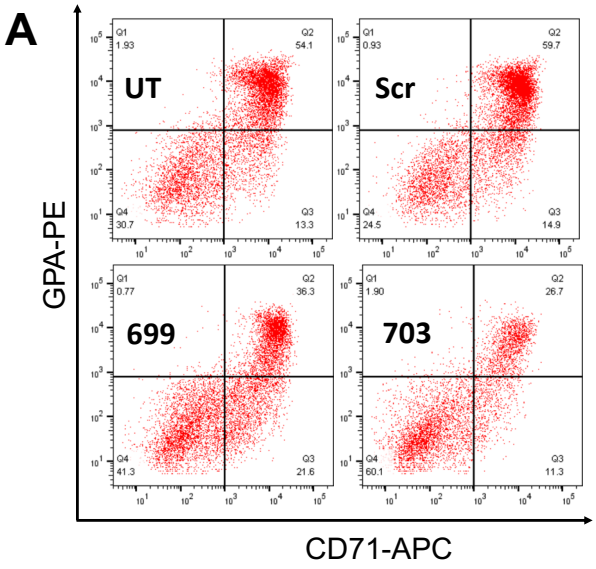
35. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 2011;27(12):1696-1697.
36. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*. 2012;40(17):e128.
37. Song J, Ugai H, Ogawa K, et al. Two consecutive zinc fingers in Sp1 and in MAZ are essential for interactions with cis-elements. *J Biol Chem*. 2001;276(32):30429-30434.
38. Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48(D1):D87-D92.
39. Lew A, Rutter WJ, Kennedy GC. Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc Natl Acad Sci U S A*. 2000;97(23):12508-12512.
40. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res*. 2008;36(6):1755-1769.
41. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J Biol Chem*. 2010;285(29):22003-22016.
42. Membrino A, Cogoi S, Pedersen EB, Xodo LE. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One*. 2011;6(9):e24421.
43. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biol*. 2007;8(2):R24.
44. Gillespie MA, Palii CG, Sanchez-Taltavull D, et al. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Mol Cell*. 2020;78(5):960-974 e911.
45. Persikov AV, Singh M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res*. 2014;42(1):97-108.
46. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
47. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794-D801.
48. Tumburu L, Thein SL. Genetic control of erythropoiesis. *Curr Opin Hematol*. 2017;24(3):173-182.
49. Huang LJ, Shen YM, Bulut GB. Advances in understanding the pathogenesis of primary familial and congenital polycythaemia. *Br J Haematol*. 2010;148(6):844-852.
50. Haller M, Au J, O'Neill M, Lamb DJ. 16p11.2 transcription factor MAZ is a dosage-sensitive regulator of genitourinary development. *Proc Natl Acad Sci U S A*. 2018;115(8):E1849-E1858.
51. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459(7243):108-112.
52. Stadhouders R, Cico A, Stephen T, et al. Control of developmentally primed erythroid genes by combinatorial co-repressor actions. *Nat Commun*. 2015;6:8893.

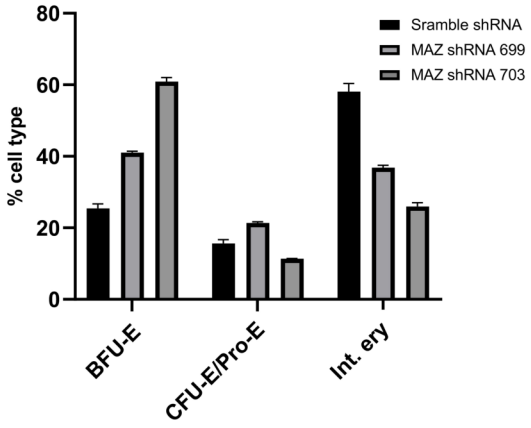
53. Tsutsui H, Geltinger C, Murata T, et al. The DNA-binding and transcriptional activities of MAZ, a myc-associated zinc finger protein, are regulated by casein kinase II. *Biochem Biophys Res Commun*. 1999;262(1):198-205.
54. Ray A, Yu GY, Ray BK. Cytokine-responsive induction of SAF-1 activity is mediated by a mitogen-activated protein kinase signaling pathway. *Mol Cell Biol*. 2002;22(4):1027-1035.
55. Ray A, Ray P, Guthrie N, Shakya A, Kumar D, Ray BK. Protein kinase A signaling pathway regulates transcriptional activity of SAF-1 by unmasking its DNA-binding domains. *J Biol Chem*. 2003;278(25):22586-22595.
56. Lee WP, Lan KH, Li CP, Chao Y, Lin HC, Lee SD. Akt phosphorylates myc-associated zinc finger protein (MAZ), releases P-MAZ from the p53 promoter, and activates p53 transcription. *Cancer Lett*. 2016;375(1):9-19.
57. Song J, Ugai H, Nakata-Tsutsui H, et al. Transcriptional regulation by zinc-finger proteins Sp1 and MAZ involves interactions with the same cis-elements. *Int J Mol Med*. 2003;11(5):547-553.
58. Her S, Claycomb R, Tai TC, Wong DL. Regulation of the rat phenylethanolamine N-methyltransferase gene by transcription factors Sp1 and MAZ. *Mol Pharmacol*. 2003;64(5):1180-1188.

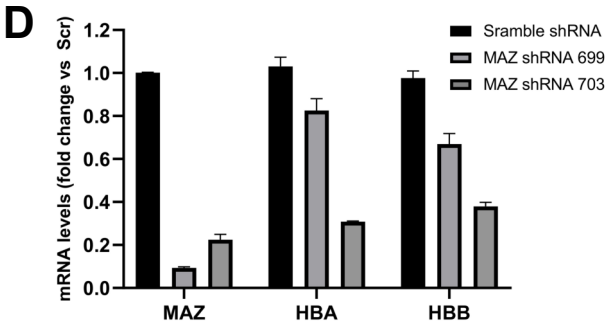
Figure 1

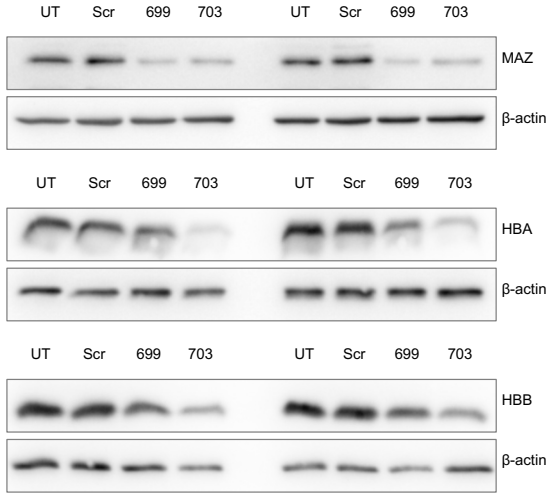

A
10 Kb
HBZ
HBM
HBA2
HBA1
HBQ1
NPRL3
HinfI
EagI
DMS footprints :
CCAAT
ATAAA
-160
-140
-120
-100
-80
-60
-40
-20
-+1
gel shift probes :
13/14
11/12
7/8
9/10
5/6
3/4
1/2
B
nuclear extract :
none
EBV-lymphoblast
K562
antibody :
αNFY
αNF1
NFY
NF1
a
b
c
d
lane:
probe :
1/2
3/4
5/6
7/8
9/10
11/12
13/14
C
nuclear extract :
K562 (uninduced)
K562 (induced)
proerythroblast
EBV-lymphoblast
HT-29
SW13
HEPG2
HeLa
d
probe :
13/14
D
antibody :
none
α-Sp1
α-Sp3
α-KLF3
α-KLF1
Sp1
Sp3
KLF3
d
probe :
13/14
nuclear extract :
K562

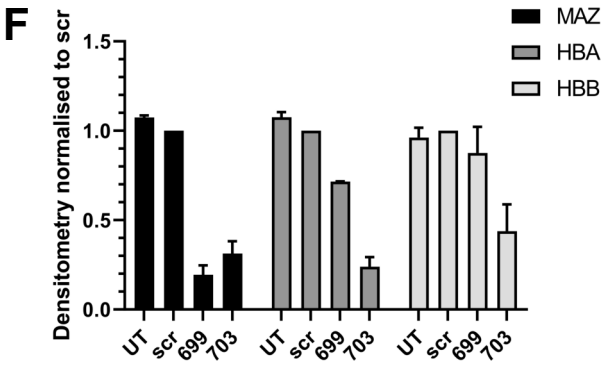

A
K562 nuclear extract
13/14 (WT)
Bio
Affinity purification with concatenated DNA wild type bait
elute, digest with trypsin, mass spec
K562 nuclear extract
13/14 (M3)
Bio
Affinity purification with concatenated DNA mutant bait
elute, digest with trypsin, mass spec
53 proteins enriched for binding to wild type
candidate validation
B
CONT siRNA
MAZ siRNA
MAZ
64
51
β actin
Nuclear Extract :
CONT siRNA
MAZ siRNA
a
b
c
d
C
MOCK
pCD-SAF1
MAZ
64
51
β actin
K562
Cos MOCK
Cos pCD-SAF1
a
b
c
d
D
RHBDF1
MPG
NPRL3
ζ
αD
α2
α1
θ
53
82
MCS-R2
150
3'θ
184
5'alpha
Pr/EX1
Ex2
Ex3
MAZ ChIP
Fold enrichment
12
9
6
3
0
EBV
K562
primary erythroblast









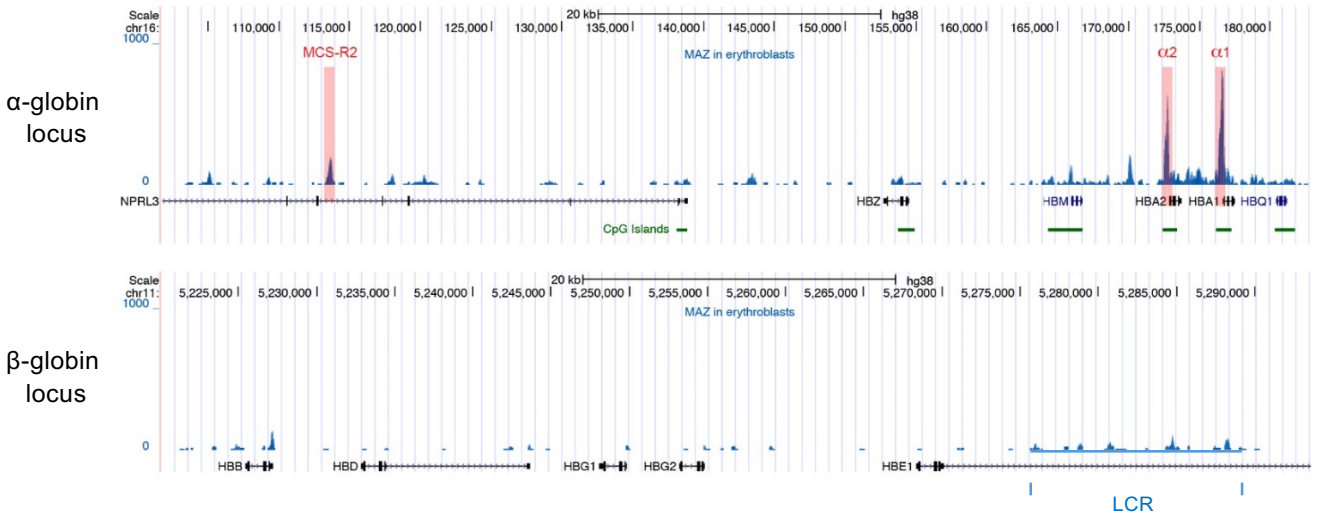


G

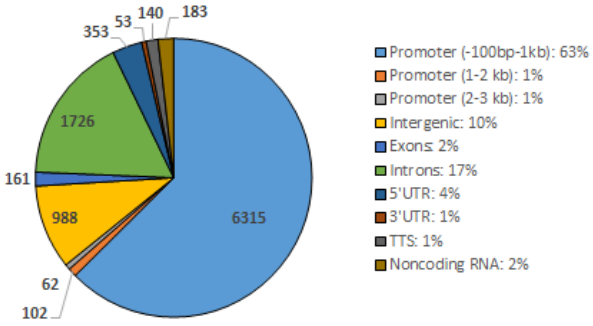
Variant	MAZ region	Erythroid trait	Change	p-value
rs572982482	MAZ promoter	Haematocrit percentage	Increase	3.54E-03
		Haemoglobin concentration	Increase	7.30E-03
		Mean reticulocyte volume	Decrease	9.66E-03
		Red blood cell (erythrocyte) count	Increase	1.96E-04
		Reticulocyte count	Increase	4.85E-04
		Reticulocyte percentage	Increase	4.63E-03
rs72798129	intron variant	Mean reticulocyte volume	Decrease	1.97E-05
		Mean sphered cell volume	Decrease	8.94E-03
rs11559000	3' UTR variant	Familial erythrocytosis		5.42E-03

Figure 4

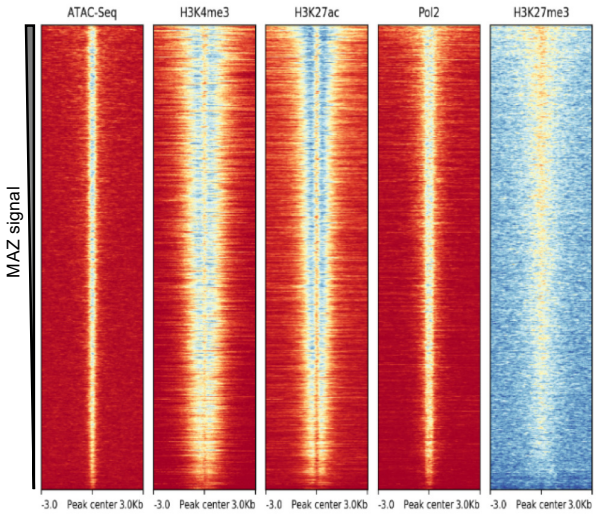
A



B



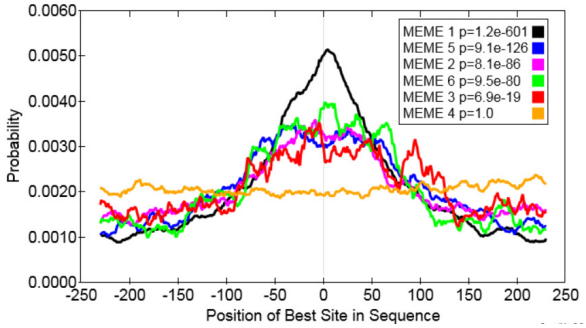
C

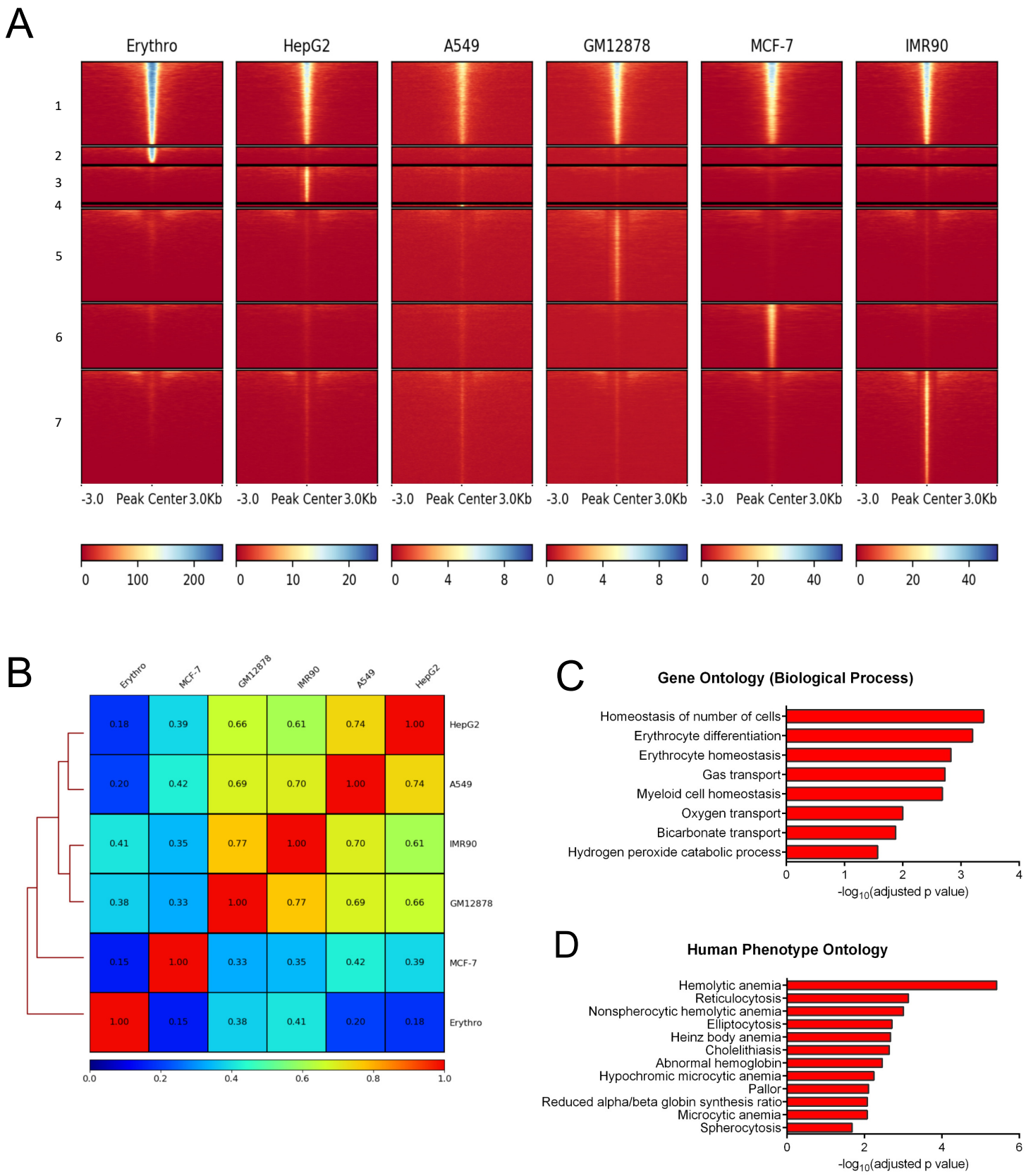


D

	Motif Logo	E-value	Canonical motif	Percentage of motif-containing peaks	Motif centrality p-value
1		$2.7e^{-303}$	SP1	92%	$1.2e^{-601}$
2		$7.1e^{-056}$	NFY	3%	$8.1e^{-86}$
3		$1.4e^{-031}$	-	7%	$6.9e^{-19}$
4		$2.1e^{-025}$	-	31%	1.0
5		$5.9e^{-016}$	-	21%	$9.1e^{-126}$
6		$4.3e^{-006}$	-	1%	$9.5e^{-80}$

E




A
Erythro
HepG2
A549
GM12878
MCF-7
IMR90
1
2
3
4
5
6
7
-3.0 Peak Center 3.0Kb
0 100 200
0 10 20
0 4 8
0 4 8
0 20 40
0 20 40
B
Erythro
MCF-7
GM12878
IMR90
A549
HepG2
0.18 0.39 0.66 0.61 0.74 1.00 HepG2
0.20 0.42 0.69 0.70 1.00 0.74 A549
0.41 0.35 0.77 1.00 0.70 0.61 IMR90
0.38 0.33 1.00 0.77 0.69 0.66 GM12878
0.15 1.00 0.33 0.35 0.42 0.39 MCF-7
1.00 0.15 0.38 0.41 0.20 0.18 Erythro
0.0 0.2 0.4 0.6 0.8 1.0
C
Gene Ontology (Biological Process)
Homeostasis of number of cells
Erythrocyte differentiation
Erythrocyte homeostasis
Gas transport
Myeloid cell homeostasis
Oxygen transport
Bicarbonate transport
Hydrogen peroxide catabolic process
0 1 2 3 4
-log10(adjusted p value)
D
Human Phenotype Ontology
Hemolytic anemia
Reticulocytosis
Nonspherocytic hemolytic anemia
Elliptocytosis
Heinz body anemia
Cholelithiasis
Abnormal hemoglobin
Hypochromic microcytic anemia
Pallor
Reduced alpha/beta globin synthesis ratio
Microcytic anemia
Spherocytosis
0 2 4 6
-log10(adjusted p value)

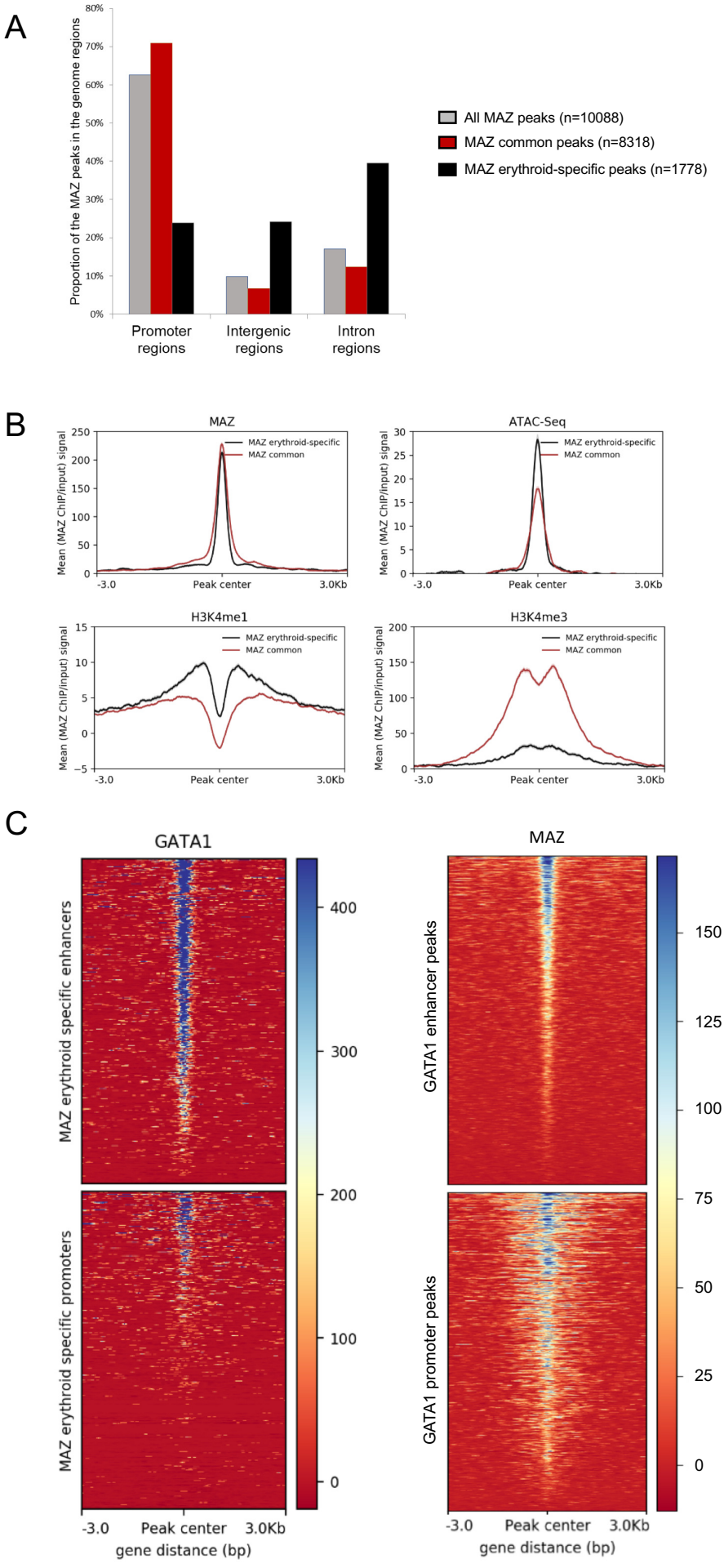

A
Proportion of the MAZ peaks in the genome regions
80%
70%
60%
50%
40%
30%
20%
10%
0%
Promoter regions
Intergenic regions
Intron regions
All MAZ peaks (n=10088)
MAZ common peaks (n=8318)
MAZ erythroid-specific peaks (n=1778)
B
MAZ
ATAC-Seq
H3K4me1
H3K4me3
Mean (MAZ ChIP/input) signal
MAZ erythroid-specific
MAZ common
-3.0
Peak center
3.0Kb
C
GATA1
MAZ
MAZ erythroid specific enhancers
MAZ erythroid specific promoters
GATA1 enhancer peaks
GATA1 promoter peaks
gene distance (bp)

Table 1. EMSA Results

Probe	Location (bp)	Predicted motifs (5)	Gel shift results
1/2	-37/-19	TATA box	No shifted species
3/4	-56/-39	α-IRP -51/-42 site	Shifted species detected in EBV and K562
5/6	-67/-50	Sp1 -61/-56 site	No shifted species
7/8	-78/-61	NFY -72/-65 site	Shifted species detected in EBV and K562 Binding impaired by α-NFY antibody
9/10	-88/-72	NF1 -85/-72 site	Shifted species detected in EBV and K562 Binding impaired by α-NF1 antibody
11/12	-100/-83		Three shifted species detected in EBV and K562 One shifted species only in K562
13/14	-128/-111	Sp1 -121/-116 site	Three shifted species detected in EBV and K562 One shifted species only in K562