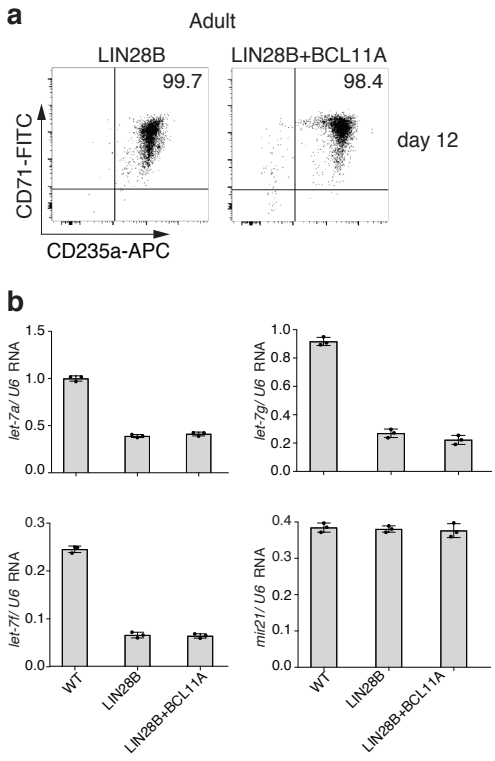
Supplementary Information for: Control of human hemoglobin switching by LIN28B-mediated regulation of BCL11A translation

Contents



Supplementary Fig. 1. LIN28B influences *BCL11A* mRNA translation independent of the let-7 pathway. **a**, Representative flow cytometry plots showing CD71 and CD235a surface expression in primary adult erythroid cells expressing LIN28B alone (left) or LIN28B and BCL11A (right) on day 12 of differentiation. n=3; 3 biologically independent experiments. **b**, *Let-7a, let-7g, let-7f* miRNA expression (normalized to *U6* expression), assessed by qRT-PCR in LIN28B (middle) or concomitant LIN28B and BCL11A (right) expressing adult HSPCs undergoing erythroid differentiation until day 12 (n=3; 3 independent experiments). *Mir21* is used as a control. 'WT' (left) denotes uninfected primary adult erythroid cells. Mean is plotted and error bars show s.d.

SUPPLEMENTARY TABLES

Supplementary Tables 1 and 2.

Tab 1. Data relevant to 18S rRNA RAP-MS. Sample size (n) = 2 biological replicates for U1 and n=2 biological replicates for 18S (using a label swap). Statistical test used: two-tailed moderated t-test adjusted for multiple hypothesis testing. Tab 2. List of proteins with relative enrichment to 18S rRNA and their relative newborn vs. adult enrichment. Sample size (n) = 2 replicates. Statistical test: generalized linear model from sleuth.

Supplementary Table 3. Conjugated antibodies and reagents used for flow cytometry

Antibody (anti-human)	Manufacturer	Catalog number	Clone	Lot number	Dilution
CD71 (Transferrin Receptor) - FITC	Fisher Scientific (eBioscience)	11-0719-42	OKT9	1930273	1:20
CD235a (Glycophorin A) -APC	Fisher Scientific (eBioscience)	17-9987-42	HIR2	4283492 1998340	1:20
CD71 (Transferrin Receptor) - PE-Cy7	Fisher Scientific (eBioscience)	25-0719-42	OKT9	2093306	1:20
CD34-PE	Fisher Scientific (eBioscience)	8012-0349-120	4H11	4286430 4311260	1:20
Propidium iodide (PI)	Biolegend	421301	-	B258516	1:1000
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Fisher Scientific (Molecular Probes)	D1306	-	1711782	1:1000

Supplementary Table 4. Primers used for qRT-PCR

Gene	Forward Primer	Reverse Primer
BCL11A (exon 3 - 4)	5'-GCCTGGGATGAGTGCAGAAT-3'	5'- ATGCACTGGTGAATGGCTGT-3'
BCL11A (exon 2 - 3)	5'-CGCCAGAGGATGACGATTGT-3'	5'-CCCCTCCAGTGCAGAAGTTT-3'
LIN28B	5' -CATCTCCATGATAAACCGAGAGG-3'	5'-GTTACCCGTATTGACTCAAGGC-3'
GATA1	5'-CTGTCCCCAATAGTGCTTATGG-3'	5'- GAATAGGCTGCTGAATTGAGGG-3'
ACTB	5'-AGAAAATCTGGCACCACACC-3'	5'-GGGGTGTTGAAGGTCTCAAA-3'
GAPDH	5'-TGCACCACCAACTGCTTAGC-3'	5'- GGCATGGACTGTGGTCATGAG-3'

HMGA2	5’-AGACCTAGGAAATGGCCACA-3’	5’-AGACCTAGGAAATGGCCACA-3’
HBB	5’-GATGGCCTGGCTCACCTG-3’	5’-TGCCCAGGAGCCTGAAGT-3’
HBG1/*HBG2*	5’-TTTGCCCAGCTGAGTGAACT-3’	5’ - CACATTTCCCAGGAGCTTGAAG-3’
HMGA1	5’-TGCTGCCAAGACCCGGAA-3’	5’-CCTCTTCCTCCTTCTCCAGT-3’
ALAS2	5’-ACCTACCGTGTGTTCAAGACT-3’	5’-AGATGCCTCAGAGAAATGTTGG-3’
LDB1	5’-CAAACGGCTTCAGAACTGGAC-3’	5’-TCCGGCCAATGGTATATCTCTT-3’
KLF1	GGTTGCGGCAAGAGCTACA	GTCAGAGCGCGAAAAAGCAC
LMO2	5’-AAGCGGATTCGTGCCTATGAG-3’	5’-AGTTGATGAGGAGGTATCTGTCA-3’

Supplementary Note

Crosslinking immunoprecipitation (CLIP)

For each CLIP experiment, 50 million cells were grown. Cells were washed once with PBS and then crosslinked on ice using 0.4 J/cm^2 of 254 nm UV light in a Stratalinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500g for 5 minutes, and flash-frozen in liquid nitrogen for storage at -80°C. To prepare cell lysates, pellets were thawed on ice and resuspended in NP40 lysis buffer (50 mM HEPES pH 7.5, 150 mM KCl, 2 mM EDTA, 1% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail) and incubated on ice for 10 minutes. Cell lysates were sonicated using a Branson Digital Sonifier with a microtip set at 5 W power for a total of 1 minute 30 seconds in intermittent pulses (0.7 sec on, 2.3 sec off), followed by RNase I (Thermo Fisher Scientific) digestion (0.5 U/µl, 10 minutes at 23°C). Subsequently, 15 µl/ml Murine RNase Inhibitor (New England Biolabs) was added, followed by DNA digestion (20 U TURBO DNase (2 U/µl; Thermo Fisher Scientific), 2.5 mM $MgCl_2$ and 0.5 mM CaCl) for 20 minutes at 37°C. Samples were incubated on ice for 10 minutes before clearing lysates by centrifugation at 15,000g for 15 minutes. Insoluble material was removed and total protein concentration was determined by BCA assay (Thermo Fisher Scientific). For each immunoprecipitation experiment, lysate (10 mg total protein) was pre-cleared by incubating with Protein A/G magnetic beads (using 30 µl/mg total protein) for 30 minutes at 4°C. In the meantime, antibodies (6 µg/mg total protein) were coupled to Protein A/G magnetic beads (using 30 µl/mg total protein) at room temperature for 45 minutes (antibodies used: LIN28B- 4196, Cell Signaling). Unbound antibody was removed and the pre-cleared lysates were added to antibody-coupled beads and incubated overnight at 4°C. The following day, the beads were washed 4 times in IP wash buffer (50 mM HEPES pH 7.5, 150 mM KCl, 0.5% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail). Protein-RNA complexes were eluted by pH shift using IgG Elution Buffer (Thermo Fisher Scientific) and incubated 30 minutes at room temperature. Protein-

RNA complexes were resolved by SDS-PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 hour, followed by transfer to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific). Protein-RNA complexes were visualized using ponceau staining and desired complexes were excised from membrane using a clean scalpel. Membrane pieces were immediately subjected to proteinase K treatment by adding 250 µl proteinase K solution (4 mg/ml Proteinase K (New England Biolabs), 100 mM Tris-HCl pH 7.5, 150 mM NaCl, 12.5 mM EDTA, 1% (w/v) SDS) and incubating 1 hour at 55°C. Following proteinase K treatment, RNA was phenol-chloroform extracted using Heavy Phase Lock Gel tubes (5Prime) and purified with the Zymo RNA Clean & Concentrator-5 kit by following the manufacturer's instructions. For the rRNA depleted sample, rRNA was removed with the NEBNext rRNA Depletion Kit (New England Biolabs). RNA sequencing libraries were constructed using the SMARTer smRNA-Seq Kit (Clontech) by following the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 instrument to an average read depth of 15-20 million reads with 50 bp read.

Size-matched input libraries (SM input)[1] were prepared by resolving 2% of input lysates by SDS-PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 hour. SDS-PAGE gels were transferred to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific) and proteins migrating at the molecular weight range of the target protein were excised using a clean scalpel. RNA was released by proteinase K treatment and purified as described in the previous section. End-repair of input RNA was performed by adjusting RNA volume to 19.5 µl with H_2O and adding 2.5 µl 10x FastAP buffer (Thermo Fisher Scientific), 2.5 U FastAP enzyme (1 U/µl; Thermo Fisher Scientific), 0.5 µl Murine RNase Inhibitor (New England Biolabs)) and incubating for 20 minutes at 37°C. In the meantime, Polynucleotide kinase mix was prepared (56 µl H_2O, 10 µl 10x PNK buffer (New England Biolabs), 1 µl Murine RNase Inhibitor, 7 µl T4 PNK (10 U/µl; New England Biolabs), 1 µl TURBO DNase) ,75 µl was

added to each 25 µl sample and samples were incubated for 20 minutes at 37°C. RNA was purified with the Zymo RNA Clean & Concentrator-5 kit using the manufacturer's instructions. SM input samples and CLIP samples were jointly carried through rRNA removal and sequencing library construction as described above.

CLIP-seq analysis

After adaptor trimming, single-end reads were aligned using STAR version 2.5.1b and subsequently filtered for uniquely mapping reads. Peaks were called using MACS2[2] with parameters "--nomodel --nolambda --bw 100 --extsize 50" for two replicates of LIN28B. To identify LIN28B-specific binding sites, peaks present in both the LIN28B and a similar library preparation of the RNA binding protein PURA were excluded. A total of 48,557 and 40,397 peaks were then called for replicate 1 and 2 of the LIN28B CLIP-seq datasets, including 12,870 and 10,855 respectively containing the canonical GGAGA motif. A consensus peak set was derived using by computing the per-peak irreproducible rate (IDR) over the two replicates, which yielded 20,001 peaks at an IDR of 1%. Genomic annotations for 5'UTR, 3'UTR, exons, and introns were downloaded from the UCSC Genome Browser. 5-mer matching was performed using the motifmatchr package as part of the chromVAR analysis framework[3].

RNA antisense purification and mass spectrometry (RAP MS)

To capture endogenous 18S and U1 transcripts, 5' biotinylated 90-mer DNA oligonucleotides (Integrated DNA Technologies) antisense to the target RNA sequence were designed and synthesized, with the exception of regions that matched to other transcripts or genomic regions as previously described[4]. For 18S and U1 antisense purifications 100 million erythroid K562 cells were grown for each RNA target in SILAC RPMI media supplemented with heavy or light labeled SILAC amino acids. Suspension cells were harvested by centrifugation, washed once with PBS and then uniformly spread out on a cell culture dish and crosslinked on ice using 0.8 J/cm2 of 254

nm UV light in a Stratalinker (Stratagene). Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500g for 5 minutes, and flash-frozen in liquid nitrogen for storage at -80°C. Preparation of total cell lysates was performed as previously described[5]. For antisense purification of crosslinked protein-RNA complexes the following modifications were included into the previously described procedure: All buffers were pre-heated to 55°C. 10 µg pooled antisense probes were used for 100 million lysed cells. For pre-clear of lysates and capture of RNA/DNA hybrids 1 ml Streptavidin Dynabeads MyOne C1 magnetic beads (Thermo Fisher Scientific) was used for 100 million cells. Elution of captured proteins from streptavidin beads was achieved by digesting nucleic acids using 250 U of Benzonase (Millipore), 25 U RNase A and 1000 U RNase T1 (Thermo Fisher Scientific) for 8 hours at 37°C. Trichloroacetic acid-precipitated proteins were reconstituted in 8 M urea and 50 mM Tris-HCl pH 8.5 and stored at -20°C. Samples were reduced (2 uL 500 mM DTT, 30 minutes, RT), alkylated (4 uL 500 mM IAA, 45 minutes, dark). Samples were digested with Lys-C using 0.1 ug of Lys-C in 100 mM Tris-HCl –incubate at room temperature for 2 hours and then diluting to 2M Urea with 100 mM Tris-Hcl (pH7.8). A second digest with 0.5 ug of Trypsin was done overnight at room temperature. Samples were quenched by adding formic acid to a final concentration of 5%. Samples were desalted using StAGE tips as described previously[6] and dried down in a speedvac. Samples were reconstituted in 3% ACN, 0.1% FA. Reconstituted peptides were separated on an online nanoflow EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific) and analyzed on a benchtop Orbitrap Q Exactive mass spectrometer (Thermo Fisher Scientific). The peptide samples were injected onto a capillary column (Picofrit with 10 µm tip opening / 75 µm diameter, New Objective, PF360-75-10-N-5) packed in-house with 30 cm C18 silica material (1.9 µm ReproSil-Pur C18-AQ medium, Dr. Maisch GmbH, r119.aq). The capillary columns were heated to 50 °C in column heater sleeves (Phoenix-ST) to reduce backpressure during UHPLC separation. Injected peptides were separated at a flow rate of 200 nL/min with a linear 80 min gradient from 100% solvent A (3% acetonitrile, 0.1% formic acid) to 30% solvent B (90% acetonitrile, 0.1% formic acid), followed by

a linear 6 min gradient from 30% solvent B to 90% solvent B. Each sample was run for 260 min, including sample loading and column equilibration times. The Q Exactive instrument was operated in the data-dependent mode acquiring HCD MS/MS scans (R=17,500) after each MS1 scan (R=70,000) on the 12 top most abundant ions using an MS1 ion target of 3x 10^6 ions and an MS2 target of 5x10^4 ions. The maximum ion time utilized for the MS/MS scans was 120 ms; the HCD-normalized collision energy was set to 25; the dynamic exclusion time was set to 20s, and the peptide match and isotope exclusion functions were enabled.

MS raw files were processed for protein identification and quantitation using MaxQuant v1.5.2.8, an open source academic software MSQuant (CEBI, open-source http://maxquant.org). The database used was Uniprot Human with a concatenated decoy database containing randomized sequences from the same database. Common contaminants like bovine serum albumin, trypsin etc. were also added to the database. Variable modifications used were oxidized methionine, arginine-13C6, lysine-13C6 15N2, and carbamidomethyl-cysteine was a fixed modification. The precursor mass tolerance used in the search was 7 ppm and fragment mass tolerance was 0.5 Da. Only proteins with more than two quantifiable peptides were included in our dataset. The protein ratios were median normalized. To assign interacting proteins we used the Limma package in the R environment to calculate moderated *t*-test *p*, as described previously[7].

Supplementary Note References

1. Van Nostrand, E.L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508-14 (2016).

2. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

3. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).

4. Engreitz, J.M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

5. McHugh, C.A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-6 (2015).

6. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-906 (2007).

7. Udeshi, N.D. *et al.* Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol Cell Proteomics* **11**, 148-59 (2012).