COUP-TFII regulates hemoglobin switching by activating the BCL11A-XL repressor LIN28B and directly binding δ and β globin promoters in fetal *versus* adult erythroid cells

Carlotta Frigo,[1]* Valentina Pastori,[1]* Gianluca Zambanini,[2,3]*° Martina Fabiano,[1] Sajeela Ahmed,[1] Elisabetta Citterio,[1,4] Claudio Cantù,[2,3,5] and Antonella Ellena Ronchi[1]

[1]Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca, Milan, Italy; [2]Wallenberg Centre for Molecular Medicine, Linköping University, Linköping, Sweden; [3]Division of Molecular Medicine and Virology, Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences, Linköping University, Linköping, Sweden; [4]Department of Life Science, Health, and Health Professions, LINK Campus University, Rome, Italy and [5]Science for Life Laboratory – SciLifeLab, Linköping University, 58185 Linköping, Sweden

CF, VP and GZ contributed equally as first authors.

°Current address: Max-Planck-Institut für molekulare Genetik, Berlin, Germany

Correspondence: A.E. Ronchi
antonella.ronchi@unimib.it

C. Cantù
claudio.cantu@liu.se

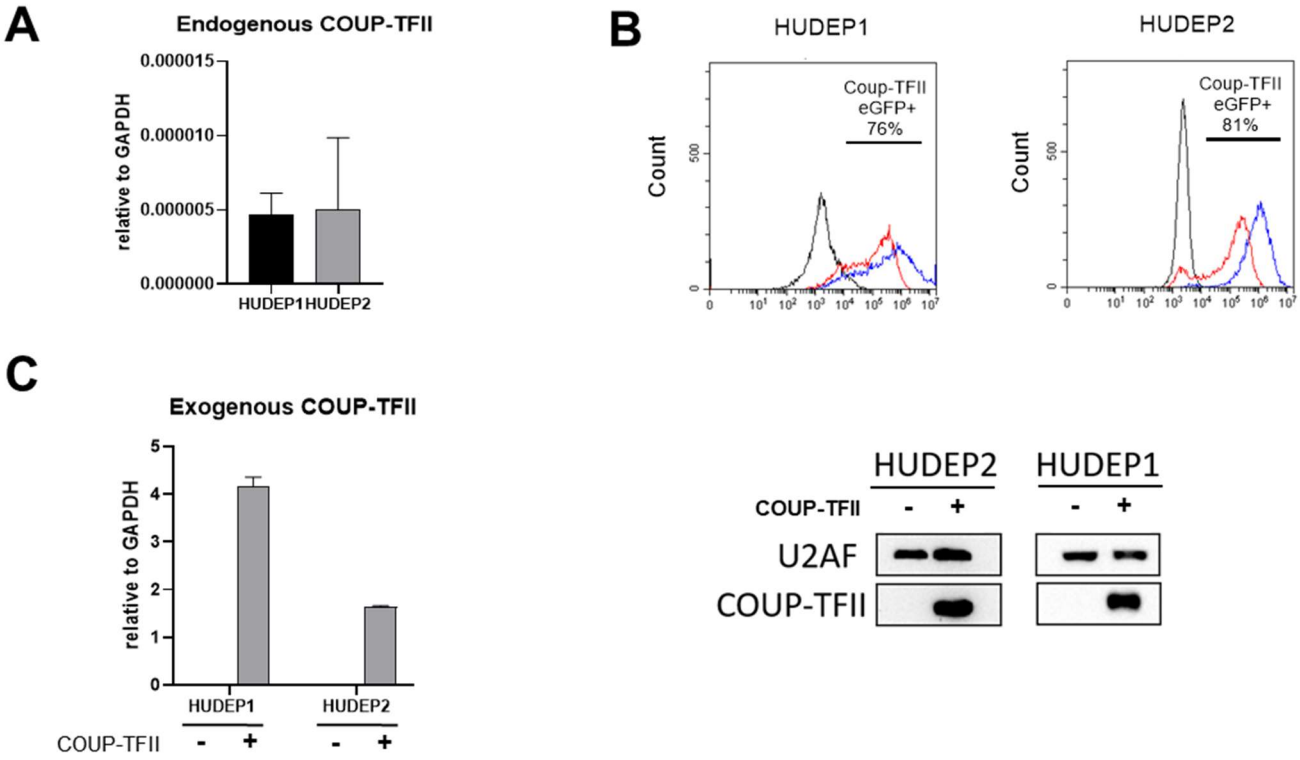
Received: June 17, 2025.
Accepted: November 17, 2025.
Early view: November 27, 2025.

https://doi.org/10.3324/haematol.2025.288485

COUP-TFII regulates hemoglobin switching by activating the BCL11A-XL repressor Lin28B and directly binding δ and β globin promoters in fetal versus adult erythroid cells *Carlotta Frigo et. al.*

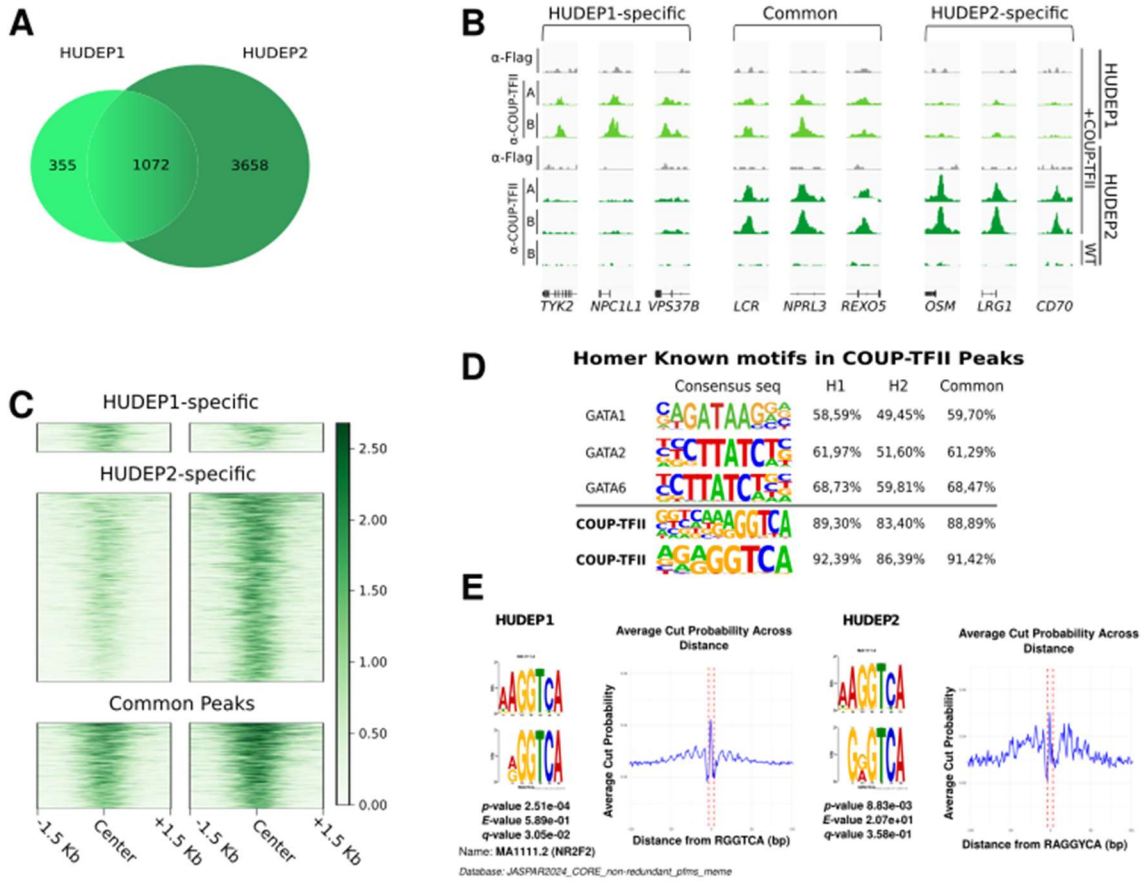
SUPPLEMENTAL FIGURES

SUPPLEMENTAL FIGURE 1



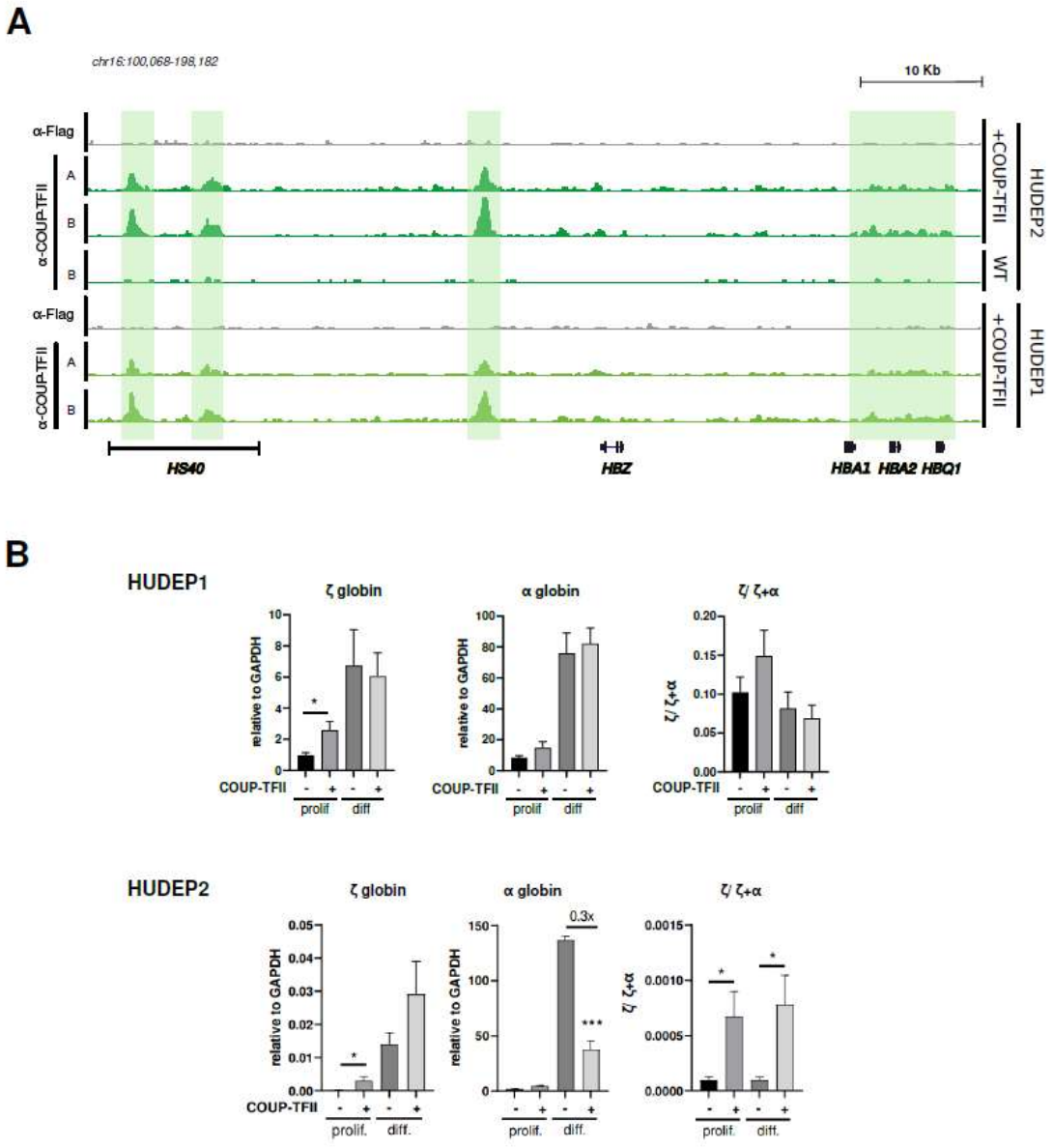
A) HUDEP1 and HUDEP2 cells do not express COUP-TFII. Endogenous COUP-TFII expression level evaluated by RTqPCR relative to GAPDH. **B) Representative transduction experiment.** The efficiency of HUDEP1 and HUDEP2 transduction was evaluated by Flow cytometry as percentage of GFP+ cells, since the expression vector contains a bicistronic Coup-TFII-IRES-GFP cassette. x axis: Mean Fluorescence Intensity; y axis: cell count. Gray: un-transduced cells. Blue: GFP^+ cells infected with the Empty Vector (-). Red: GFP^+ cells infected with the Coup-TFII-expressing vector (+). **C)** Corresponding exogenous Coup-TFII RNA (left) and protein (right) expression.

SUPPLEMENTAL FIGURE 2



COUP-TFII genomic occupancy in HUDEP1 and HUDEP2 cells. A) Venn diagram showing the overlap between COUP-TFII high-confidence peaks called in HUDEP1 and HUDEP2. **B)** IGV representation of three examples of COUP-TFII peaks in each category: HUDEP1-specific, common, HUDEP2-specific. Antibodies, cell lines, and treatment are indicated. Two anti-COUP-TFII antibodies, A (α-COUP-TFII A: ABIN6928040) and B (α-COUP-TFII B: ab41859) were used, each of them in two replicates and we considered high-confidence peaks those detected in at least 3 out of 4 replicates. The negative controls include an α-Flag antibody in COUP-TFII expressing cell lines and COUP-TFII antibody B (α-COUP-TFII B) in un-transduced HUDEP2 (WT). +COUP-TFII: cells transduced with the COUP-TFII expressing vector. **C)** Signal intensity plots summarizing the replicate average signal intensities across different genomic regions for common and unique COUP-TFII peaks in HUDEP1 (left) and HUDEP2 (right). **D)** Known motifs discovery by HOMER analysis within COUP-TFII peaks. Each motif, its likely binding factor and its frequency in HUDEP1 (H1), HUDEP2 (H2), or in shared peaks (common) are indicated. **E)** *De novo* motifs (corresponding to the COUP-TFII Jaspar Matrix MA1111.2) identified genome-wide by COUP-TFII antibodies in CUT&RUN experiments in HUDEP1 (left) and HUDEP2 (right). Average cut probability across the COUP-TFII site is plotted in the corresponding right panels.

SUPPLEMENTAL FIGURE 3



α-globin locus analysis. A) COUP-TFII CUT&RUN profile at the α-locus in HUDEP1 and HUDEP2 visualized in IGV. **B)** RT-qPCR analysis of mRNA levels for the indicated globin genes in HUDEP1 (upper panel) and HUDEP2 (lower panel) upon transduction with the empty vector (-) or with the COUP-TFII expressing vector (+) in proliferation and differentiation conditions. Expression levels are expressed relative to GAPDH and in relative ratios (n≥3; *=p≤0.05; **=p≤0.01; ***=p≤0.001)

SUPPLEMENTAL METHODS

Library preparation

Library preparation was performed using the KAPA Hyper Prep Kit for Illumina platforms (Cat. #KK8504, KAPA Biosystems) according to the manufacturer's guidelines with the following modifications. End repair and A-tailing was performed with 0.4x reactions with 20 µl of purified DNA. The thermocycler conditions were set to 12 °C for 15 min, 37 °C for 15 min and 58 °C for 25 min to prevent thermal degradation of the shortest fragments. Adapter ligation was done with 0.4x reactions. KAPA Dual Indexed adapters were used at 0.15 µM. A post-ligation clean-up was performed with Mag-Bind TotalPure NGS beads at 1.2x. Resuspension was done in 10 mM Tris-HCl pH 8.0. Library amplification was performed with 0.5x reactions. The thermocycler was set with the following conditions: initial denaturation at 98 °C for 45 sec, denaturation at 98 °C for 15 sec, annealing/elongation at 60 °C for 10 sec, final extension at 72 °C for 1 min, hold at 4 °C, with 13 cycles. After amplification, a post-amplification cleanup was performed with 1.2x beads. Libraries were then run on an E-Gel EX 2% agarose gel (Cat. #G402022, Invitrogen) for 10 min using the E-Gel Power Snap Electrophoresis System (Invitrogen). Bands of interest between 150 and 500 bp were cut out and purified using the QIAquick Gel Extraction Kit (Cat. #28706, QIAGEN) according to manufacturer's instructions. Libraries were quantified with the Qubit (Thermo Scientific) using their high sensitivity DNA kit (Cat #Q32854, Thermo Scientific), pooled and sequenced 36 bp pair-end on the NextSeq 550 (Illumina) using the Illumina NextSeq 500/550 High Output Kit v2.5 (75 cycles) (Cat. #20024906, Illumina). The CUT&RUN datasets (raw and processed files) have been deposited at ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-14720.

Cell line	Condition	Antibody
HUDEP1	COUPTFII vector	Flag F7425
HUDEP1	COUPTFII vector	COUPTFII A ABIN6928040 (Lot HD09MA0327-B)
HUDEP1	COUPTFII vector	COUPTFII B ab41859
HUDEP2	WT	COUPTFII B ab41859
HUDEP2	COUPTFII vector	Flag F7425
HUDEP2	COUPTFII vector	COUPTFII A ABIN6928040 (Lot HD09MA0327-B)
HUDEP2	COUPTFII vector	COUPTFII B ab41859

Data Analysis

Trimming was performed using bbmap bbduk(1) (version 39.0) removing adapters, artifacts, poly AT, G and C repeats. Reads were aligned to the hg38 genome with bowtie (2) (version 1.3.1) using options -v 0 -m 1 -X 500. Samtools(3) (version 1.6) view, fixmate, markdup and sort were used to create bam files, mark and remove duplicates, and sort bam files. Mitochondrial reads were removed together with the problematic of problematic regions as described in ref. (4). Individual track bedgraphs were created using bedtools(5) (version 2.30.0) genomecov on pair-end mode. Normalized signal per million reads tracks for visualization were created by using the –SPRM function of macs2(6) (version 2.2.6) for each replicate with the options -f BAMPE –SPMR and –bdg. After normalization 2 replicate per Antibody were averaged out using the bigwigAverage function of Deeptools(7) (version 3.5.2). Peaks were called using SEACR(8) (version 1.3) against the corresponding negative control using the options “norm, stringent”. Venn diagrams and overlap peak sets were created using Intervene(9) (version 0.6.5). Motif analysis was done using Homer(10) (version 4.11) findMotifsGenome to find motifs in the hg38 genome using -size given. Peak set gene annotation was done using GREAT(11) (version 4.0.4) with default parameters. Signal intensity plots were created using the Deeptools(7) (version 3.5.2) functions computeMatrix reference-point, with the parameters -H chromosome length, -referencePoint center, -a 1500, -b 1500 and plotHeatmap. Footprint analysis: to identify potentially enriched sequences within the CUT&RUN peaks, we employed MEME Suite(18) (version 5.5.6) for de novo motif discovery. Sequences spanning from -100 bp to +100 bp relative to the intermediate positions of 1000 randomly selected peaks were analyzed. The top 10 enriched motifs were extracted using DREME (-dreme-m 10). Motifs were annotated using tomtom with the JASPAR database. The sequences containing the motifs where then extracted using FIMO and the cut frequency matrix was computed via CUT-RUNTools-2.0 package(12) with make_cut_matrix (version 0.1.6). The variation in cut probabilities relative to the distance from the motif center was visualized using R (version 4.3.3). Single locus footprinting analysis was performed adapting get_cuts_single_locus.sh, also from CUT-RUNTools-2.0 package.

Comparison between the CUT&RUN occupancy of BCL11A-XL and COUP-TFII.

For the comparison with BCL11A-XL we downloaded raw data from Liu N. et al. (13) (accession number GEO: GSE104676) and we aligned them with Bowtie2, as described in the paper with the following modification: all tracks were filtered the suspect of problematic regions as

described in ref.(4). We selected tracks for 2 BCL11A-XL replicates in WT and to have a fair comparison we realigned our data (only antibody B) together with the BCL11A-XL dataset.

Sample/Dataset	Cell line and condition	Antibody
SRR6144299	HUDEP2	BCL11A
SRR6144304	HUDEP2	BCL11A

For visualization purposes the 2 replicate per condition were averaged out using the bigwigAverage function of Deeptools (Ramírez et al., version 3.5.2).

1. Bushnell B, Rood J, Singer E. BBMerge - Accurate paired shotgun read merging via overlap. PloS one. 2017;12(10):e0185056.
2. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
3. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
4. Nordin A, Zambanini G, Pagella P, Cantu C. The CUT&RUN suspect list of problematic regions of the genome. Genome Biol. 2023;24(1):185.
5. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
6. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
7. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5.
8. Meers MP, Tenenbaum D, Henikoff S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. Epigenetics Chromatin. 2019;12(1):42.
9. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics. 2017;18(1):287.
10. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, 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):576-89.
11. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495-501.
12. Yu F, Sankaran VG, Yuan GC. CUT&RUNTools 2.0: a pipeline for single-cell and bulk-level CUT&RUN and CUT&Tag data analysis. Bioinformatics. 2021;38(1):252-4.
13. Liu N, Hargreaves VV, Zhu Q, Kurland JV, Hong J, Kim W, et al. Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell. 2018;173(2):430-42 e17.