SUPPLEMENTARY APPENDIX

Integrated analyses of translatome and proteome identify the rules of translation selectivity in *RPS14*-deficient cells

Ismael Boussaid,[1] Salomé Le Goff,[1,2,*] Célia Floquet,[1,*] Emilie-Fleur Gautier,[1,3] Anna Raimbault,[1] Pierre-Julien Viailly,[4] Dina Al Dulaimi,[1] Barbara Burroni,[5] Isabelle Dusanter-Fourt,[1] Isabelle Hatin,[6] Patrick Mayeux,[1,2,3,#] Bertrand Cosson[7,#] and Michaela Fontenay[1,2,3,4,8]

[1]Université de Paris, Institut Cochin, CNRS UMR 8104, INSERM U1016, Paris; [2]Laboratoire d'Excellence du Globule Rouge GR-Ex, Université de Paris, Paris; [3]Proteomic Platform 3P5, Université de Paris, Paris; [4]Centre Henri-Becquerel, Institut de Recherche et d'Innovation Biomedicale de Haute Normandie, INSERM U1245, Rouen; [5]Assistance Publique-Hôpitaux de Paris, Centre-Université de Paris - Cochin, Service de Pathologie, Paris; [6]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université de Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette Cedex; [7]Université de Paris, Epigenetics and Cell Fate, CNRS UMR 7216, Paris and [8]Assistance Publique-Hôpitaux de Paris, Centre-Université de Paris - Hôpital Cochin, Service d'Hématologie Biologique, Paris, France

**SLG and CF contributed equally to this work.*

#PM and BC contributed equally as co-senior authors.

 doi:10.3324/haematol.2019.239970

Received: October 7, 2019.

Accepted: March 26, 2020.

Pre-published: April 23, 2020.

Correspondence: *MICHAELA FONTENAY* - michaela.fontenay@inserm.fr

Supplemental Material

Boussaid et al

Supplemental methods

Human $CD34^+$ progenitors were purified on MidiMacs system (Miltenyi Biotech, Bergisch Gladbach, Germany) and cultured in Iscove's modified Dulbecco medium (IMDM, Life Technologies, Grand Island, NY) supplemented with 2 mM L-glutamine, 100 U/mL penicillin/streptomycin (Life Technologies), and 15% bovine serum albumin insulin transferrin (BIT 9500) serum substitute (StemCell Technologies, Vancouver, Canada). For amplification, $CD34^+$ progenitors were cultured with 1U/ml interleukin-3 for 3 days, and with 100 ng/mL stem cell factor and 10 ng/mL interleukin-6 for 6 days. Recombinant EPO at 1U/ml was then added to the medium for the differentiation phase. The UT-7/EPO erythro-megakaryocytic cells were cultured in αMinimum Essential Medium (αMEM) containing 10% fetal calf serum (FCS), 1mM glutamine, 100UI/mL penicillin-streptomycin and 1 UI/mL EPO. K562 human erythroid cells (ATCC) were maintained at a density between 0.1 x 10^6 and 1 x 10^6 cells per milliliter in RPMI 1640 medium supplemented with 10% FCS and 1mM glutamine, 100UI/mL penicillin-streptomycin. Cytokines were purchased from Miltenyi Biotech and chemical products from Sigma Aldrich (Saint-Quentin Fallavier, France). Cells were incubated at 37°C with 5% CO2.

Lentiviral vector

For UT-7 and K562 cell line infection, a scramble (SCR) control shRNA and *RPS14* shRNA (640 and 641) were cloned in pLKO.1 Tet-On vector. Cells were selected with puromycine (1 mg/mL). shRNA expression was induced by doxycycline (0.2 µg/mL) for 3 days (UT7) or 5 days (K562). For primary cell infection, sh*RPS14* (640 and 641) and shSCR were cloned in the non-inducible pLenti X1 vector encoding green fluorescent protein (GFP). Cells were sorted on GFP expression at 3 days after infection and harvested.

Quantitative proteomics

Peptides were prepared from whole-cell lysates, using the filter-aided sample preparation (FASP) method. In a first experiment, peptides were pre-fractionated by strong cationic exchange (SCX) StageTips.[1] In the subsequent experiments, peptides were analyzed without prefractionation since the study did not require the quantification of proteins expressed to very low levels. Mass spectrometry (MS) analyses were performed on a Dionex U3000 RSLC nano-LC-system coupled to Orbitrap Fusion (Thermo Fisher Scientific, Waltham, MA). Peptides were separated on a C18 reverse-

phase resin (75-mm inner diameter and 15-cm length) with a 3 h gradient. The mass spectrometer acquired data throughout the elution process and operated in a data-dependent scheme.

Polysome profiling

At respectively 3 and 5 days post-induction, 30 millions of UT-7/EPO and K562 cells were incubated with 100µg/ml cycloheximide for 10 min at 37°C and, washed, and then pelleted by centrifugation. Lysis was obtained by vortexing glass beads (diameter 0.25-0.5 mm; ref. A553.1; Carl Roth GmbH, Karlsruhe, Germany) in 50 mM Tris-HCl pH 7.4, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 100 µg/mL cycloheximide, and 200 U/mL RNasine (Promega) in RNase free H_20. Total RNA was quantified by spectrophotometry at 260 nm in the supernatants collected by centrifugation for 5 min at a 16,000 rcf. The samples were loaded onto a 10 to 50% sucrose gradient in 50mM Tris-acetate pH 7.6, 50mM NH_4Cl, 12mM $MgCl_2$, 1mM DTT and centrifuged at 36,000 rpm for 3 h (SW41 rotor in XE-90 ultracentrifuge, Beckman Coulter Miami, FL). Fractions were collected and their absorbance was measured using a UA-6 detector (Teledyne ISCO, Lincoln, NE). For ribosomal fractions, one volume of Trizol (Ambion) was added to the sucrose fractions. After vortexing, chloroform (200 µL) was added and the mixture was centrifuged for 10 min at 15000 rpm. RNA was extracted and recovered in RNase free H_2O and a 1 µL aliquot was reverse-transcribed using M-MLV Reverse Transcription kit (Invitrogen, #28025-013).

Oligonucleotide Microarrays: transcriptome and translatome analysis

Briefly, 18S and 28S RNA quality was checked using a Bioanalyzer (Agilent). Linear amplification of 20 ng of total RNA was performed using the Ovation Biotin RNA Amplification and Labelling System (NuGEN, San Carlos, CA). The resulting fragmented, labeled cDNA was hybridized to oligonucleotide microarrays. Raw fluorescence intensity values were normalized using Robust Multi-array Average (RMA) algorithm in R to generate the normalized data matrix by performing background correction, quantile normalization and $\log_2$ transformation of raw fluorescence intensity values of each gene. Data were normalized using custom brainarray CDF files (v20 ENTREZG). Heatmaps were drawn using the Morpheus software (http://www.broadinstitute.org/morpheus/). Gene set enrichment analysis (GSEA) was performed on normalized data matrix deriving from the translatome and transcriptome using the GSEA desktop application® v.2.0.12 (www.broadinstitute.org/gsea). Enrichment rates were considered significant when the *P*-value < 0.05 and the FDR ≤ 0.05. *GATA1* target gene set derived from integrated ChIP-seq and transcriptomic data deposited in the Gene Expression Omnibus (GEO) database under access numbers GSE16594 and GSE18042.[2] Genes bound by GATA1 and expressed with a fold change > 2 or < -2 with a *P*-value < 0.05 after 14h or 30h were considered for the enrichment analysis. For Network integration analysis, data were imported into GSEA software to

generate enrichment scores for gene sets in Hallmark, C2.all MSigDB datasets. Cytoscape (v3.2.1) and the Enrichment Map plug-in was used to generate networks for gene sets enriched with a FDR cut-off of < 0.1. Nodes represent the gene sets that were enriched at the top or bottom ranking of the differentially expressed genes. The node size corresponds to the number of genes in the set. Edges indicate overlaps between gene sets and the line thickness indicates the similarity coefficient between gene sets in red for the less affected genes and blue for downregulated genes. The node color depends on the FDR value as shown on the scale. Clusters of nodes are determined with Autoannotate App and denoted by the circles. The datasets generated and analyzed are provided in the Supplemental data.

Codon usage, uORFs and structure prediction analysis

To calculate the Codon adaptation index, a human codon usage table has been calculated from a set of 19250 human genes from the Ensembl database (Release 57).[3] Only genes with UniProtKB/SwissProt ID have been included. In cases of alternative splicing variants, only the longest splice variant has been included. Genes encoding proteins of length shorter than 99 amino acids have been also removed. Average of Relative Synonymous Codon usage (RSCU) were calculated for 19,000 human transcripts. A list of 1094 transcripts with uORF were selected from Van Heesch *et al* study.[4] This list corresponds to 920 unique genes of which 908 were detected in our translatome dataset.

Transcript properties and thermodynamic characteristics (free folding energy ΔG, UTR and CDS length) were collected on UCSC database and obtained from the RNAfold program ViennaRNA Package.[5] A highly structured sequence was defined as a sequence with strongly negative energy of folding in $Kcal.mol^{-1}$. To validate this base pair probabilities approach we also used the CROSS prediction of structuration method based on artificial neural network to calculate the structural profile of an RNA sequence at single-nucleotide resolution and without sequence length restrictions. This method takes advantage of recent approaches exploiting biochemical reactions to perform high-throughput profiling of the RNA structure: Parallel Analysis of RNA Structure (PARS) Selective 2 - Hydroxyl Acylation analyzed by Primer Extension (icSHAPE) and NMR/X-ray structures.[6] With the CROSS algorithm we determined the structuration propensity “Global-Score” along 5’UTR CDS and 3’UTR of the 100 transcripts with the most upregulated or downregulated TE.

mRNA stability datasets

mRNA half-life was obtained from Time-lapse-Seq in K562 (GSE95854) from Schofield et al., study.[7] mRNA decay rates was obtained from SLAM-Seq in K562 from Wu et al., 2019 study.

Ribosome profiling data analysis

We took advantage of two published datasets. First, a dataset of DBA model was obtain from GSE89183. We used the processed count table to determine differentially translated (e.g., changes in TE) genes. We performed the same analytical method indicated in Khajuria et al., 2018 study: We used Xtail (https:// github.com/xryanglab/xtail), which first uses the negative binomial distribution to estimate either the log2 fold changes separately for mRNAs and RPFs between conditions (i.e., ΔmRNA and ΔRPF) or the log2 fold changes for mRNA to RPF within conditions (i.e., $TE_{control}$ and TE_{RPH}), and then estimates a discrete joint probability distribution of either ΔmRNA and ΔRPF or $TE_{control}$ and TE_{RPH}. Testing of differential translation (i.e., ΔTE in both cases) was then performed, the less significant result of the two methods was kept. The Benjamini–Hochberg FDR was used to control for multiple testing. Only genes with > 150 mRNA counts and > 90 RPF counts were analyzed in order to obtain more stable estimates of ΔTE. The second dataset was obtained from GSE85864. This ribosome profiling was carried out as described in Ingolia et al., 2012 and Mills et al., 2016 studies and have been used to determine the most translated genes in reticulocytes.

Luciferases constructs and assays

The Firefly luciferase of the pmirGlo vector (Promega) were switched for the Firefly luciferase of the pCINeoFLuc vector using primers, ApaI and SalI restriction sites and the NEBuilder HiFi DNA Assembly Cloning Kit (NEB, Ipswich, MA). Constructions were controlled using restriction enzymes SacI, KpnI, XbaI, and EcoRV and size verified by gel electrophoresis, and using dideoxy sequencing. All primers and sequences are listed in Table S4. At day 3 post doxycycline, these vectors were transfected in K652 shSCR sh640 and sh641 conditions using lipofectamine 3000 following the manufacturer protocol. At day 5 post doxycycline, K562 cells were collected 48 h after transfection and assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) and a Multilabel Plate Reader Fluostar Optima (BMG Labtech, Ortenberg). Firefly luciferase activity was normalized against Renilla luciferase activity or inversely in each sample.

Flow cytometry

The expression of glycophorin A (GPA, CD235a; clone GA-R2, BD Biosciences, Franklin Lakes, NJ, USA) or c-Kit (CD117; clone YB5.B8, BD Bioscience) was followed and quantified on a BD LSRFortessa™ cytometer (BD Biosciences). Results were analyzed with Kaluza software 1.2 (Beckman Coulter, Miami, FL, USA). For cell cycle experiments, iodure propidium-positive cells were quantified on a Flow cytometer BD Accuri™ C6 and data were analyzed with the use of CFlow software (BD Biosciences). Apoptosis was determined by assaying PE-coupled annexin V (Ref 556421, BD Biosciences), quantified on a Guava cytometer (easyCyte™ Merck Millipore, Darmstadt, Germany),

and data were analyzed using Cytosoft 5.1 (Guava Technologies Inc., Hayward, CA). Click-iT® Plus OPP Protein Synthesis Assay Alexa Fluor™ 647 was performed following the manufacturer protocol 3 days after induction by doxycycline in shSCR and sh*RPS14* samples and in a control shSCR+ 30min CHX. Data were analyzed using Cytosoft 5.1 (Guava Technologies Inc., Hayward, CA)

Western blot

UT7/EPO cells (5.10^5), were lyzed in 50µL 1X LaemmLi buffer [62.5 mM Tris HCl pH 6.7, 10%glycerol, 2% sodium dodecylsulfate (SDS), 40 mM dithiotreitol (DTT), bromophenol blue], heated at 90°C for 5 min and resolved on a 12.5% SDS-polyacrylamide gel. Protein signals were revealed by chemoluminescence (ECL Biorad) and detected using a CCD camera (LAS 3000 Fujifilm). Primary antibodies are listed in the table S4.

Immunofluorescent labelling

Cells were fixed for 20 min in 2% paraformaldehyde, permeabilized for 5 min with 0.5% Triton X100 in PBS (pH 7.4), saturated for 30 min in saline sodium citrate solution containing 3% BSA, 0.1% Tween-20 and incubated with primary antibodies for 1h at room temperature and then with a secondary fluorochrome-coupled antibody for 45 min. After dehydratation in ethanol, nuclei were stained with diaminidophenylindol (DAPI; Sigma-Aldrich, Saint-Louis, MI). Slides were sealed in Fluoromount-G medium (SouthernBiotech, Birmingham, AL). Images were obtained on a Leica DMI6000 inverted microscope with spinning disk (Yokogawa, Tokyo, JP) and analyzed with Image J or 3D Imaris (Oxford Instruments, Abingdon-on-Thames, UK).

Immunohistochemistry

Fixed bone marrow biopsy from del5q or control patients were dehydrated and embedded in paraffin for sectioning. Paraffin sections 3 µm in thickness were prepared. For immunohistochemistry, anti-GATA1 (dilution 1:100, Cell signaling 3535) was used, with counterstaining with hematoxylin and eosin (Leica Autostainer XL). Light microscopy was performed with Lamina slide scanner, Perkin Elmer.

RT-qPCR

RNA was extracted using the Kit QIAamp® RNA Blood MiniKit. Reverse transcription was performed using the Maxima First Strand cDNA Synthesis Kit Thermo Scientific™ on an Applied Biosystems 2720® thermocycler (25°C 10 min, 50°C 15 min, 85°C 5 min). PCR was subsequently performed in triplicate using FastStart Universal SYBR Green Master (Roche). An amplification efficiency >95% was obtained after 40 cycles (95°C, 55°C, 72°C) on a LightCycler480 (Roche Molecular Diagnostics,

Pleasanton, CA). For quantification, CP values were used to calculate the normalized ratio quantities (NRQ) according to the formula: NRQ=RQ/NF with RQ being the $E^{\Delta CP}$ where ΔCP is the difference between the CP of the gene of interest and the mean CP of this gene in all samples, E is the primer efficacy and NF is the mean of RQ of the reference genes. For ribosomal fractions, RNA levels were summed across all fractions and presented as percentage of this total. The primers for RT-qPCR used are listed in the table S4.

Supplemental references

1. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods 2014;11(3):319–324.
2. Yu M, Riva L, Xie H, et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol Cell 2009;36(4):682–695.
3. Sharp PM, Li WH. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15(3):1281–1295.
4. van Heesch S, Witte F, Schneider-Lunitz V, et al. The Translational Landscape of the Human Heart. Cell 2019;178(1):242-260.e29.
5. Lorenz R, Bernhart SH, Höner zu Siederdissen C, et al. ViennaRNA Package 2.0. Algorithms Mol Biol 2011;6(1):26.
6. Delli Ponti R, Marti S, Armaos A, Tartaglia GG. A high-throughput approach to profile RNA structure. Nucleic Acids Res 2017;45(5):e35.
7. Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat Methods 2018;15(3):221–225.

Integrated analyses of translatome and proteome identify the rules of translation selectivity in *RPS14*-deficient cells.

List of supplemental Tables

Table S1: Data relative to translatome, transcriptome and proteome analysis. Sheet 1: Translatome, transcriptome and proteome datasets, related to figures 1,2,3,6. **Sheet 2**: List of genes which are significantly down regulated in the transcriptome analysis, related to Figures 2 and 3. **Sheet 3**: List of genes which are significantly up regulated in the transcriptome analysis, related to Figures 2 and 3. **Sheet 4**: List of genes which are significantly up regulated in the translatome analysis, related to Figures 2 and 3. **Sheet 5**: List of genes which are significantly down regulated in the translatome analysis, related to Figures 2 and 3. **Sheet 6**: List of genes up regulated in the analysis of FC(translatome)/(transcriptome), related to Figures 2 and 3. **Sheet 7**: List of genes down regulated in the analysis of FC(translatome)/(transcriptome), related to Figures 2 and 3. **Sheet 8**: Post-transcriptionally up regulated genes (Proteome/translatome >1.5), related to Figure 6. **Sheet 9**: Post-transcriptionally down regulated genes (Proteome/translatome < 1/1.5) with RPs, related to Figure 6. **Sheet 10**: Post-transcriptionally down regulated genes (Proteome/translatome < 1/1.5) without RPs, related to Figure 6. **Sheet 11**: Levels of Ribosomal Proteins in sh*RPS14* vs shSCR conditions, related to **Figure 1**. **Sheet 12**: Raw data of proteomic analysis for M1 and M2 replicates, related to Figures 1 and 6. **Sheet 13**: Raw data of proteomic analysis for M0 replicates, related to Figures 1 and 6. **Sheet 14:** Translatome dataset of K562 cells (sh*RPS14* 640 641 and shSCR), related to Figure S3. **Sheet 15**: Analysis mix(sh*RPS19*/*RPL5*) vs shluc in human primary CD34 cells in erythroid differentiation (Khajuria et al., 2018) related to Figure S3. **Sheet 16**: Analysis of sh*RPS19* vs shluc human primary CD34 cells in erythroid differentiation (Khajuria et al., 2018) related to Figure S3. **Sheet 17**: Analysis of sh*RPL5* vs shluc human primary $CD34^+$ cells in erythroid differentiation (Khajuria et al., 2018) related to Figure S3. **Sheet 18**: List of uORF used in the cumulative frequency analysis (Van Heesch et al., 2018) related to Figure S5. **Sheet 19**: List of mRNA half-life and decay rate used in the cumulative frequency analysis (Wu et al., 2019) related to Figure S5.

Table S2: GATA1 targets used as geneset for GSEA.

Table S3: Data relative to GSEA, Cytoscape and GO enrichment analysis.

Table S4: List of primers and sequences used. Sheet 1: qPCR Primers and Westernblot antibodes related to **Figures 1 and S1**. **Sheet 2** Luciferase sequences related to **Figures 4 and S4**. **Sheet 3** Overlapping primers, luciferase sequences used for constructs related to **Figures 4 and S4**.

List of supplemental Figures

Figure S1: **A RPS14 downregulation results in limited ribosome availability and reduced GATA1 translation**. Related to Figure 1. (**A-C**) Knockdown of RPS14 by shRNA in human primary erythroblasts of healthy donors. (**A**) Glycophorin A (GPA)+ cells by flow cytometry expressed as mean percentages ± SEM (n=3). (**B**) Immunofluorescence detection of GATA1 in cultured erythroblasts at day 10. Nuclei were counterstained with DAPI. Original magnification, X63. (**C**) GATA1 mRNA level determined by MILE transcriptomic analysis (GSE15061) of bone marrow mononuclear cells from 5q- syndrome patients (n=11) and healthy controls (n=69). (**D-J**) UT-7/EPO cell lines were treated or not with doxycycline for 3 days to induce sh*RPS14* 640, sh*RPS14* 641 or shSCR expression. (**D**) Quantification of *GATA1* transcripts by RT-qPCR expressed as the normalized ratio quantities (NRQ) to *UBC* and *B2M* shown as means ± SEM (n=3). (**E**) Proliferation curves of UT7/EPO cells during 3 days post-induction. (**F-H**) Flow cytometry.(**F**) Cell cycle analysis expressed as percentages ± SEM of cells in G1 phase (n=3). (**G**) Apoptosis expressed as mean percentages ± SEM of annexin V-positive cells (n=3). (**H**) GPA expression shown as mean fluorescence intensity ± SEM (n=3). (**I**) Transcriptomic analysis of UT-7/EPO sh*RPS14 versus* UT-7/EPO shSCR cell lines. *GATA1* target transcript quantities expressed as fold-change (FC) in sh*RPS14* conditions relative to shSCR in microarrays. (**J-K**) Quantification of GYPA and KIT transcripts by RT-qPCR expressed as the NRQ to UBC and B2M in means ± SEM (n=3). (**L-Q**) K562 cell lines were treated or not with doxycycline for 5 days to induce sh*RPS14* 640, sh*RPS14* 641 or shSCR expression. (**L**) Schematic workflow. (**M**) Quantification of *GATA1* transcripts by RT-qPCR expressed as the NRQ to *UBC* and *B2M* shown as means ± SEM (n=3). (**N**) GATA1 protein expression determined by western blot. (**O**) Polysome profiling. (**P-Q**) Relative distribution of GATA1 and MYC mRNA in ribosomal fractions. Quantification by RT-qPCR is expressed as mean ± SEM (n=3). * for $p<0.05$; ** for $p<0.01$; *** for $p<0.001$.

Figure S2: **Selective translation in UT-7/EPO shRPS14 cell lines and correlations with published datasets**. Related to Figure 2. (**A**) The GO terms with a p-value < 0.01 were considered in the analysis. (**B**) Correlations between translational changes induced by shRPL5, shRPS19, shRPS14 in human primary erythroblasts or UT-7/EPO cells. Linear fits are shown in red. Pearson correlations are indicated.

Figure S3: CAI, Transcript length and 3'UTR structure explain the translation rules in limiting ribosome availability conditions. Related to Figure 3. **(A-B)** Thermodynamic landscape: Bi-parametric scatter-plots of the 5'UTR and 3'UTR fold energy, fold energy per base and length parameters of

every transcripts of the UCSC database (grey dots). **(C-D)** Monoparametric representation of energy per base, length of the 5'UTR, 3'UTR and CDS+UTRs expressed as means ± SD in up or downregulated transcripts in the transcriptome and translatome analysis. **(E)** Monoparametric representation of transcript length and energy per base of the 5'UTR and 3'UTR expressed as means ± SD in up or downregulated transcripts in the ΔTE analysis. **(F)** Monoparametric representation of energy per base of the 5'UTR and 3'UTR, transcript length and CAI expressed as means ± SD in up or downregulated transcripts in the translatome analysis. The p-values were calculated using two-sided Mann–Whitney U-test. (**G**) Annotation of CDS and UTRs characteristics of transcripts translationally up and down regulated in other models of ribosomopathies: shRPL5/shRPS19 human primary cells (GSE89183). Thermodynamic landscape: Bi-parametric scatter-plots with energy per base on y axis and length on x axis in a log scale of 5'UTRs and 3'UTRs. Bi-parametric scatter-plot with CAI on x axis and total length of the transcript on y axis. All transcripts described in the UCSC database are represented by grey dots. Red dots for the top 300 upregulated transcripts and blue dots for the top 300 downregulated transcripts in ribosome profiling experiments. (**H**) CAI, Transcript length and 3'UTR structure explain the translation rules in limiting ribosome availability conditions. Bi-parametric scatter-plots of these different characteristics. Grey points represent all UCSC transcripts. Red and blue points represent top 300 up and down regulated transcripts at translational level in our models and in published models. (**I**) Observed differences between SSU and LSU targeting are, in part, linked to the 5'UTR structuration. Dot plots representing the 5'UTR energy per base of differential targets between SSU and LSU targeting. The p-values were calculated using two-sided Mann–Whitney U-test. * P <0.05; ** P <0.01; *** P <0.001; **** P <0.0001.

Figure S4: Validation of codon bias as a determinant of translation selectivity. Related to Figure 4. **A.** Average of Relative Synonymous Codon Usage (RSCU) in all human transcripts ((grey dots) calculated in 19,000 transcripts), in top 100 transcripts with increased ΔTE (red dots) and in top 100 transcripts with downregulated ΔTE (blue dots) in UT-7/EPO model. Codons and (amino acids) on X axis were arranged based on the ascending mean RSCU values in all human transcripts. In this metric, a RSCU value equal to 1 means no codon bias. **B.** Constructs with low CAI and high CAI Firefly in pmirGLO vector backbone **C.** Firefly luciferase (Fluc) and Renilla luciferase (Rluc) signals expressed as relative light unit means ± SEM (n=9).

Figure S5: **Limited ribosome availability slightly up-regulates uORF translation efficiency and has no impact on mRNA stability.** Related to Figure 5. (**A**) Cumulative frequency representation of translation efficiency changes of transcripts with uORF (**B**) Correlation between mRNA half-life and decay rate with CAI and 3'UTR $\Delta G.base^{-1}$. (**C**) Correlation between mRNA half-life and decay rate

with Log_2 (FC) transcriptome of UT7/EPO model. Linear regression (with confidence intervals) are shown in dashed lines. Pearson correlations are indicated. mRNA half-life was obtained from Time-lapse-Seq in K562 (GSE95854). mRNA decay rates was obtained from SLAM-Seq in K562 (GSE126523).

Figure S6: Proteome analyses confirm identified rules. Related to Figure 6. (**A**) Scatter-plot showing the 3'UTR thermodynamic characteristics of each transcript encoding ribosomal proteins (cyan for 40S proteins, purple for 60S proteins) in the thermodynamic landscape (grey). (**B**) Characterization of transcripts translationally up and down regulated in a lymphoid model of RPS15 mutated CLL .Monoparametric representation of energy per base of 3'UTR and CAI expressed as means ± SD. The p-values were calculated using two-sided Mann–Whitney U-test.**P<0.01.

Figure S7: Codon Adaptation Index, coding sequence length and thermodynamic characteristics of UTRs are key determinants of translation in normal erythropoiesis. Related to Figure 7. Bi-parametric scatter-plots of 5'UTR and 3'UTR structure, CAI and total length (CDS+UTRs) of every transcript of the UCSC database (grey dots). Top 15 protein most expressed in red blood cells are overlaid in green (Gautier et al., PXD009258).

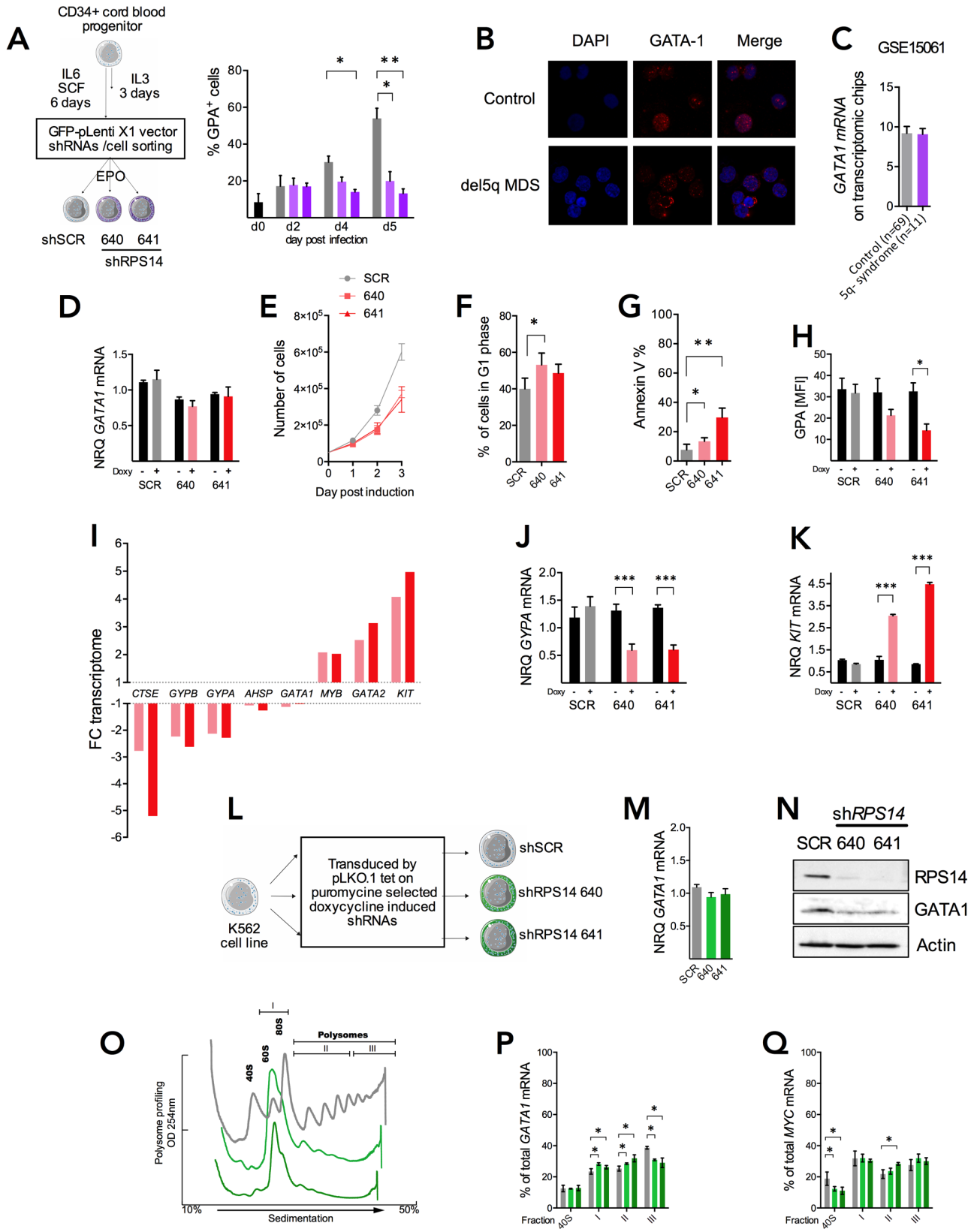
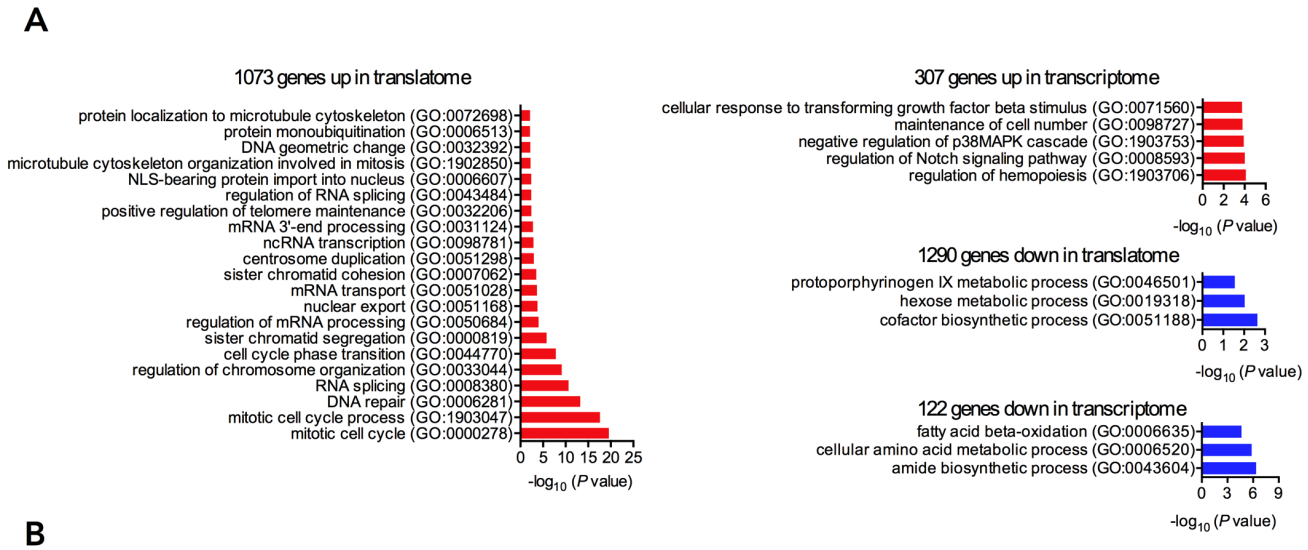

A
CD34+ cord blood progenitor
IL6 SCF 6 days
IL3 3 days
GFP-pLenti X1 vector shRNAs /cell sorting
EPO
shSCR
640
641
shRPS14
% GPA+ cells
day post infection
B
DAPI
GATA-1
Merge
Control
del5q MDS
C
GSE15061
GATA1 mRNA on transcriptomic chips
Control (n=69)
5q- syndrome (n=11)
D
NRQ GATA1 mRNA
Doxy
SCR
640
641
E
SCR
640
641
Number of cells
Day post induction
F
% of cells in G1 phase
G
Annexin V %
H
GPA [MFI]
I
FC transcriptome
CTSE
GYPB
GYPA
AHSP
GATA1
MYB
GATA2
KIT
J
NRQ GYPA mRNA
K
NRQ KIT mRNA
L
K562 cell line
Transduced by pLKO.1 tet on puromycine selected doxycycline induced shRNAs
shSCR
shRPS14 640
shRPS14 641
M
NRQ GATA1 mRNA
N
shRPS14
SCR 640 641
RPS14
GATA1
Actin
O
Polysome profiling OD 254nm
40S
60S
80S
Polysomes
I
II
III
10%
50%
Sedimentation
P
% of total GATA1 mRNA
Fraction
40S
Q
% of total MYC mRNA

Figure S1


A
1073 genes up in translatome
protein localization to microtubule cytoskeleton (GO:0072698)
protein monoubiquitination (GO:0006513)
DNA geometric change (GO:0032392)
microtubule cytoskeleton organization involved in mitosis (GO:1902850)
NLS-bearing protein import into nucleus (GO:0006607)
regulation of RNA splicing (GO:0043484)
positive regulation of telomere maintenance (GO:0032206)
mRNA 3'-end processing (GO:0031124)
ncRNA transcription (GO:0098781)
centrosome duplication (GO:0051298)
sister chromatid cohesion (GO:0007062)
mRNA transport (GO:0051028)
nuclear export (GO:0051168)
regulation of mRNA processing (GO:0050684)
sister chromatid segregation (GO:0000819)
cell cycle phase transition (GO:0044770)
regulation of chromosome organization (GO:0033044)
RNA splicing (GO:0008380)
DNA repair (GO:0006281)
mitotic cell cycle process (GO:1903047)
mitotic cell cycle (GO:0000278)
0 5 10 15 20 25
-log10 (P value)
307 genes up in transcriptome
cellular response to transforming growth factor beta stimulus (GO:0071560)
maintenance of cell number (GO:0098727)
negative regulation of p38MAPK cascade (GO:1903753)
regulation of Notch signaling pathway (GO:0008593)
regulation of hemopoiesis (GO:1903706)
0 2 4 6
-log10 (P value)
1290 genes down in translatome
protoporphyrinogen IX metabolic process (GO:0046501)
hexose metabolic process (GO:0019318)
cofactor biosynthetic process (GO:0051188)
0 1 2 3
-log10 (P value)
122 genes down in transcriptome
fatty acid beta-oxidation (GO:0006635)
cellular amino acid metabolic process (GO:0006520)
amide biosynthetic process (GO:0043604)
0 3 6 9
-log10 (P value)

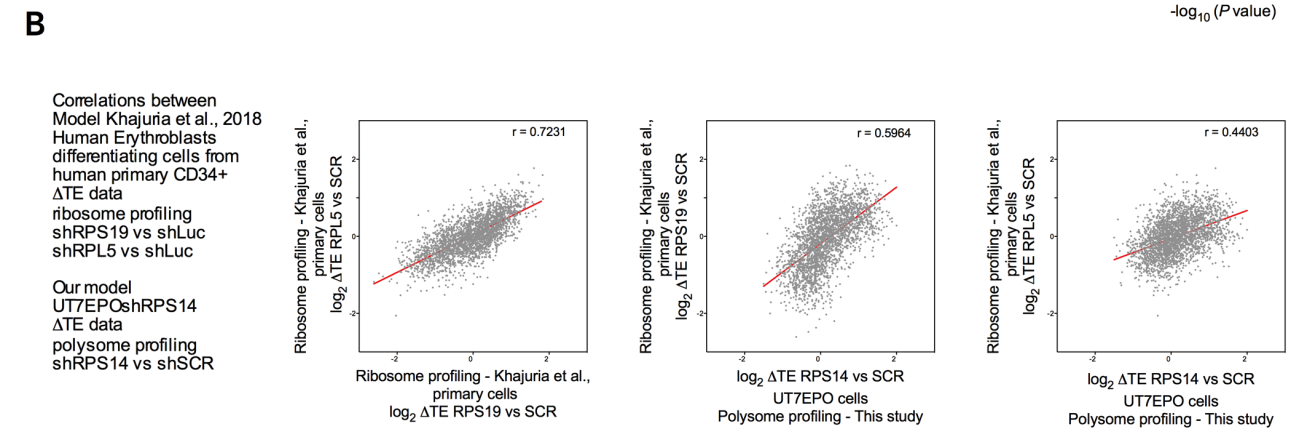

B
Correlations between
Model Khajuria et al., 2018
Human Erythroblasts
differentiating cells from
human primary CD34+
ΔTE data
ribosome profiling
shRPS19 vs shLuc
shRPL5 vs shLuc
Our model
UT7EPOshRPS14
ΔTE data
polysome profiling
shRPS14 vs shSCR
r = 0.7231
Ribosome profiling - Khajuria et al., primary cells log2 ΔTE RPL5 vs SCR
Ribosome profiling - Khajuria et al., primary cells log2 ΔTE RPS19 vs SCR
r = 0.5964
Ribosome profiling - Khajuria et al., primary cells log2 ΔTE RPS19 vs SCR
log2 ΔTE RPS14 vs SCR
UT7EPO cells
Polysome profiling - This study
r = 0.4403
Ribosome profiling - Khajuria et al., primary cells log2 ΔTE RPL5 vs SCR
log2 ΔTE RPS14 vs SCR
UT7EPO cells
Polysome profiling - This study

Figure S2

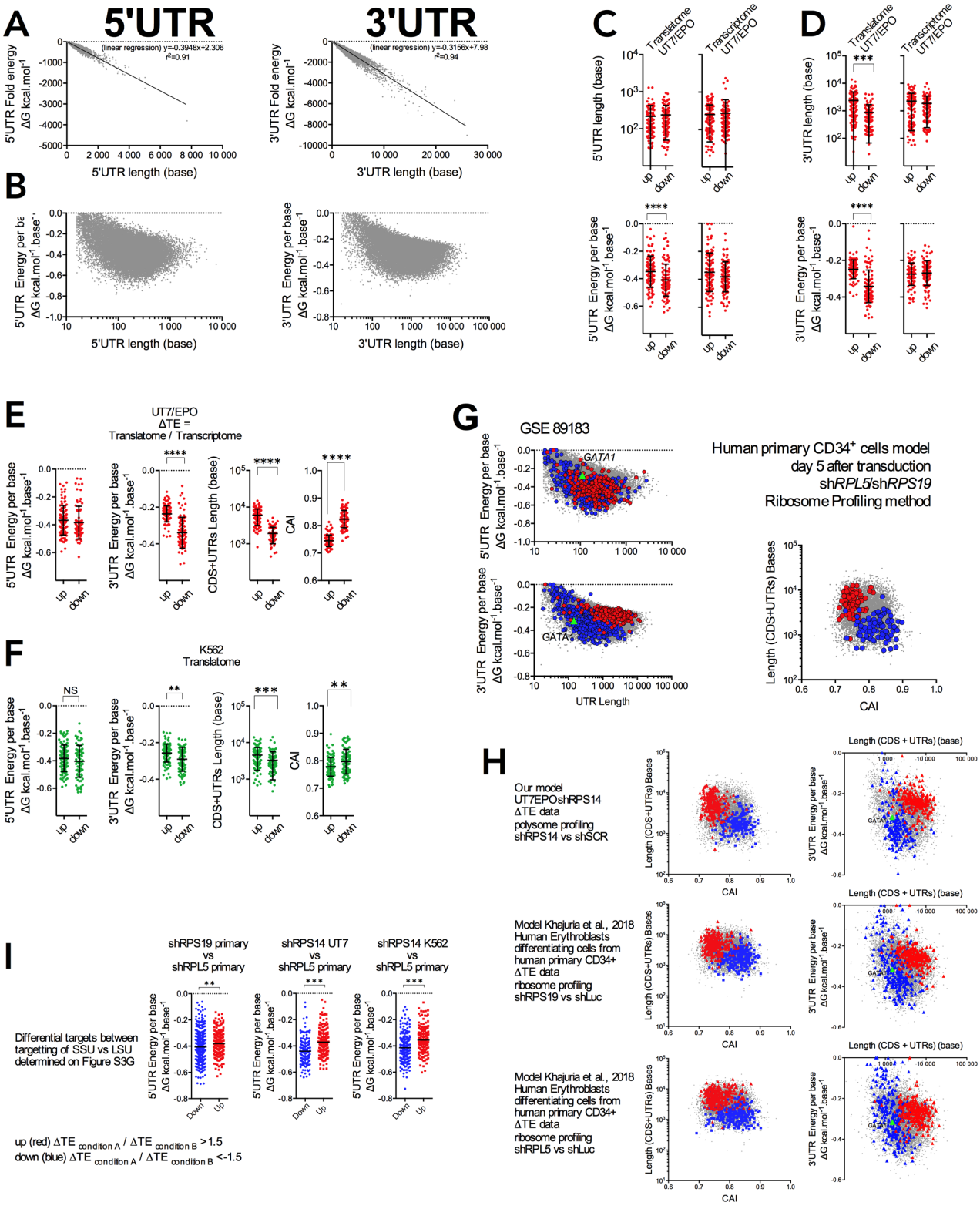

A
5'UTR
(linear regression) y=-0.3948x+2.306
r²=0.91
5'UTR Fold energy ΔG kcal.mol⁻¹
5'UTR length (base)
3'UTR
(linear regression) y=-0.3156x+7.98
r²=0.94
3'UTR Fold energy ΔG kcal.mol⁻¹
3'UTR length (base)
B
5'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
5'UTR length (base)
3'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
3'UTR length (base)
C
Translatome UT7/EPO
Transcriptome UT7/EPO
5'UTR length (base)
5'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
****
up
down
D
Translatome UT7/EPO
Transcriptome UT7/EPO
***
3'UTR length (base)
3'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
****
E
UT7/EPO
ΔTE =
Translatome / Transcriptome
5'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
3'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
CDS+UTRs Length (base)
CAI
F
K562
Translatome
NS
**
***
G
GSE 89183
GATA1
5'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
3'UTR Energy per base ΔG kcal.mol⁻¹.base⁻¹
UTR Length
Human primary CD34+ cells model
day 5 after transduction
shRPL5/shRPS19
Ribosome Profiling method
Length (CDS+UTRs) Bases
CAI
H
Our model
UT7EPOshRPS14
ΔTE data
polysome profiling
shRPS14 vs shSCR
Length (CDS + UTRs) (base)
Model Khajuria et al., 2018
Human Erythroblasts
differentiating cells from
human primary CD34+
ΔTE data
ribosome profiling
shRPS19 vs shLuc
shRPL5 vs shLuc
I
shRPS19 primary vs shRPL5 primary
shRPS14 UT7 vs shRPL5 primary
shRPS14 K562 vs shRPL5 primary
Differential targets between targetting of SSU vs LSU determined on Figure S3G
Down
Up
up (red) ΔTE condition A / ΔTE condition B >1.5
down (blue) ΔTE condition A / ΔTE condition B <-1.5

Figure S3

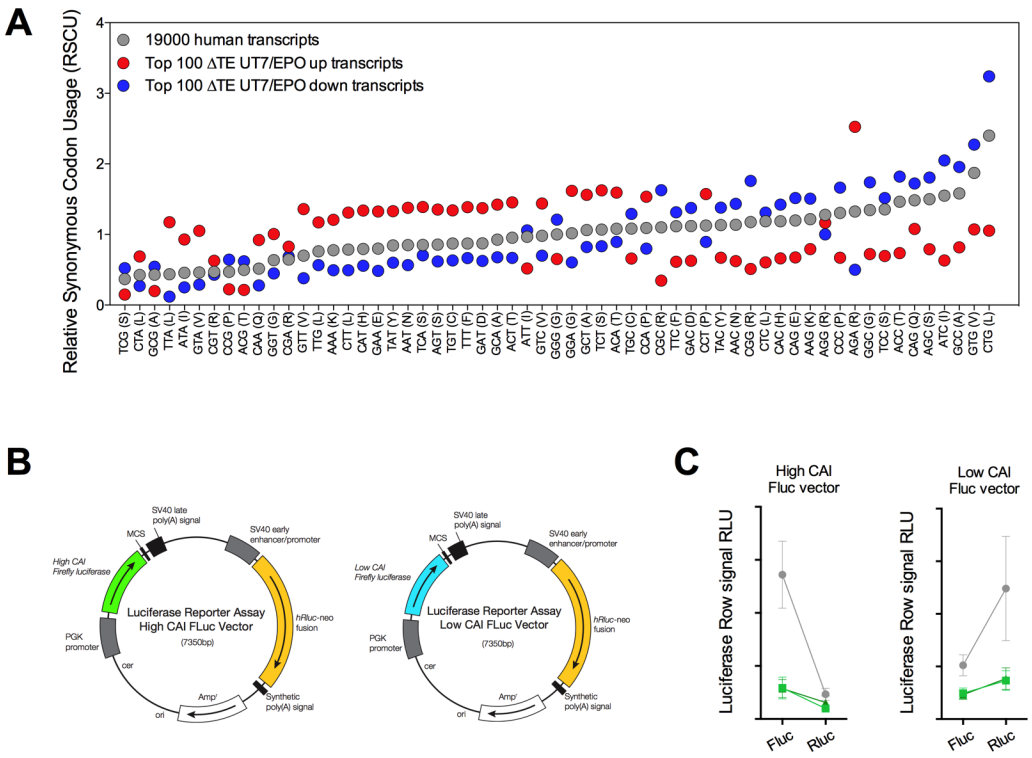

A
Relative Synonymous Codon Usage (RSCU)
19000 human transcripts
Top 100 ΔTE UT7/EPO up transcripts
Top 100 ΔTE UT7/EPO down transcripts
TCG (S) CTA (L) GCG (A) TTA (L) ATA (I) GTA (V) CGT (R) CCG (P) ACG (T) CAA (Q) GGT (G) CGA (R) GTT (V) TTG (L) AAA (K) CTT (L) CAT (H) GAA (E) TAT (Y) AAT (N) TCA (S) AGT (S) TGT (C) TTT (F) GAT (D) GCA (A) ACT (T) ATT (I) GTC (V) GGG (G) GGA (G) GCT (A) TCT (S) ACA (T) TGC (C) CCA (P) CGC (R) TTC (F) GAC (D) CCT (P) TAC (Y) AAC (N) CGG (R) CTC (L) CAC (H) CAG (E) AAG (K) AGG (R) CCC (P) AGA (R) GGC (G) TCC (S) ACC (T) CAG (Q) AGC (S) ATC (I) GCC (A) GTG (V) CTG (L)
B
SV40 late poly(A) signal
MCS
SV40 early enhancer/promoter
High CAI Firefly luciferase
Luciferase Reporter Assay
High CAI FLuc Vector
(7350bp)
hRluc-neo fusion
PGK promoter
cer
Amp
ori
Synthetic poly(A) signal
Low CAI Firefly luciferase
Luciferase Reporter Assay
Low CAI FLuc Vector
(7350bp)
C
High CAI Fluc vector
Low CAI Fluc vector
Luciferase Row signal RLU
Fluc
Rluc

Figure S4

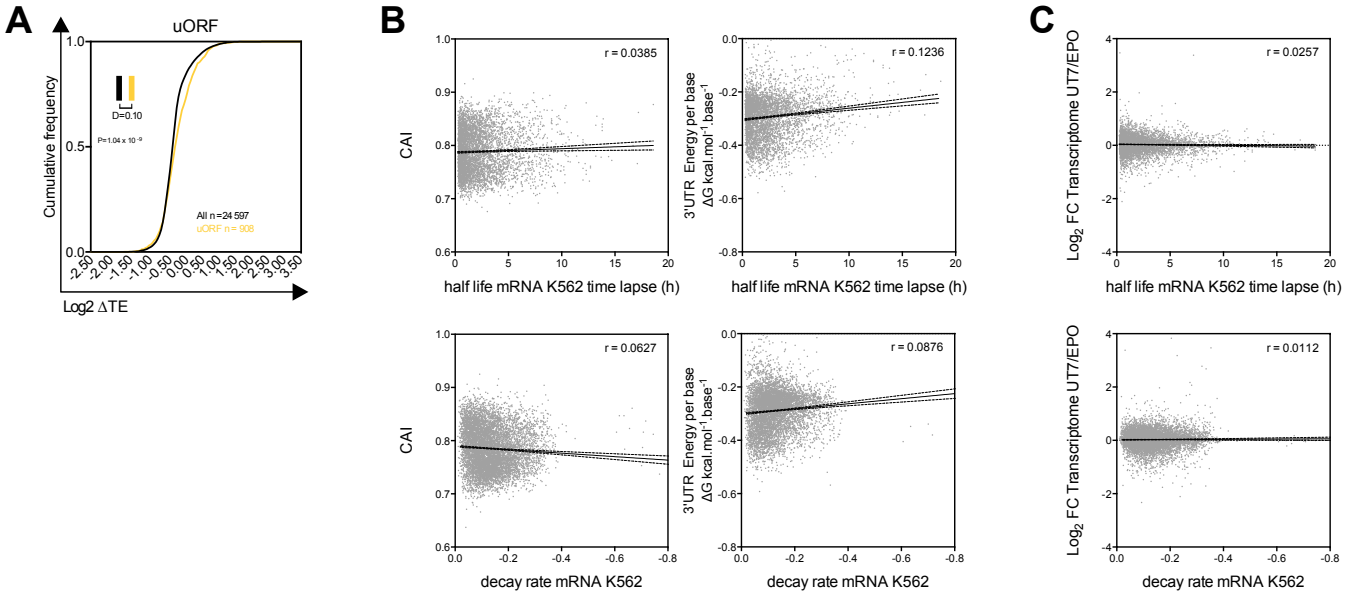

A
uORF
Cumulative frequency
1.0
0.5
0.0
D=0.10
P=1.04 x 10 -9
All n = 24 597
uORF n = 908
-2.50
-2.00
-1.50
-1.00
-0.50
0.00
0.50
1.00
1.50
2.00
2.50
3.00
3.50
Log2 ΔTE
B
CAI
r = 0.0385
half life mRNA K562 time lapse (h)
3'UTR Energy per base ΔG kcal.mol-1.base-1
r = 0.1236
half life mRNA K562 time lapse (h)
CAI
r = 0.0627
decay rate mRNA K562
3'UTR Energy per base ΔG kcal.mol-1.base-1
r = 0.0876
decay rate mRNA K562
C
Log2 FC Transcriptome UT7/EPO
r = 0.0257
half life mRNA K562 time lapse (h)
Log2 FC Transcriptome UT7/EPO
r = 0.0112
decay rate mRNA K562

Figure S5

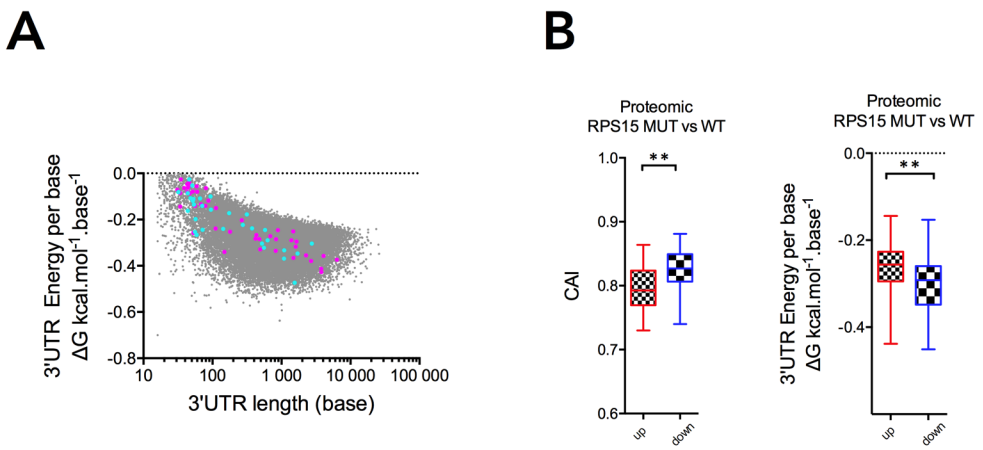

A
B
3'UTR Energy per base
ΔG kcal.mol-1.base-1
0.0
-0.2
-0.4
-0.6
-0.8
10
100
1 000
10 000
100 000
3'UTR length (base)
Proteomic
RPS15 MUT vs WT
CAI
1.0
0.9
0.8
0.7
0.6
**
up
down
Proteomic
RPS15 MUT vs WT
3'UTR Energy per base
ΔG kcal.mol-1.base-1
0.0
-0.2
-0.4
**
up
down

Figure S6

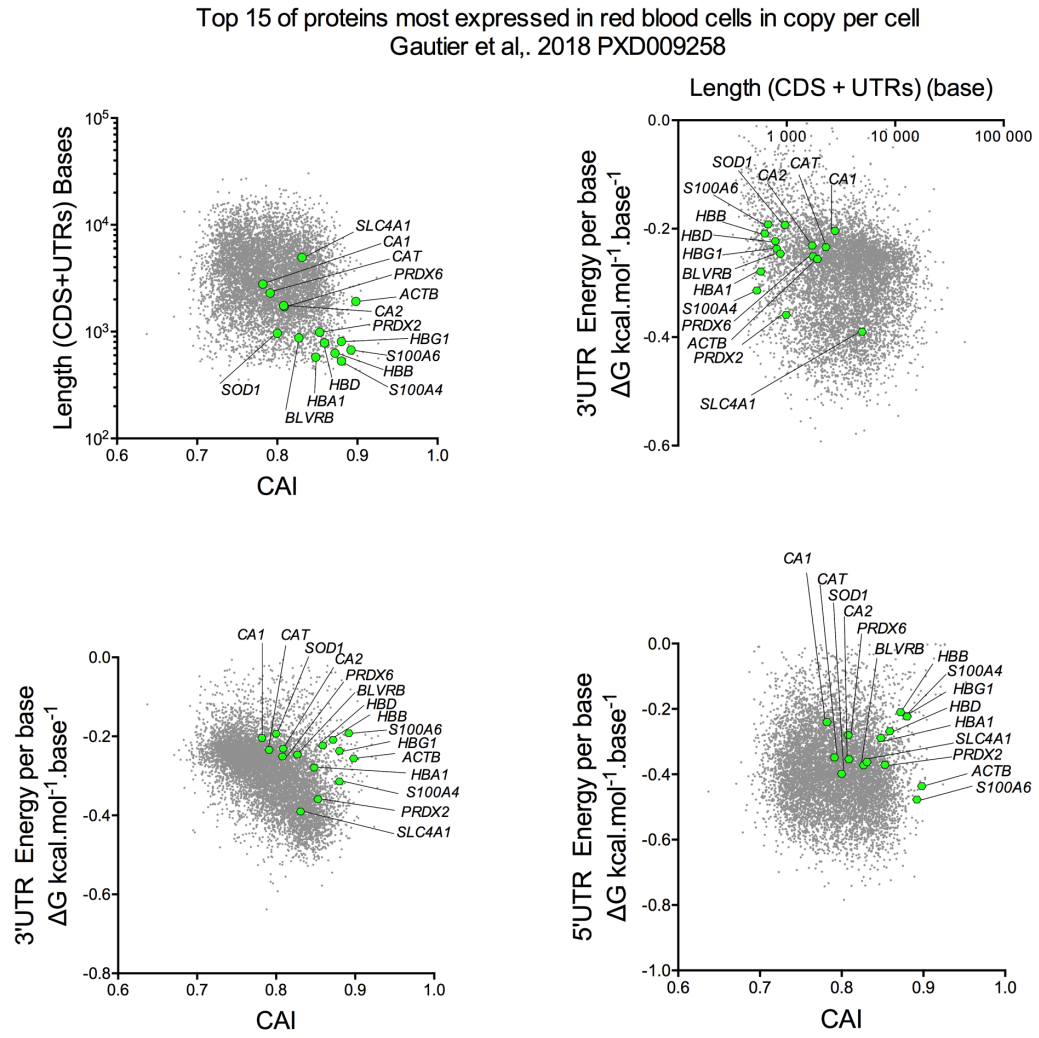

Top 15 of proteins most expressed in red blood cells in copy per cell
Gautier et al,. 2018 PXD009258
Length (CDS+UTRs) Bases
CAI
SLC4A1
CA1
CAT
PRDX6
ACTB
CA2
PRDX2
HBG1
S100A6
HBB
HBD
S100A4
HBA1
BLVRB
SOD1
Length (CDS + UTRs) (base)
3'UTR Energy per base ΔG kcal.mol-1.base-1
1 000
10 000
100 000
5'UTR Energy per base ΔG kcal.mol-1.base-1

Figure S7