Codon usage optimization in pluripotent embryonic stem cells

Susanne Bornelöv[3#], Tommaso Selmi[1#], Sophia Flad[1,2], Sabine Dietmann[3], Michaela Frye[1,2*]

[1] Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, United Kingdom; [2] German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; [3] Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, United Kingdom

[*] Corresponding: M.Frye@dkfz.de

[#] Equal contribution

Abstract

Background: The uneven use of synonymous codons in the transcriptome regulates the efficiency and fidelity of protein translation rates. Yet, the importance of this codon bias in regulating cell state-specific expression programs is currently debated. Here, we ask whether different codon usage controls gene expression programs in self-renewing and differentiating embryonic stem cells.

Results: Using ribosome and transcriptome profiling, we identify distinct codon signatures during human embryonic stem cell differentiation. We find that cell state-specific codon bias is determined by the GC-content of differentially expressed genes. By measuring the codon frequencies at the ribosome active sites interacting with transfer RNAs (tRNA), we further discover that self-renewing cells optimize translation of codons that depend on the inosine tRNA modification in the anti-codon wobble position. Accordingly, inosine levels are highest in human pluripotent embryonic stem cells. This effect is conserved in mice and is independent of the differentiation stimulus.

Conclusions: We show that GC-content influences cell state-specific mRNA levels, and we reveal how translational mechanisms based on tRNA modifications change codon usage in embryonic stem cells.

Background

Understanding normal tissue development and disease susceptibility requires the knowledge of mechanisms determining lineage fate decisions in embryonic stem and progenitor cells. While the transcriptional networks governing the naïve and lineage committed states are now increasingly understood [1-3], it remains largely unknown whether and how translational mechanisms contribute to early cell fate decisions.

Translation of mRNA takes place on ribosomes, and distinct sets of tRNAs link each nucleotide triplet in mRNA to a corresponding amino acid. The genetic code is degenerate, because the 64 codons (triplets) encode for only 20 amino acids. Thus, single amino acids are often encoded by multiple synonymous codons that are not used randomly. The uneven use of synonymous codons in the transcriptome is commonly found in different organisms and called codon bias [4]. The selective use of optimal codons that correspond to abundant tRNAs has been suggested to improve translational efficiency by fine tuning translation elongation rates [5-7]. Yet, the importance of codon bias in regulating gene expression programs remains debated, because other features of the coding sequence, such as GC-content and mRNA secondary structure also strongly influence translation elongation efficiency [8, 9].

The hundreds of tRNA genes interspersed in the human genome encode only 47 out of the 64 possible tRNAs anticodons. Codons lacking a corresponding tRNA are translated via wobble base pairing. While the standard Watson-Crick base-pairing is required at the first and second positions, so-called wobbling at the third position can allow otherwise-disfavored base pairing such as between guanosine and uracil [10, 11]. The affinity by which synonymous codons are recognized via wobble base-pairing can vary, and the translation kinetics of different codon-anticodon pairs is further influenced by modified nucleotides present in tRNAs [12].

Specifically, position 34 in tRNAs, corresponding to the wobble position, carries a wide range of chemical modifications, including 2'O-methylribose, 5-methylcytidine, 5-methoxycarbonylmethyl, 5-carbamoylmethyl, and their derivatives, as well as direct adenine-to-inosine editing (A34I) [13-15]. These modifications can dramatically alter codon translation rates and for instance adapt protein synthesis to external stress stimuli [16-23].

Whether optimal codon usage through tRNA modifications can regulate cell fate has been unexplored. Therefore, using human embryonic stem cells (hESC) as a model system, we analysed codon usage optimization during self-renewal and differentiation. We triggered the exit of pluripotency by exposure to retinoic acid [24, 25]. We found that the differentially expressed genes between the cell states were strongly biased by GC-content. By analysing the translation kinetics at the ribosomes, we found that self-renewing embryonic stem cells optimized codon usage of specific codons that depended on inosine modified tRNAs for translation. The differential codon bias was independent of the nature of the differentiation stimulus. In summary, we reveal a codon-bias driven by tRNA modifications that defines pluripotent embryonic stem cells.

Results

RNA expression and translation highly correlate in self-renewing and differentiating cells

To determine mRNA expression and translation levels in self-renewing and differentiating human embryonic stem cells (hESCs), we performed RNA-sequencing (RNA-seq) and ribosome footprint profiling (Ribo-seq). We induced the exit of pluripotency of human embryonic stem cells (hESCs) by removing the growth factor FGF-2 and exposing the cells to retinoic acid (RA), one of the best characterized stimuli to induce early embryonic differentiation (Fig. 1A) [26, 27].

After sequencing, we determined the quality of the Ribo-seq data. The library composition of self-renewing and differentiating hESCs was comparable and contained a considerable number of reads mapping to ribosomal RNAs (rRNA), transfer RNAs (tRNAs), and adapter dimers, as expected (Additional file 1: Fig. S1A). We found a large number of genes differentially expressed (Additional file 1: Fig. S1B; Table S1). The biological replicates were highly correlated (r = 0.98-1.00) and clustered together (Additional file 1: Fig. S1C). In contrast, the correlation between self-renewing and differentiating samples were much lower (r=0.70-0.77 in Ribo-seq; r=0.85-0.88 in RNA-seq). Both the read length distribution and the strong reading frame periodicity for reads of 27 to 29 nucleotides in length confirmed sufficient coverage of mRNAs to analyse translation at single codon resolution (Additional file 1: Fig. S1D-F) [28]. Metagene analysis further confirmed a strong enrichment of reads over the coding sequences, with only a small number of reads mapping to the 5' UTRs and nearly no reads at the 3' UTRs (Additional file 1: Fig. S1G). Finally, the differentially abundant genes in the Ribo-seq and RNA-seq datasets highly correlated (Additional file 1: Fig. S1H), suggesting that transcriptional mechanisms might be the dominant factor changing mRNA levels upon exit from pluripotency [29].

Induction of stem cell differentiation using RA in the absence of FGF-2 induced transcripts of all primary germ layers (Table S2). Both RNA-seq and Ribo-seq confirmed the existence of heterogeneous cell populations expressing markers of ectoderm, endoderm, and mesoderm, while pluripotency markers were consistently repressed (Fig. 1B-D) [30-33]. The up-regulated genes further included the *Hox* family, which is known to be regulated through RA-signaling in early embryonic development [34]. To further confirm that we efficiently differentiated the hESCs, we also grew hESCs in suspension to induce their differentiation into embryoid bodies (EBs) for 5 and 7 days [35]. The change of mRNA levels of pluripotency and lineage markers were comparable to RA-induced differentiation (Fig. 1E-G). Thus, RA-treated hESCs exited the pluripotent state and underwent cell differentiation.

Codon composition of cell state-specific mRNAs is biased towards GC-content

We next asked whether self-renewing and differentiating cells optimized their translational programs by using cell state-specific codons. First, we selected all well-annotated coding sequences from the consensus coding sequence project [36]. Then, we calculated the relative codon frequency of each gene, thereby each gene was represented as vector of 64 codon frequencies. Using our data, we defined two groups of genes: (*i*) significantly up-regulated genes in self-renewing hESCs and (*ii*) significantly up-regulated genes in differentiating hESCs, and then calculated the overall codon usage compared to all genes (Fig. 2).

We reasoned that if codon bias contributed to altering mRNA translation between self-renewing and differentiating cells, then it would be most likely to occur at the wobble positions due to the relaxed base pairing and the degeneracy of the genetic code. Therefore, we subdivided the codons according to the base at the wobble positions (*nn*A, *nn*T, *nn*C, and *nn*G)

(Fig. 2A). The codon enrichment ratios were highly similar for both the up-regulated genes in self-renewing and differentiating hESCs (Fig. 2A). Genes that changed most in expression upon retinoic acid treatment were commonly enriched in G and C at the wobble position (Fig. 2A; dotted circle) and depleted in A and T. While *nn*C codons clustered closely together, *nn*G codon were more diffuse; and this was correlated to the codon degeneracy. Only four-fold degenerate *nn*G codons clustered closely with *nn*C codons (Fig. 2B). Next, we asked whether the codon bias correlated with a different amino acid usage in the two cell states. However, neither up- nor down-regulated genes showed an enrichment for a specific amino acid usage (Fig. 2C).

We considered the possibility that the enrichment of *nn*C and *nn*G codons was driven by the translational machinery. However, we observed a similar distribution of codon usage in the up-regulated genes of both cell states obtained by RNA-seq (Fig. 2D-F). Thus, our data indicated that the difference in codon usage in self-renewing and differentiating hESCs was already present at mRNA level. Possible underlying mechanism causing the distinct codon usage in self-renewing and differentiating hESCs are for example transcription or RNA decay.

The use of optimal codons can enhance mRNA stability and translation, and this effect is evolutionary conserved [37-40]. To address whether optimal codons enhanced mRNA stability in the two cell states, we calculated the relative enrichment of stabilizing and destabilizing codons (Additional File 1: Fig. S2A) using a codon classification made in zebrafish [39]. The up-regulated mRNAs in both conditions showed a slight depletion of strongly destabilizing codons (Additional File 1: Fig. S2B). When we measured the difference between the self-renewing and differentiating genes, we found that up-regulated mRNAs in self-renewing hESCs contained significantly fewer destabilizing codons and slightly more highly stabilizing

codons (Additional File 1: Fig. S2B; right hand panels). The association of GC-content at the third nucleotide of a codon to high levels of mRNAs was however more pronounced than the association to stabilizing and destabilizing codons (Additional File 1: Fig. S2C, D). Together these data demonstrated that abundant mRNAs in both the self-renewing and differentiated state were biased towards a higher GC-content, but that in addition, codon optimality contributed to the codon usage differences between the two cell states.

GC-content defines groups of genes with common function

Underlying GC-content can substantially contribute to codon usage differences among different cell states [9, 41, 42]. We therefore analysed codon usage in genes belonging to various functional categories, including those deployed during self-renewal and differentiation of hESCs. We selected all well-annotated coding sequences and grouped them according to their Gene Ontology (GO) terms. We calculated the average of codon frequencies per GO term and used principal component analysis (PCA) to reveal the underlying sources of variance. The GO terms primarily separated by GC-content (Fig. 3A); indeed, the first principal component closely corresponded to the GC-content of the third codon base (Fig. 3B). The second principal component reflected the GC-content at the first and second bases (Fig. 3C). Accordingly, genes belonging to similar GO terms such as cell differentiation and neuron fate commitment had similar and high GC-content (blue) (Fig. 3C; right panels).

It was unexpected that all three codon positions separated the GO terms according to GC-content. While the third and variable base position of codons can be readily varied to optimize translation rates without impacting the encoded amino acids [5], changes in the first and second position almost always result in a different amino acid. Nevertheless, we found that genes

belonging to different functional categories were characterized by different codon usage and that this correlated with GC-content.

Using a PCA loadings plot, we identified the codons driving the global separation by GC-content (Fig. 3D). Codons with cytosine at the third position contributed to the left and codons with adenosine or thymidine at the third position contributed to the right of the plot. As described for our own differentiation data (Fig. 2B, E), only four-fold degenerate codons with guanosine at the third position contributed to the left of the PCA plots (Fig. 3D). The high level of similarity between the global PCA analysis of all genes and our differentiation experiment suggested that genes with similar functions share similar inherent genomic features. In conclusion, genomic GC-content influences codon choice in the differentiation gene expression program.

Self-renewing ESCs share a common codon signature at the ribosome A- and P-sites

While GC-content was one factor for codon usage optimization, protein translation efficiency is likely to be regulated by additional factors. The translation kinetics of codon-anticodon pairs in the ribosome are complex and determined by the availability and binding affinity of matching and mismatching tRNAs [5, 43, 44]. In part, translation rate is directed by the competition between near-cognate and cognate aminoacyl tRNAs entering the ribosome. Therefore, we asked whether self-renewing and differentiating hESCs showed differential codon stalling at the active sites of the ribosome.

During translation, only three codons actively pair with tRNAs, known as the ribosome E-, P- and A- sites (Fig. 4A). Cognate aminoacyl-tRNAs enter the ribosome at the A-site. After transfer of the polypeptide chain to the tRNA, the tRNA moves from the P- to the E-site and is

then released from the ribosome. Although the ribosome only actively engages with three codons within an RNA, it covers more than 27 nucleotides (Fig. 4A). Due to the three-nucleotide periodicity reflecting the triplet nature of the genetic code (Additional file 1: Fig. S1E-F) [45], the actively bound codons at the E-, P- and A-sites can be calculated by counting nucleotide triplets from the 3' end of the read fragments. We numbered the ribosome-protected codons from -5 to +3, with 0 corresponding to the A-site (Fig. 4A). We first calculated the codon frequencies per footprint position. To do so, we compared the relative frequency of a codon at the ribosome P (or A) site to the frequency of this codon in all Ribo-seq reads from the sample.

As expected, some codons occurred more frequently than others (Additional File 1: Fig. S3A-C). Since the codon frequencies at the positions flanking the tRNA-bound sites (-5 to -3, and +1 to +3) are expected to follow the genomic codon distribution, they served as a negative control. We then normalized the raw frequencies to the frequency across all nine positions (Fig. 4B). This will correct for differences in mRNA abundance and neither transcriptional nor differences in mRNA decay should affect the following analysis. The strong enrichment of stop codons at the A-site and their depletion at the P- and E-sites validated this approach (Fig. 4C-E). Although Proline incorporates more slowly in translation [46], its codons were not strongly enriched at the A-, P-, or E site (Fig. 4C-E). When we compared the enrichment of codon frequencies along all sites in the ribosome, we found the highest variation at the A-site (Fig. 4F, G), indicating that the stalling at the A-site was rate limiting for translation elongation [47, 48].

Next, we tested whether the normalized codon frequencies differed in self-renewing and differentiating hESCs (Fig. 4H). We observed differences at both the A- and the P-sites, yet

most significant differences occured at the P-site (Fig. 4H; Additional File 1: Fig. S3D). To test whether the codon usage differences in human ESCs was an evolutionarily conserved feature of translational control, we performed similar analyses using published Ribo-seq and RNA-seq data obtained from self-renewing and differentiating mouse embryonic stem cells (mESCs) [49]. The exit of pluripotency and differentiation of the mESCs was induced using two well-established methods: the withdrawal of leukemia inhibitory factor (LIF) from the culture medium and by growing the mESCs as embryoid bodies (EBs) [50-52]. As described for hESCs, we excluded rRNA and tRNA reads and selected read lengths with high reading frame periodicity (Additional File 1: Fig. S4A-E). Again, most differences occurred at the A- and P-sites (Fig. 4I). Although the mouse data were noisier, the significantly depleted codons at the P-site in differentiated hESCs were often also less common in differentiated mESCs (Fig. 4H, I; Additional File 1: Fig. S4F; blue dots). Thus, we found a distinct set of codons with a higher likelihood of being stalled at the ribosome P-site in mouse and human self-renewing embryonic stem cells.

A prominent self-renewing codon signature is hetADAT-sensitive

A key factor modulating codon translation is tRNA selection [19]. To enhance codon binding, the tRNA wobble nucleoside (position 34) frequently carries chemical modifications [53]. One such prominent and essential tRNA modification is wobble inosine [54, 55]. Inosine is formed by a deamination reaction of adenosine and catalyzed by the heterodimeric enzyme adenosine deaminase acting on tRNA (hetADAT) composed of two subunits ADAT2 and ADAT3 (Fig. 5A) [56]. While tRNAs carrying adenosine at position 34 (A34) pair with uracil, inosine (I34) is capable of pairing with adenosine, uracil as well as cytosine and thereby, enhances translation efficiency (Fig. 5A) [11, 57]. In eukaryotes, the modification is present in eight different tRNAs that recognize the eight *nn*T codons. Inosine in tRNAs is required to translate

*nn*C codons that have none or few tRNA gene copies (Fig. 5B) [58]. In principle, the same tRNAs can also bind eight *nn*A codons, yet these codons are also translated by their own complementary tRNAs.

We noticed that codons showing significant differences in self-renewing hESCs were often translated by inosine-modified tRNAs (Fig. 5C). All *nn*T and *nn*C codons that depended on inosine tRNA for translation occurred less often at the P-site in differentiated hESCs (Fig. 5C). Moreover, these differential codon frequencies highly correlated between human and mouse (Fig. 5D). Thus, mouse and human differentiating ESCs showed a depletion of hetADAT-dependent codons at the P-site. As described above, due to the high variability of codon frequencies at the A-site, we were unable to identify significant changes (Fig. 4H). However, when we now plotted the A34I-dependent codons, we found the vast majority of them enriched at the A-site in both mouse and human differentiating ESCs (Fig. 5E, F; Table S3a,b). Finally, we confirmed that this distribution was specific to the A- and P-site and did not occur at any other site in the ribosome (Fig. 5G). In summary, we found that in self-renewing ESCs hetADAT-dependent codons occurred less often at ribosome A-site but more often at the P-site when compared to differentiated ESCs.

The time it takes the ribosome to ensure that a tRNA in its A-site is indeed a cognate tRNA for the codon, is considered rate limiting for translation elongation [59, 60]. In addition, whenever a ribosome is bound to an mRNA waiting for a cognate tRNA at its A-site, a tRNA is already bound to its P-site, attached to the growing polypeptide chain. Thus, the lower frequency of hetADAT-dependent codons at the A-site indicated their faster translocation to the P-site in self-renewing ESCs.

hetADAT regulates codon usage in self-renewing embryonic stem cells

The lower frequencies of hetADAT-dependent codons at the A-site of ribosomes in self-renewing ESCs might indicate an enhanced activity of the catalytic complex. Indeed, the mRNA level of ADAT2 was significantly higher in self-renewing hESCs when compared to RA-differentiated cells or EBs (Fig. 6A). To quantify the levels of inosine in hetADAT-dependent tRNAs, we performed tRNA sequencing in self-renewing and differentiated hESCs (Table S4). We used the fact that inosine will form the strongest bond to cytosine and is therefore reported as a guanine during sequencing. The level of A34I modifications can thus be quantified by measuring the A-to-G substitutions at the anticodon wobble position (Table S5). The A34I tRNA modification levels were significantly higher in self-renewing (83-86%) than differentiated (60-79%) hESCs (p=0.031; Student's t-test) (Fig. 6B). Accordingly, the A34I modification occurred less often in the majority of hetADAT-dependent tRNA isotypes (Fig. 6C). Thus, self-renewing hESCs have higher levels of A34I tRNA modifications than differentiating cells.

Next, we confirmed that also the protein levels of ADAT2 and ADAT3 decreased with differentiation of hESCs (Fig. 6D-F; Additional File 1: Fig. S5A-E). ADAT2/3 protein expression was significantly reduced after only one day of treatment with retinoic acid (Fig. 6E; Additional File 1: Fig. S5D). Down-regulation of hetADAT was further confirmed when the hESCs were grown in suspension to induce their differentiation into EBs (Fig. 6F; Additional File 1: Fig. S5E). Similarly, translation of mouse ADAT2 and ADAT3 was down-regulated upon differentiation, yet the differences were only significant when compared to differentiated EBs (Fig. 6G).

Finally, we asked whether the optimized codon usage in self-renewing ESCs resulted in differential translation of hetADAT-dependent codons. We calculated the total frequency of the eight *nn*C codons that rely on the inosine modification for translation in all genes (Fig. 5B). Then, we selected all coding sequences containing more than 28.4% (top 1%) *nn*C codons and asked whether they showed translational differences in self-renewing versus differentiated hESCs that were independent of mRNA levels. Comparing the two cell states, we found that translation of 13.5% *nn*C-high coding sequences was enhanced and 7.1% repressed (>2-fold) (Fig. 6H; left panel). This difference in translation of *nn*C-enriched genes was significant different when compared to all coding sequences (Fig. 6H; right panel). Thus, the differential translation of *nn*C-high coding sequences in self-renewing hESCs cannot be explained by differences in RNA abundance alone.

Together our data suggest that self-renewing ESCs translate hetADAT-dependent codons with higher efficiency when compared to differentiating ESCs. This difference in hetADAT-dependent codon usage occurred independently of the nature of the differentiation stimulus and was conserved in mice.

Discussion

Protein synthesis is a fundamental process in all cells, yet recent studies revealed distinct regulatory functions of the mRNA translation machinery in stem cells. Low global protein translation rates are commonly found in adult undifferentiated stem and progenitor cells and are required to maintain a fully functioning stem cells state [61-63]. Mouse embryonic stem cell differentiation also correlates with an overall increase in protein synthesis and enhanced mRNA loading into polyribosomes [64]. One explanation for the differences in protein

synthesis might be the expression of distinct tRNA pools that optimize codon usage in different cell states [6]. However, the use of tRNA concentrations to understand codon usage is hampered by the fact that in addition of being crucial adaptors during translation, tRNAs are important regulators of a wide range of biological processes [65]. An alternative method to analyse codon usage is ribosome profiling, which has been applied to various organisms and cell contexts, for example amino acid starvation, oxidative stress, the perturbation of signalling pathways, and embryonic stem cell differentiation [49, 66-69].

Here, we use ribosome profiling to analyse the importance of optimized codon usage in regulating gene expression in different cell states. Our work identifies a codon bias in self-renewing and differentiating stem cells, and we propose that this is driven by both mRNA levels and translational mechanisms. Genes belonging to the same functional category use similar codons characterized by differences in CG content. Since we measured similar GC-content enrichment in up- and down-regulated genes, our observed differences are unlikely to be directly linked to mRNA degradation. Our data is consistent with previous studies in Neurospora, where codon usage strongly correlated with protein and mRNA levels genome-wide but was largely independent of mRNA translation and stability [70].

Thus, GC-content is one important factor that influences codon choice during gene expression. The concept that codon optimization is due to the GC-content of the underlying genes has been demonstrated before for both specific genes and genome-wide [9, 41, 42]. However, we further propose that the regulation of the protein synthesis machinery through chemically modified tRNAs plays important additional roles in regulating gene expression programs. We find that codons whose translation depend on hetADAT-modified tRNAs were under-represented at the ribosome A-sites in self-renewing mouse and human ESCs. We provide further evidence that

a faster translation of hetADAT codons in self-renewing cells might be facilitated by higher levels of inosines at the tRNA wobble position. While a role for hetADAT in embryonic stem cell fate has not been described before, it is well known that tRNA wobble modifications ensure efficient protein synthesis, maintain protein homeostasis, and promote cell adaptation and survival [71].

The existence of a nuclear (GC-content) and a cytoplasmic (translation) mechanism that both determine codon usage during stem cell differentiation raises the question of how they are connected. It is plausible that a cytoplasmic tRNA modification-driven codon bias may help to overcome a GC-content-driven codon usage for cell type-specific mRNAs, thereby enhancing translation of certain groups of mRNAs. However, the functional analyses on how precisely hetADAT regulates codon usage and how this influences gene expression in pluripotent embryonic stem cells is hampered by the fact that ADAT2/3 is essential for cell survival [56]. Knock-down of hetADAT causes cell death in human cells. Over-expression also affects cell survival because ADAT2/3 is highly mutagenic enzyme [72]. Therefore, ADAT2/3-mediated C to U deamination at position 32 requires the prior formation of 3-methylcytosine in *Trypanosoma brucei* [73]. Thus, increasing the hetADAT levels might not be sufficient to increase inosines specifically at the wobble positions.

Together, we provide evidence for an hetADAT-dependent codon bias in self-renewing embryonic stem cells that might suppress differentiation and lineage commitment.

Conclusion

In this study, we used RNA-seq and Ribo-seq to decipher transcriptional and translational mechanisms regulating codon bias in self-renewing and differentiating human embryonic stem cells. We revealed that codon usage during stem cell differentiation is regulated at the mRNA levels and during translation. We confirm that codon usage of differentially expressed genes is primarily characterized by genomic GC-content. Furthermore, we reveal a novel mechanism based on tRNA modifications that regulate codon usage in pluripotent embryonic stem cells. Translation of codons that depend on the hetADAT-mediated inosine formation in the anti-codon loop of tRNAs are under-represented at ribosome A-site. The reduced stalling of these codons at the A-site implies their enhanced translation in self-renewing embryonic stem cells when compared to differentiated cells. Thus, we reveal that tRNA modifications contribute to optimized codon usage in self-renewing embryonic stem cells.

Material and Methods

Human embryonic stem cell culture and differentiation

The human embryonic stem cell (hESC) line Hues9 (H9) was obtained from the Wicell Research Institute (Madison, WI). hESC were maintained in Essential 8 media (Thermo Fisher Scientific) on hESC-Qualified Matrigel (Corning) coated plates at 37°C, 5 % CO_2. Cultures were dissociated in clumps using 0.5mM EDTA in PBS every 4 days and media was renewed daily. Differentiating hESC were cultured in MEF-conditioned KSR media plus 1 uM retinoic acid in DMSO for five days, while the control hESC population was maintained either in MEF-conditioned KSR media completed with 4 ng/mL FGF2 or Essential 8 media (Thermo Fisher Scientific). During the course of the experiment media wash replaced daily. KSR media consists of 85% KO-DMEM (Thermo Fisher Scientific), 15% KO-serum replacement (Life Technlogies), 1mM Glutamax (Thermo Fisher Scientific), 0.1mM 2-mercaptoethanol (Thermo Fisher Scientific), 0.1 mM non-essential aminoacids (Thermo Fisher Scientific). In the embryoid bodies experiments, 70% confluent HESC were dissociated in clumps using 0.5 mM EDTA in PBS and seeded in ultra-low attachment well plates (Corning) maintaining a 1:1 dilution factor. hESC were cultured in Essential 6 media (Thermo Fisher Scientific) plus 10 uM rock-inhibitor (Y-27632) (Stem Cell Technologies Canada) for the first 24 hours and for further 4 to 6 days in Essential 6 media (Thermo Fisher Scientific), when samples were collected for qPCR or Western Blot experiments.

RNA, qPCR and Western Blot

Total RNA was extracted using TRIZOL (Thermo Fisher Scientific) according to manufacturer instructions. Reverse transcription was performed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) and random primers (Promega). Quantitative PCR were run using TaqMan probes (Thermo Fisher Scientific) for eukaryotic 18S rRNA (X03205.1), ADAT2

(Hs00699339_m1) CDX2 (Hs01078080_m1), DLX3 (Hs00270938_m1), DLX5 (Hs01573641_mH), DNMT3B (Hs00171876_m1), FOXD3 (Hs00255287_s1), GATA6 (Hs00232018_m1), HOXA1 (Hs00939046_m1), NANOG (Hs02387400_g1), POU5F1 (Hs03005111_g1), PRDM14 (Hs01119056_m1), TDGF1 (Hs02339497_g1).

Protein extracts used for western blotting were prepared in RIPA buffer (50 mM sodium chloride, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0. The antibodies used in the experiments were anti HSP90 (sc-13119) 1:5000, anti OCT3/4 (sc-5279) 1:1000 both from Santa Cruz Biotechnologies and anti ADAT2 (ab135429) 1:1000 and anti ADAT3 (ab125514) 1:1000, both from Abcam. The Western Blot band intensity was measured using ImageJ.

Immunofluorescence

For immunofluorescence experiments, hESC were plated on Matrigel coated coverslips and cultured as described. At the desired time point, cells were washed in PBS and fixed for 10' in ice cold 4% PFA. Cell permeabilisation was performed in 0.01% Triton x-100 in PBS for 5' at room temperature. Coverslips were blocked for 1 hour at room temperature using 10% donkey serum in PBS. Anti ADAT2 (ab135429) 1:100 and anti ADAT3 (ab125514) 1:100 primary antibodies were incubated overnight in the same blocking solution. Alexa Fluor secondary antibodies (Thermo Fisher Scientific) were used at a dilution of 1:500 in blocking solution. Nuclei were counterstained using DAPI. Images were acquired on a Leica SP8 Confocal microscope and process

Preparation of RNA-seq, Ribo-seq, and tRNA seq libraries

Total RNA from self-renewing and differentiated hESCs (4 biological replicates per condition) was extracted using Trizol according to the manufacturer's instruction. Total RNA was treated with the RiboZero Magnetic Kit (Epicentre, MRZH11124) to remove ribosomal RNA. Libraries for sequencing were prepared using the NEB ultra directional library prep kit (NEB).

Ribosome protected RNA was isolated as described before [61, 74]. Briefly, at the indicated time points, H9 cells (4 replicates) were washed with PBS and lysed in 20 mM Tris-Cl (pH 7.4), 150 mM NaCl, 5 mM MgCl2, 1 mM dithiothreitol (DTT) (Sigma), 1% Triton X-100 (Sigma), 25 U ml−1 of Turbo DNase I (Thermo Fisher Scientific) and 100 μg ml−1 of cycloheximide (Sigma). After passing the lysates ten times through a 26-G needle, they were spun down at 13,000 r.p.m. for 10 min. Digestion with RNaseI (100 U μl−1, Thermo Fisher Scientific) for 45 min at room temperature was used to produce ribosome mRNA footprints. The RNaseI digestion was inhibited with SuperaseIN (Thermo Fisher Scientific) and lysates were fractionated on a 1 M sucrose cushion by ultracentrifugation at 45,000 r.p.m. in a 70Ti rotor for 9 h at 4 °C. The ribosome mRNA footprints were further purified using Qiazol reagent. Footprints with a length of 26–34 nucleotides were size selected on 15% TBE-urea gel (Thermo Fisher Scientific) and 3′-dephosphorylated with T4 polynucleotide kinase (10 U, NEB). All samples were multiplexed and sequenced on the HiSeq2500 platform (Illumina). The NGS data are up-loaded onto GEO (GSE123611): Bornelöv S, Selmi T, Flad S, Dietmann S, Frye M. Codon usage optimization in pluripotent embryonic stem cells. Datasets. Gene Expression Omnibus.https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123611(2019).

For tRNA sequencing, differentiating and self-renewing H9 (4 replicates per condition) were collected in TRIZOL (Thermo Fisher Scientific) at day 5 of differentiation. 50 ug of total RNA samples were DNaseI digested (Ambion) at 37°C for 30 min. The treated samples were then

subjected to phenol-chloroform extraction and isopropanol precipitation and resuspended in 1mM EDTA 0.1M Tris-Hcl (pH9.0). 1 ug of each replicate was used to generate cDNAs and perform qPCR to confirm differentiation. The remaining sample was heated at 37°C for 30 min in order to de-aminoacylate tRNAs. Subsequently, samples were heated for 5' at 80°C and then loaded on a Novex ^{TM} TBE-Urea polyacrylamide gel (Thermo Fisher Scientific). A gel fragment spanning from 50 to 100 nucleotides, according to the Abnova RNA marker low easy, was excised from the gel and resuspended in 400 uL of a solution of 300 mM NaAc pH 5.5, 1.0 mM EDTA, 0.25% SDS. RNA was extracted by a cycle of 30' freezing at -80°C followed by overnight gentle mixing at room temperature in the same buffer and subsequently ethanol precipitated. RNA ends of any tRNA fragments were phosphorylated by treatment with 20 U of T4 Polynucleotide Kinase (NEB) supplemented with ATP (NEB). Following heat inactivation and phenol chloroform extraction and isopropanol precipitation samples were subjected to library preparation using the Illumina Truseq small RNA kit (Illumina). Samples were multiplexed and sequenced on the HiSeq4000 platform (Illumina).

Processing of RNA-seq, Ribo-seq data and tRNA-seq data

Human RNA-seq and Ribo-seq data were obtained as described above. Mouse RNA-seq and Ribo-seq data with accession numbers SRR315620-SRR315627 and SRR31591-SRR31600 were downloaded from the NCBI short read archive [49]. *Trim_galore!* was used to trim small RNA adapters from the human data (auto-detected under default settings) or polyadenylation sequences from the mouse data (using the parameters "--stringency 1" and "--adapter" with ten 'A' followed by multiple 'N'). Trimmed reads shorter than 20 nt were discarded. Alignment was done using Tophat2 (v2.1.0) [75]. The RNA-seq reads were aligned directly to the reference genome (hg38 or mm10). Ribo-seq reads were firstly aligned to a set of known rRNA and tRNAs (selected from the UCSC RepeatMasker tracks), followed by alignment of all

unmapped reads to the reference genome. An index with known transcripts (Gencode v23 or Gencode vM9) were provided for the genome alignment and novel splice junctions were permitted [76, 77]. Multi-mapping read alignments were not used.

For tRNA-seq data analyses *Trim galore!* with "--paired --stringency 6 -a TGGAATTCTCGG -a2 GATCGTCGGACT" was used to trim small RNA adapters and to remove reads shorter than 20 nucleotides. Alignment was done using bowtie [78] with "-n 2 -y -k 1 --nomaqround --allow-contain" to allow for up to two mismatches in the alignment and to report one alignment per read. The reads were aligned to the 430 tRNAs from the high-confidence set in GtRNAdb [79]. Introns were removed and the 'CCA' tail was added before creating the bowtie index.

Modified bases were identified using samtools mpileup with "-BQ 0 -t AD -vuf" to report base coverage per positions. Inosine modifications were identified as positions with an A-to-G substitution at position 33-35. The presence of the expected anticodon in the reference sequence was used to verify that only substitutions at the correct nucleotide were counted. We identified 21 Ala, 7 Arg, 8 Ile, 9 Leu, 9 Pro, 8 Ser, 7 Thr, and 9 Val tRNAs with at least one mapped read supporting the substitution. These are the same eight tRNA isotypes that were expected to be modified. Next, we estimated to overall modification level per sample as the percentage of reads supporting the modification (A-to-G substitution) across all 78 modified tRNAs.

Meta-gene profile

DeepTools were used to quantify the Ribo-seq coverage across the whole reference genome [80]. Each strand was quantified separately and a blacklist file containing all rRNA, tRNA, snoRNA, snRNA and miRNA was provided. The bin size was set to 1 and an offset of 12 was

used to only consider a single nucleotide corresponding to the 'P' site predicted from each read. A meta-gene profile for all protein coding genes was then computed using *computeMatrix* with the parameters "scale-regions -b 500 --unscaled5prime 200 --regionBodyLength 1000 --unscaled3prime 200 -a 500" to define relative coordinates and "--metagene --exonID CDS --transcriptID gene --transcript_id_designator gene_id" to define the coding sequence of each gene. Finally, a trimmed mean was used to exclude the most extreme value at each position.

Differential mRNA levels and translation

Gene models were downloaded from Gencode (v23 for human and vM9 for mouse). *FeatureCounts* was used to quantify the number of reads per gene, represented by either all annotated exons (RNA-seq) or all annotated coding sequences (Ribo-seq) [81]. Only reads aligning to the sense strand of the gene and with mapping quality of at least 20 was counted. For the paired-end RNA-seq the additional flags "-B -C" was specified to exclude chimeric reads and/or reads mapping with only one end. Differential mRNA expression and translation analyses were done with *DESeq2* [82]. The heatmaps were created using the R package pheatmap (https://CRAN.R-project.org/package=pheatmap) and the amino acids logo was created using the DiffLogo package [83].

Determining read periodicity and the position of the ribosome P-site

Ribosome profiling data display 3-nt periodicity, but the reading frame of the 5' read ends differ between different experiments and even between read lengths within the same experiment. The human data showed the strongest periodicity for reads of length 27-29 (Additional file 1: Fig. S1E-F) and the 5' end of those reads were highly enriched for the first codon position and the most frequent starting position was located 12 positions upstream of

the TIS. We therefore decided to base our codon analysis in human on these three reads lengths and to extract read position 12-14 as the P-site codon.

The mouse data showed different dominating reading frames for different read lengths (Additional File 1: Fig. S4C). To predict the strongest reading frame for each read length we again calculated the number of reads starting from each position around the TIS (see Additional File 1: Fig. S4D for an example). Next, we counted the most frequent reading frame both per codon (from codon position -7 to +32 relative to the TIS) and across the whole region. We excluded read lengths where the reading frame that was most abundant in the highest number of codons was different from the one that was most abundant in total. Furthermore, we required the most abundant reading frame was at least 5-fold more abundant than the second most abundant both in terms of the number of codons and in total. This resulted in read lengths 26, 29, 31-36 and 39 being selected for the mouse data. Out of these, read length 39 was later excluded due to having substantially fever reads and showing poor correlation to the others. Since the highest number of reads started 13 or 14 positions upstream of the TIS (depending on if frame 3 or 2 was most abundant), we trimmed one or two nucleotides of the reads in the subsequent analysis so that the P-site would still be located at read position 12-14 similar to the human data.

Extracting codon counts per position relative the A-site

The bam file for each sample with uniquely aligned reads was converted to bed format. *Bedtools intersect* was used to select reads with at least 50% overlap to Gencode-annotated coding sequences. Next, the reading frame of the 5' end of each read was determined using the frame information in the Gencode annotation. If the frame did not agree with the expected reading frame for that read length, the read was discarded. For the mouse data reads starting

with the third or second position of a codon had one or two nucleotides trimmed off the read to put all reads into the same reading frame. Then, nucleotide positions 1-27 were extracted from each read as nine codons, numbered as codon position -5 to +3, where 0 corresponded to the A-site. Next, we counted the number of occurrences of each codon per position and we validated the approach by verifying that codon counts calculated using different read lengths but corresponding to the same codon position (e.g., the P-site) correlated better to each other than to any other codon counts. Furthermore, counts from position -4 to -2 and +1 to +3 correlated well to the genome-wide distribution of codons in the human and mouse translatomes, whereas counts from the predicted P-site and A-site did not. We therefore concluded that we were counting codons from the correct frame for each included read length for both human and mouse.

Codon enrichment

The number of codon occurrences were counted separately for each ribosome-protected codon position and converted into a fraction of the total number. Normalized codon counts were obtained by dividing the codon fraction at a specific position by the mean fraction across all nine positions. Significant differences in normalized codon enrichment were calculated using Student's t test using the four biological replicates per cell state. The p-values were corrected for multiple testing using fdr correction. The figures are showing the signal across pooled replicates.

GO term analysis

Coding sequences were downloaded from the consensus CDS project. GO relations (go-basic.obo) were downloaded from the Gene Ontology Project (http://www.geneontology.org) and parsed using the Perl module GO::Parser. Relative codon frequencies were calculated per

gene and then averaged per GO term. Only GO terms with at least 40 genes were considered. The PCA analysis was performed in *R* using the method *princomp* using the correlation matrix.

Declarations

Acknowledgment

Funding

Availability of data and materials

The sequencing data used in our study have been deposited in NCBI's Gene Expression Omnibus and are accessible through the GEO accession number GSE123611:

Bornelöv S, Selmi T, Flad S, Dietmann S, Frye M. Codon usage optimization in pluripotent embryonic stem cells. Datasets. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123611(2019).

Authors' contributions

TS and SF designed and conducted the experiments. SB analyzed data and performed statistical analysis. SD provided guide for the analyses of the genome-wide data. TS, SB, MF designed and supervised the study and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

MF is consulting for STORM Therapeutics. All other authors declare that they have no competing interests.

Additional Files

Additional File 1: Figures S1 to S5 including figure legends (PDF; 1.7 MB)

Table S1: RNA-seq and Ribo-seq data in the four replicates of self-renewing (self) and differentiating (diff) human embryonic stem cells (XLSX; 4.2 MB)

Table S2: Gene enrichment analyses (EnrichR) identifies highly expressed genes in tissues derived from all three germ layers using ARCHS4 enrichment tool (XLSX; 12 KB).

Table S3: Distribution of hetADAT targets at ribosome covered codons in (a) differentiated versus self-renewing hESCS. The A-site is highlighted in yellow and (b) differentiated embryoid bodies (EB) versus self-renewing mESCS. The A-site is highlighted in yellow (XLSX; 14 KB).

Table S4: tRNA sequencing in self-renewing and differentiated (5 days) human embryonic stem cells (XLSX; 10 KB).

Table S5: tRNA sequencing to quantify inosine levels in differentiated (diff) and self-renewing (self) human embryonic stem cells (XLSX; 27 KB).

References

1. Kalkan, T. and A. Smith, *Mapping the route from naive pluripotency to lineage specification.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1657).
2. Dunn, S.J., et al., *Defining an essential transcription factor program for naive pluripotency.* Science, 2014. **344**(6188): p. 1156-1160.
3. Li, M. and J.C. Belmonte, *Ground rules of the pluripotency gene regulatory network.* Nat Rev Genet, 2017. **18**(3): p. 180-191.
4. Sharp, P.M. and W.H. Li, *An evolutionary perspective on synonymous codon usage in unicellular organisms.* J Mol Evol, 1986. **24**(1-2): p. 28-38.
5. Quax, T.E., et al., *Codon Bias as a Means to Fine-Tune Gene Expression.* Mol Cell, 2015. **59**(2): p. 149-61.
6. Gingold, H., et al., *A dual program for translation regulation in cellular proliferation and differentiation.* Cell, 2014. **158**(6): p. 1281-92.
7. Ingolia, N.T., *Ribosome profiling: new views of translation, from single codons to genome scale.* Nat Rev Genet, 2014. **15**(3): p. 205-13.
8. Kudla, G., et al., *Coding-sequence determinants of gene expression in Escherichia coli.* Science, 2009. **324**(5924): p. 255-8.
9. Rudolph, K.L., et al., *Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States.* PLoS Genet, 2016. **12**(5): p. e1006024.
10. Soll, D., et al., *Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique.* J Mol Biol, 1966. **19**(2): p. 556-73.
11. Crick, F.H., *Codon--anticodon pairing: the wobble hypothesis.* J Mol Biol, 1966. **19**(2): p. 548-55.
12. Agris, P.F., F.A. Vendeix, and W.D. Graham, *tRNA's wobble decoding of the genome: 40 years of modification.* J Mol Biol, 2007. **366**(1): p. 1-13.
13. Torres, A.G., E. Batlle, and L. Ribas de Pouplana, *Role of tRNA modifications in human diseases.* Trends Mol Med, 2014.
14. Schaffrath, R. and S.A. Leidel, *Wobble uridine modifications-a reason to live, a reason to die?!* RNA Biol, 2017. **14**(9): p. 1209-1222.
15. Rafels-Ybern, A., C.S. Attolini, and L. Ribas de Pouplana, *Distribution of ADAT-Dependent Codons in the Human Transcriptome.* Int J Mol Sci, 2015. **16**(8): p. 17303-14.
16. Van Haute, L., et al., *Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3.* Nat Commun, 2016. **7**: p. 12039.

17. Nakano, S., et al., *NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA(Met).* Nat Chem Biol, 2016. **12**(7): p. 546-51.
18. Haag, S., et al., *NSUN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation.* EMBO J, 2016.
19. Nedialkova, D.D. and S.A. Leidel, *Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity.* Cell, 2015. **161**(7): p. 1606-18.
20. Chan, C.T., et al., *Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins.* Nat Commun, 2012. **3**: p. 937.
21. Gu, C., T.J. Begley, and P.C. Dedon, *tRNA modifications regulate translation during cellular stress.* FEBS Lett, 2014. **588**(23): p. 4287-96.
22. Endres, L., P.C. Dedon, and T.J. Begley, *Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses.* RNA Biol, 2015. **12**(6): p. 603-14.
23. Chionh, Y.H., et al., *tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence.* Nat Commun, 2016. **7**: p. 13302.
24. Lu, J., et al., *All-trans retinoic acid promotes neural lineage entry by pluripotent embryonic stem cells via multiple pathways.* BMC Cell Biol, 2009. **10**: p. 57.
25. Bain, G., et al., *Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture.* Biochem Biophys Res Commun, 1996. **223**(3): p. 691-4.
26. Maden, M., *Retinoic acid in the development, regeneration and maintenance of the nervous system.* Nat Rev Neurosci, 2007. **8**(10): p. 755-65.
27. Eiselleova, L., et al., *A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells.* Stem Cells, 2009. **27**(8): p. 1847-57.
28. Ingolia, N.T., *Ribosome Footprint Profiling of Translation throughout the Genome.* Cell, 2016. **165**(1): p. 22-33.
29. Freimer, J.W., T.J. Hu, and R. Blelloch, *Decoupling the impact of microRNAs on translational repression versus RNA degradation in embryonic stem cells.* Elife, 2018. **7**.
30. Nakamura, M., et al., *Differentiation patterns of mouse embryonic stem cells and induced pluripotent stem cells into neurons.* Cytotechnology, 2016. **68**(3): p. 409-17.
31. Boroviak, T., et al., *Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis.* Dev Cell, 2015. **35**(3): p. 366-82.
32. Kalkan, T., et al., *Tracking the embryonic stem cell transition from ground state pluripotency.* Development, 2017. **144**(7): p. 1221-1234.
33. Mohammed, H., et al., *Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation.* Cell Rep, 2017. **20**(5): p. 1215-1228.
34. Cunningham, T.J. and G. Duester, *Mechanisms of retinoic acid signalling and its roles in organ and limb development.* Nat Rev Mol Cell Biol, 2015. **16**(2): p. 110-23.
35. Itskovitz-Eldor, J., et al., *Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.* Mol Med, 2000. **6**(2): p. 88-95.
36. Pruitt, K.D., et al., *The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes.* Genome Res, 2009. **19**(7): p. 1316-23.
37. Boel, G., et al., *Codon influence on protein expression in E. coli correlates with mRNA levels.* Nature, 2016. **529**(7586): p. 358-363.
38. Presnyak, V., et al., *Codon optimality is a major determinant of mRNA stability.* Cell, 2015. **160**(6): p. 1111-24.
39. Bazzini, A.A., et al., *Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition.* EMBO J, 2016. **35**(19): p. 2087-2103.
40. Wu, Q., et al., *Translation affects mRNA stability in a codon dependent manner in human cells.* Elife, 2019. **8**.
41. Newman, Z.R., et al., *Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9.* Proc Natl Acad Sci U S A, 2016. **113**(10): p. E1362-71.
42. Barahimipour, R., et al., *Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii.* Plant J, 2015. **84**(4): p. 704-17.
43. Gromadski, K.B., T. Daviter, and M.V. Rodnina, *A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity.* Mol Cell, 2006. **21**(3): p. 369-77.
44. Hanson, G. and J. Coller, *Codon optimality, bias and usage in translation and mRNA decay.* Nat Rev Mol Cell Biol, 2018. **19**(1): p. 20-30.
45. Ingolia, N.T., et al., *Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes.* Cell Rep, 2014. **8**(5): p. 1365-79.
46. Pavlov, M.Y., et al., *Slow peptide bond formation by proline and other N-alkylamino acids in translation.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 50-4.
47. Fluitt, A., E. Pienaar, and H. Viljoen, *Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis.* Comput Biol Chem, 2007. **31**(5-6): p. 335-46.
48. Varenne, S., et al., *Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains.* J Mol Biol, 1984. **180**(3): p. 549-76.
49. Ingolia, N.T., L.F. Lareau, and J.S. Weissman, *Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes.* Cell, 2011. **147**(4): p. 789-802.

50. Smith, A.G., et al., *Differentiation inhibiting activity (DIA/LIF) and mouse development.* Dev Biol, 1992. **151**(2): p. 339-51.

51. Williams, R.L., et al., *Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells.* Nature, 1988. **336**(6200): p. 684-7.

52. Wobus, A.M., et al., *Embryonic stem cells as a model to study cardiac, skeletal muscle, and vascular smooth muscle cell differentiation.* Methods Mol Biol, 2002. **185**: p. 127-56.

53. Agris, P.F., *The importance of being modified: an unrealized code to RNA structure and function.* RNA, 2015. **21**(4): p. 552-4.

54. Tsutsumi, S., et al., *Wobble inosine tRNA modification is essential to cell cycle progression in G(1)/S and G(2)/M transitions in fission yeast.* J Biol Chem, 2007. **282**(46): p. 33459-65.

55. Rafels-Ybern, A., et al., *Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla.* RNA Biol, 2017: p. 1-8.

56. Gerber, A.P. and W. Keller, *An adenosine deaminase that generates inosine at the wobble position of tRNAs.* Science, 1999. **286**(5442): p. 1146-9.

57. Torres, A.G., et al., *A-to-I editing on tRNAs: biochemical, biological and evolutionary implications.* FEBS Lett, 2014. **588**(23): p. 4279-86.

58. Sprinzl, M., et al., *Compilation of tRNA sequences and sequences of tRNA genes.* Nucleic Acids Res, 1998. **26**(1): p. 148-53.

59. Khade, P.K. and S. Joseph, *Messenger RNA interactions in the decoding center control the rate of translocation.* Nat Struct Mol Biol, 2011. **18**(11): p. 1300-2.

60. Shah, P., et al., *Rate-limiting steps in yeast protein translation.* Cell, 2013. **153**(7): p. 1589-601.

61. Blanco, S., et al., *Stem cell function and stress response are controlled by protein synthesis.* Nature, 2016. **534**(7607): p. 335-40.

62. Signer, R.A., et al., *Haematopoietic stem cells require a highly regulated protein synthesis rate.* Nature, 2014. **509**(7498): p. 49-54.

63. Llorens-Bobadilla, E., et al., *Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury.* Cell Stem Cell, 2015. **17**(3): p. 329-40.

64. Sampath, P., et al., *A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation.* Cell Stem Cell, 2008. **2**(5): p. 448-60.

65. Raina, M. and M. Ibba, *tRNAs as regulators of biological processes.* Front Genet, 2014. **5**: p. 171.

66. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.* Science, 2009. **324**(5924): p. 218-23.

67. Gerashchenko, M.V., A.V. Lobanov, and V.N. Gladyshev, *Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress.* Proc Natl Acad Sci U S A, 2012. **109**(43): p. 17394-9.

68. Hsieh, A.C., et al., *The translational landscape of mTOR signalling steers cancer initiation and metastasis.* Nature, 2012. **485**(7396): p. 55-61.

69. Thoreen, C.C., et al., *A unifying model for mTORC1-mediated regulation of mRNA translation.* Nature, 2012. **485**(7396): p. 109-13.

70. Zhou, Z., et al., *Codon usage is an important determinant of gene expression levels largely through its effects on transcription.* Proc Natl Acad Sci U S A, 2016. **113**(41): p. E6117-E6125.

71. Ranjan, N. and M.V. Rodnina, *tRNA wobble modifications and protein homeostasis.* Translation (Austin), 2016. **4**(1): p. e1143076.

72. Rubio, M.A., et al., *An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 7821-6.

73. Rubio, M.A., et al., *Editing and methylation at a single site by functionally interdependent activities.* Nature, 2017. **542**(7642): p. 494-497.

74. Ingolia, N.T., et al., *The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments.* Nat Protoc, 2012. **7**(8): p. 1534-50.

75. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.

76. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res, 2012. **22**(9): p. 1760-74.

77. Mudge, J.M. and J. Harrow, *Creating reference gene annotation for the mouse C57BL6/J genome assembly.* Mamm Genome, 2015. **26**(9-10): p. 366-78.

78. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

79. Chan, P.P. and T.M. Lowe, *GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes.* Nucleic Acids Res, 2016. **44**(D1): p. D184-9.

80. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.

81. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.

82. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

83. Nettling, M., et al., *DiffLogo: a comparative visualization of sequence motifs.* BMC Bioinformatics, 2015. **16**: p. 387.

Figure Legends

Figure 1. **Profiling mRNA expression and translation in of self-renewing and differentiating hESCs.** (**A**) Treatment regime to differentiate the hESCs (H9) by removing FGF2 from and adding retinoic acid (RA) to the culture medium. Four replicates of self-renewing (H9^{self}) and differentiating (H9^{diff}) hESCs were used for RNA-seq and Ribo-seq analyses. (**B**) Expression heatmap of pluripotency (blue) and lineage (red) markers measured by RNA-seq and Ribo-seq. Expression shown as log$_2$ difference to the mean. Markers in each group are ordered by decreasing expression. (**C, D**) MA-plot showing the log$_2$ fold change of differentiated (diff) versus self-renewing (self) hESCs against mean expression in either RNA-seq (**C**) or Ribo-seq (**D**). Examples of significant genes related to differentiation are shown in red. (**E-G**) RT-qPCR confirming similar change in RNA levels of pluripotency markers (upper panels) and lineage markers (lower panels) in hESCs RA-differentiated for 5 days (**E**) and in embryoid bodies grown for 5 days (**F**), or 7 days (**G**). Shown is the mean (n=3-4). Error bars s.d.

Figure 2. **Genomic GC-content influences codon usage**. (**A-F**) Overview of codon (**A, B, D, E**) and amino acid (**C, F**) enrichment in differentially expressed genes measured by Ribo-seq (**A-C**) and RNA-seq (**D-F**). Enrichment was calculated as log$_2$ fold-change of codon or amino acid frequency in differentiation or self-renewal genes relative to all genes. Codons are color coded according to their third nucleotide (**A, D**) and are further separated by *nn*G codon degeneracy (**B, E**). Dotted circles highlight the most enriched codons in response to the retinoic stimulus.

Figure 3. **Genes from different GO categories show different codon frequencies.** (**A-C**) Codon frequencies were calculated per Gene Ontology (GO) term and PCA was used to reveal a global separation into high and low GC genes (**A**). GO terms are colored according to total GC (**A**), GC at the third codon position (**B**), or at the first and second position (**C**). Examples of GO terms are listed next to the graphs. (**D**) PCA loadings plot identifying the codons contributing the location of the GOs in (**A-C**). Codons are color coded by third position base and/or codon degeneracy. The separation across the x axis is comparable to the separation across the diagonal in figure 2A, B, D-E.

Figure 4. **Distinct codon signature at the ribosome A- and P-site in self-renewing ESCs**. (**A**) Illustration of a ribosome binding to mRNA protecting at least nine codons (position -5 to +3) from degradation by the RNase digestion. Positions -2, -1 and 0 correspond to the E-, P-, and A-site that host the charged tRNA. (**B**) The raw codon frequency was calculated for each specific codon (e.g. A-site) and normalized to the mean frequency across all nine positions. Calculations were done separately for self-renewing and differentiating samples. (**C-E**) Codon enrichment at the E- (**C**), P- (**D**), and A-site (**E**). Stop codons are highlighted in red and Proline codons are highlighted in blue. The codons are sorted according to their frequency (shown in **Additional File 1: Fig. S3A-C**). (**F, G**) Codon enrichment of stop (**F**) or all other codons (**G**) across all ribosome-protected codons. (**H**) Log_2 fold-change of normalized codon usage in differentiated versus self-renewing hESCs across all ribosome-protected codons. Significantly different codons (Welch's t-test; fdr correction) are marked in red (enriched) and blue (reduced) and occur mainly at the P-site. (**I**) Log_2 fold-change of normalized codon usage in mouse differentiated (embryoid bodies) versus self-renewing embryonic stem cells (mESCs). Blue and red dots represent the significantly changed codon shown in (**H**).

Figure 5. **Distinct codon usage signature in pluripotent cells are hetADAT-sensitive codons.** (**A**) Illustration of hetADAT (ADAT2 and ADAT3)-mediated adenosine-to-inosine tRNA editing at position 34 in the anti-codon loop and the codons read by unmodified and modified tRNAs. (**B**) HetADAT-dependent codons (red) that sometimes are lacking corresponding tRNA genes (light gray). (**C-F**) Volcano plot (**C, E**) showing significance versus codon usage fold change at the ribosome P-site (**C**) and A-site (**E**) in differentiated versus self-renewing hESCs. HetADAT-dependent codons are highlighted in red. Correlation (**D, F**) between $\log_2$ fold change of codon usage frequencies in differentiated human (hESCs) with differentiated mouse (mESCs) embryonic stem cells through embryoid body culture (EB; left hand panel) or by removal of LIF (right hand panel) at the P-site (**D**) and A-site (**F**). (**G**) Log_2 fold change of codon usage frequency across all sites protected by the ribosome in human (left hand panel) and mouse ESCs differentiated through EB culture (middle panel) or removal of LIF (right hand panel). HetADAT-dependent codons are highlighted in red.

Figure 6. **HetADAT-dependent translation in mouse and human ESCs**. (**A**) RT-qPCR confirming down-regulation of ADAT2 mRNA levels in differentiated hESCs (Diff) and embryoid bodies (EB) compared to self-renewing hESCs (Self). * $p < 0.05$ one-way ANOVA. (**B**) Distribution of wobble inosine (A34I) levels per tRNA (>10 reads) shown as Violin plot with median (black line) and quartiles (dotted line). (**C**) Differences of inosine levels at A34 in differentiated versus self-renewing hESC per tRNA isotype. (**D**) Light microscopic image of self-renewing (self) human embryonic stem cells (hESCs), differentiated hESCs for 5 days (5d), embryoid bodies (EBs) after 5, and 7 days in suspension. (**E**) Quantification of protein expression using Western blotting of OCT4, ADAT2, and ADAT3 in self-renewing (grey) and differentiating (red) hESCs for one day. ** $p < 0.01$ unpaired t-test. (**F**) Quantification of protein levels using Western blotting of ADAT2 and ADAT3 after differentiating hESCs into EBs for

7 days (7d). * $p<0.05$; ** $p<0.01$ unpaired t-test. (**G**) Log_2 fold-change gene translation (Ribo-seq) of indicated proteins in mouse self-renewing ESCs versus ESCs differentiated by removal LIF or grown as embryoid bodies (EB). ****$p<0.0001$. (**H**) . Ribo-seq fold changes (diff/self) was compared with RNA-seq fold changes (diff/self). Genes with ribosome occupancy difference 2-fold higher (red), 2-fold lower (blue) or similar (grey) to the mRNA level difference was identified among genes rich in hetADAT-dependent *nn*C codons (left) and among all genes (right). The hetADAT-dependent group was significantly enriched for genes with altered (higher and lower) ribosome occupancy.

Figure 1
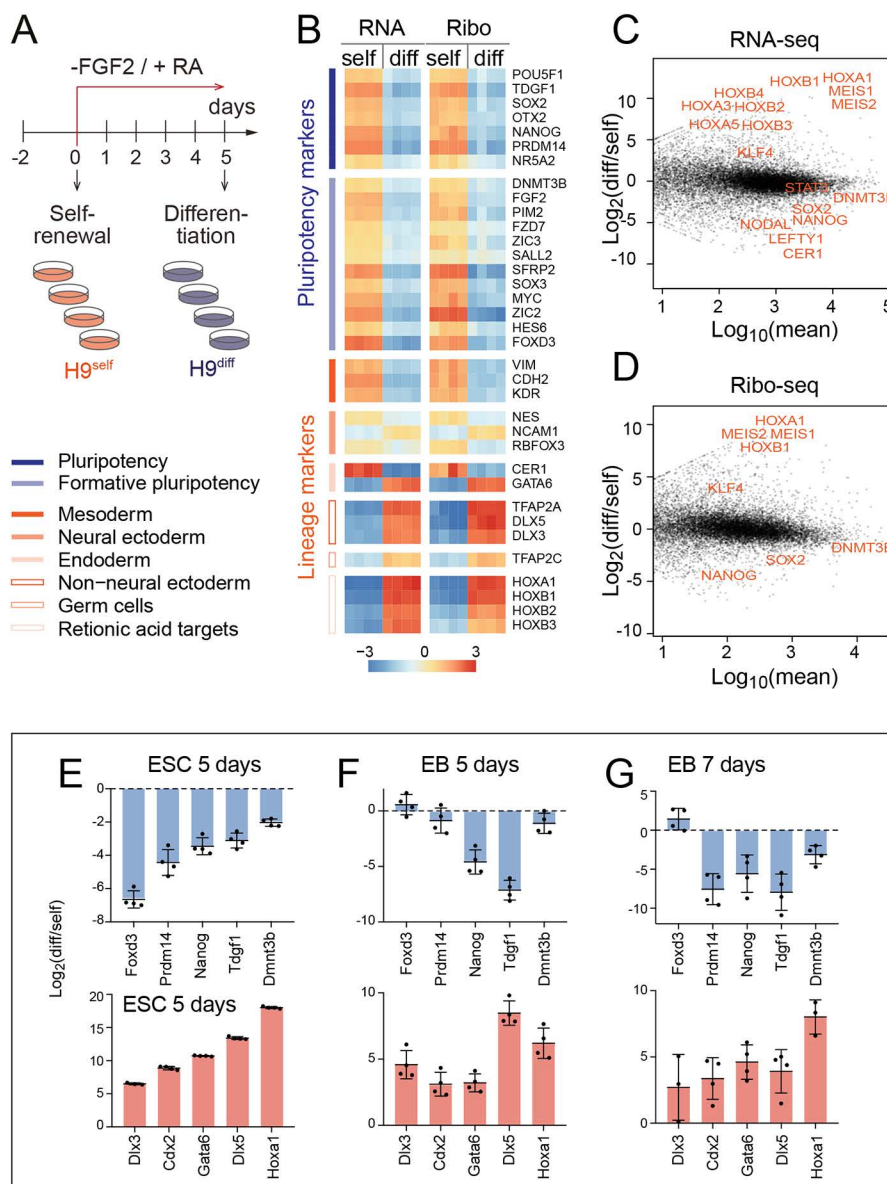
Figure 2
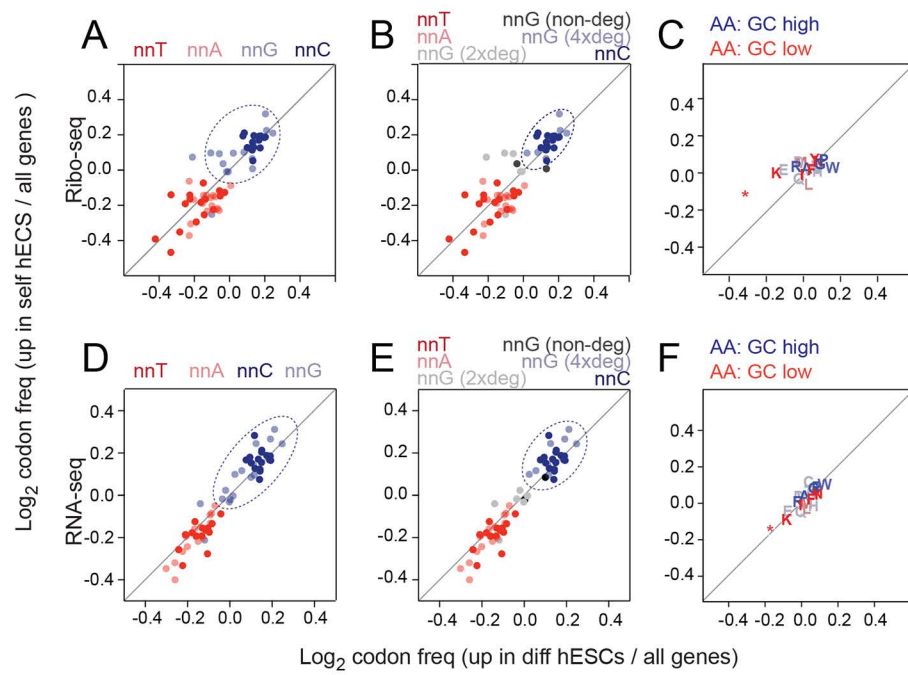
Figure 3
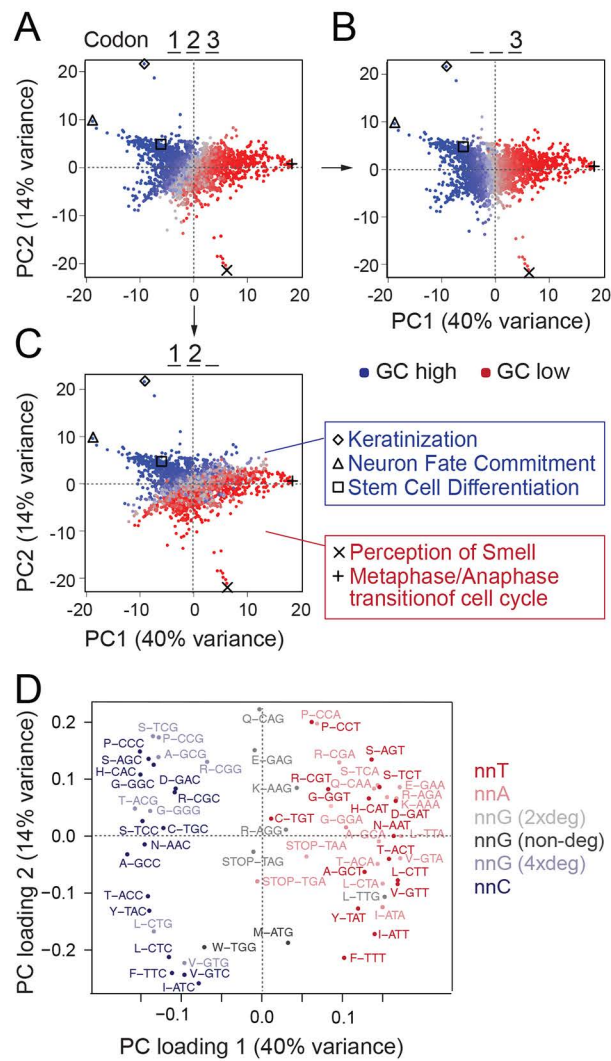
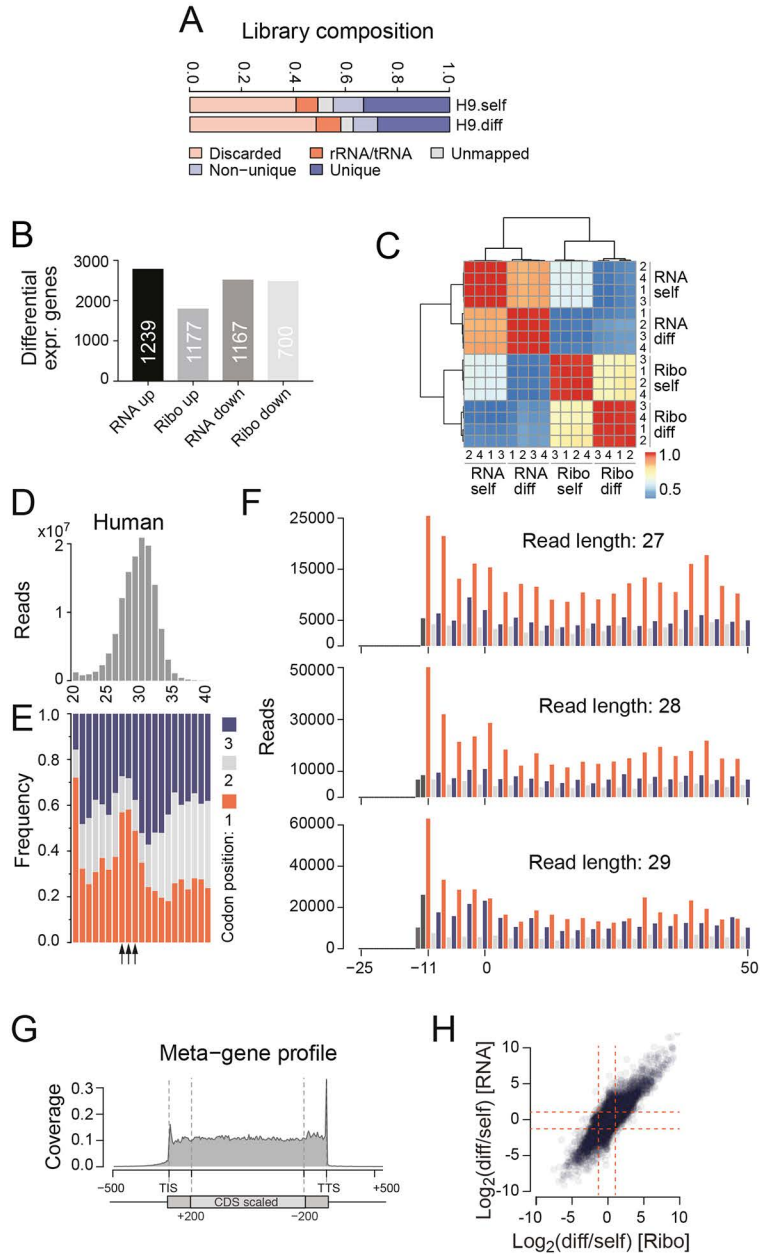# Figure 4
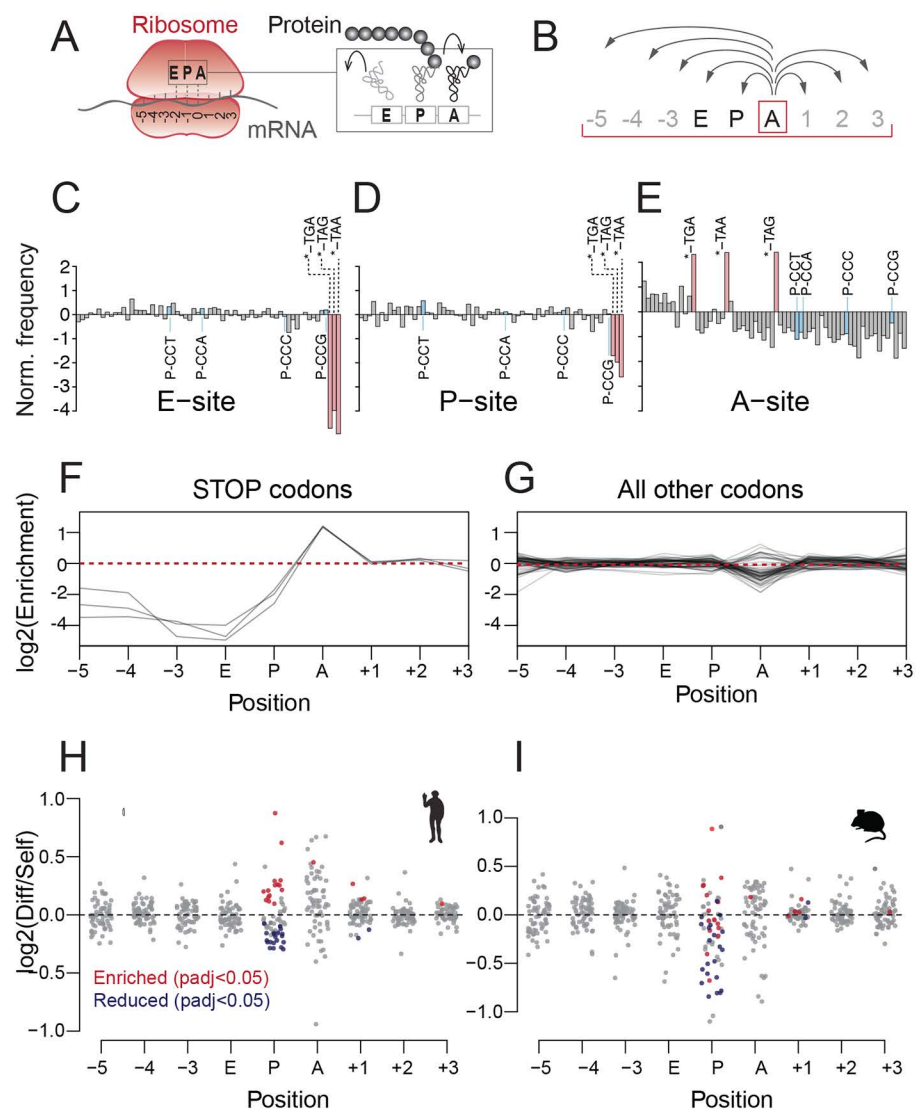
Figure 5
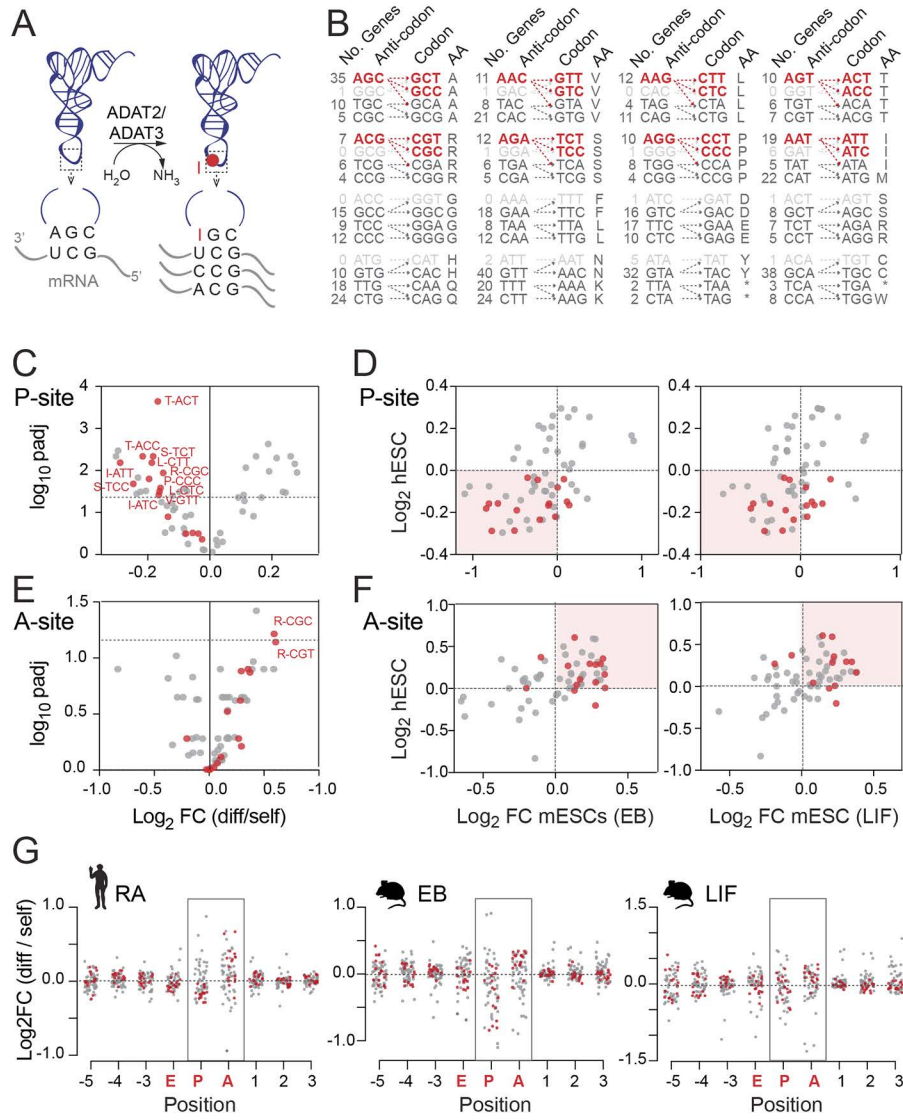
Figure 6
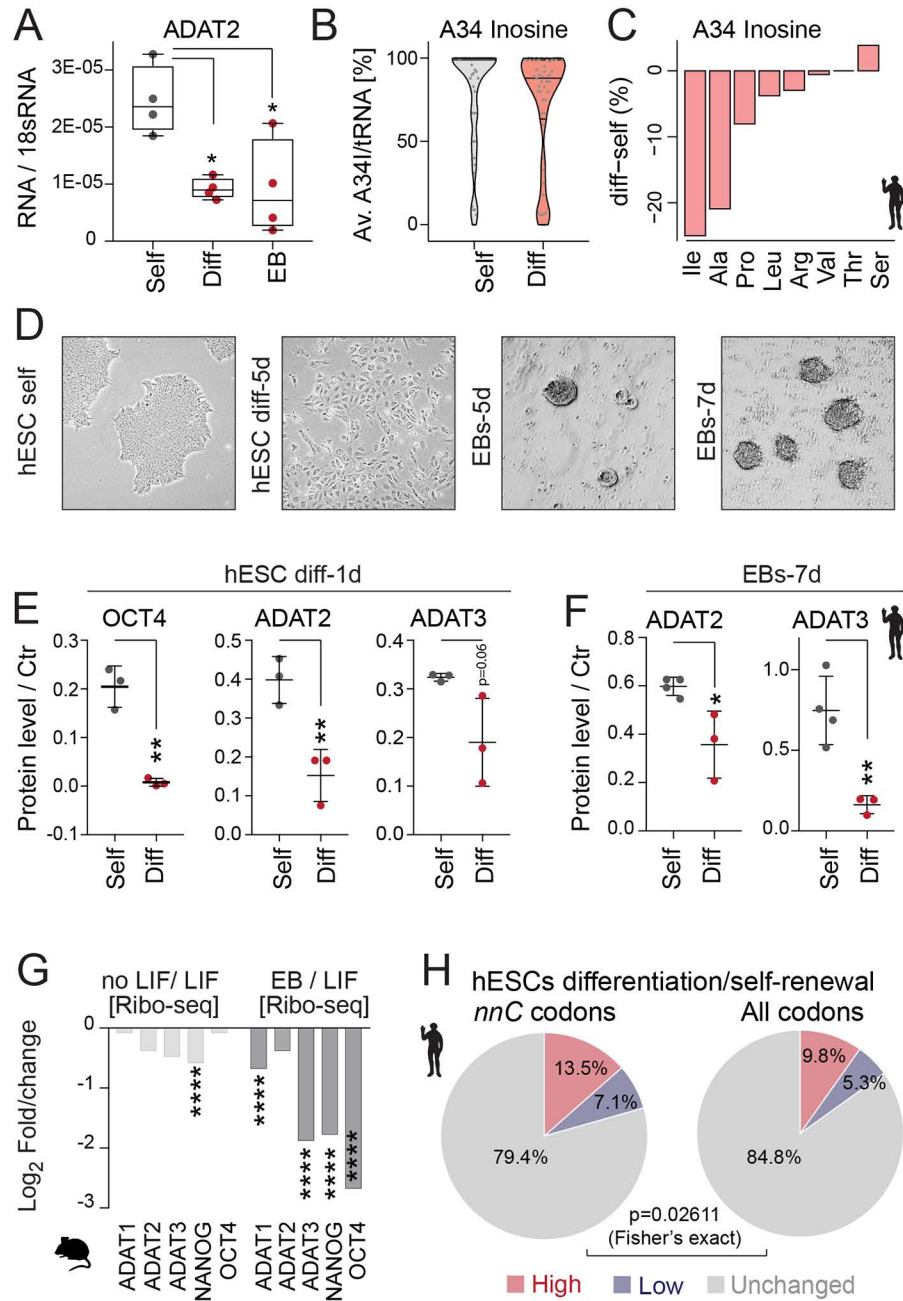
Figure S1. Quality check of ribosome profiling data. (A) Library composition of Ribo-seq dataset. **(B)** Number of differentially expressed genes (padj<0.05; log$_2$ FC>2 or FC<-2) in RNA- and Ribo-seq datasets. **(C)** Correlation plot between replicates and datasets (RNA-seq and Ribo-seq) in self-renewing (self) and differentiating (diff) hESCs. **(D, E)** Read length distribution **(D)** and reading frame **(E)** of the 5' end of the reads. **(F)** High triplet-periodicity near the start codon (position 1-3) of aligned reads with lengths of 27 to 29 nucleotides. **(G)** Metagene profile of aligned reads over the coding sequence (CDS) and the adjacent 5' and 3' untranslated regions. The first and last 200 nucleotides of the CDS are shown unscaled and the remaining CDS was scaled to equal length for all genes. TIS: translation initiation site. TTS: translation termination site. **(H)** Correlation between differentially expressed (RNA-seq) and translated (Ribo-seq) genes.
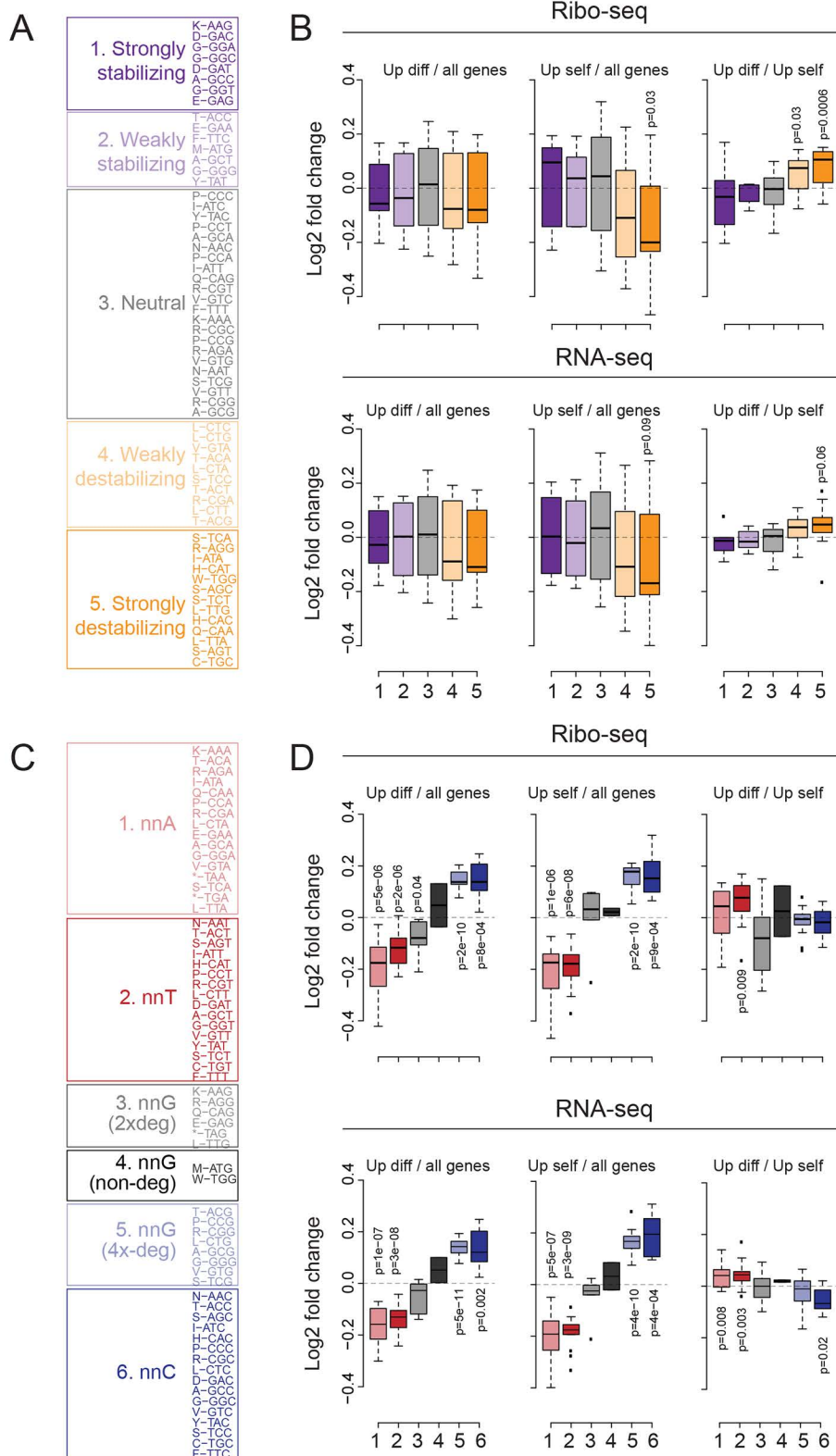
Figure S2. Optimal codons and GC-content contribute to mRNA levels. (A, B) Codons were classified according to their stabilizing or de-stabilizing effect on mRNAs (**A**) and their enrichment measured in up-regulated genes in self-renewing (self) and differentiating (diff) human embryonic stem cells (hESCs) using Ribo-seq or RNA-seq datasets (**B**). (**C**, **D**) Codon enrichment by the third nucleotide and degeneracy (**C**) in up-regulated genes in self and diff hESCs using Ribo-seq or RNA-seq datasets. The righthand columns (**B, D**) compare the difference of stabilizing or de-stabilizing codons (**B**) or codon composition (**D**) between up-regulated RNAs in diff and self hESCs.
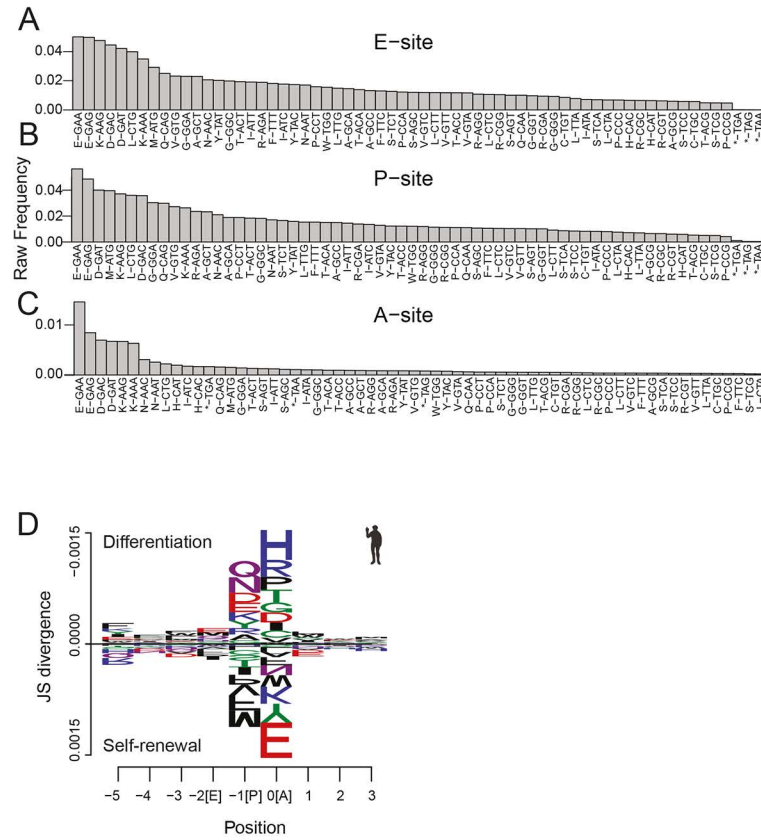
Figure S3. Codon and amino acid frequencies. (A-C) Ratio of raw codon frequency in differentiating versus self-renewing human embryonic stem cells (four pooled replicates) at the ribosome E- (**A**), P- (**B**), and A-(**C**) sites. (**D**) DiffLogo analyses showing the difference in normalized amino acid occupancy between differentiating and self-renewing hESCs at the E, P, and A positions and the flanking positions (-5..-3, +1..+3) covered by the ribosome.
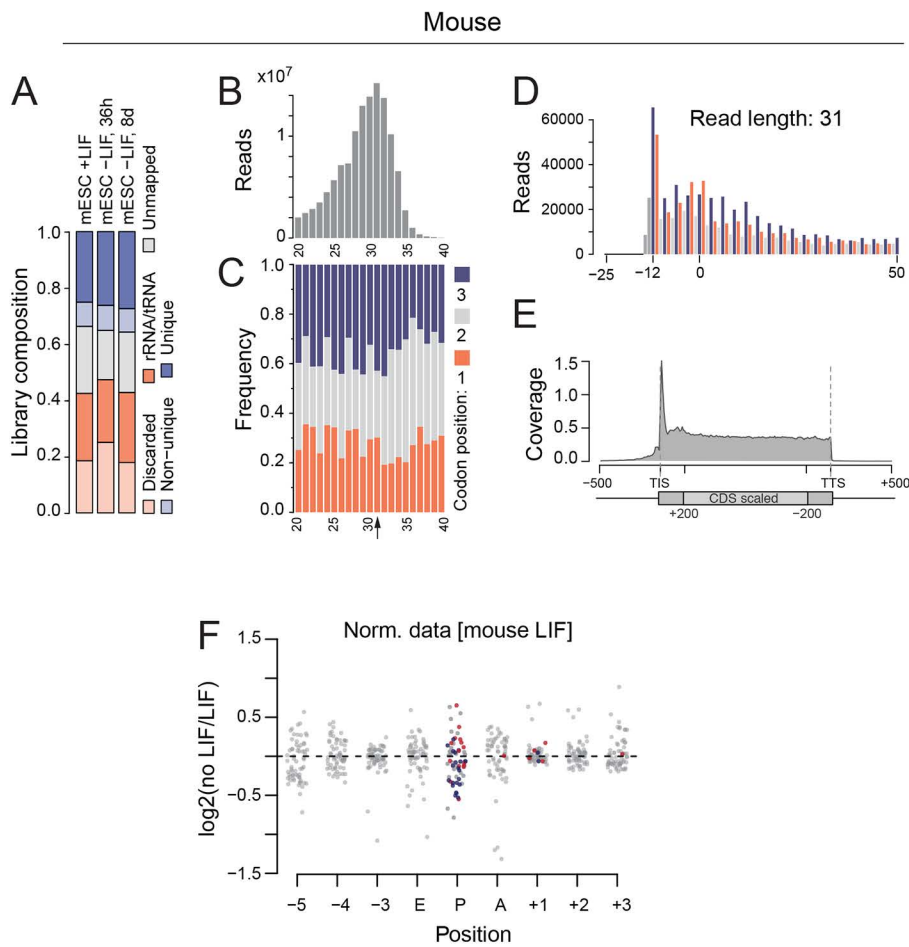
Figure S4. Quality check of mouse ribosome profiling data. (A) Library composition of Ribo-seq dataset. (**B, C**) Read length distribution (**B**) and reading frame (**C**) of the 5' end of the reads. (**D**) Example of triplet-periodicity near the start codon (position 1-3) of aligned reads with lengths of 31 nucleotides. Read lengths 26, 29 and 31 to 36 were used for the analyses. (**E**) Metagene profile of aligned reads over the coding sequence (CDS) and the adjacent 5' and 3' untranslated regionsThe first and last 200 nucleotides of the CDS are shown unscaled and the remaining CDS was scaled to equal length for all genes. TIS: translation initiation site. TTS: translation termination site. (**F**) Log$_2$ fold-change of normalized codon usage in mouse differentiated (removal of LIF) versus self-renewing embryonic stem cells (mESCs). Blue and red dots represent the significantly changed codon shown in figure 3H.
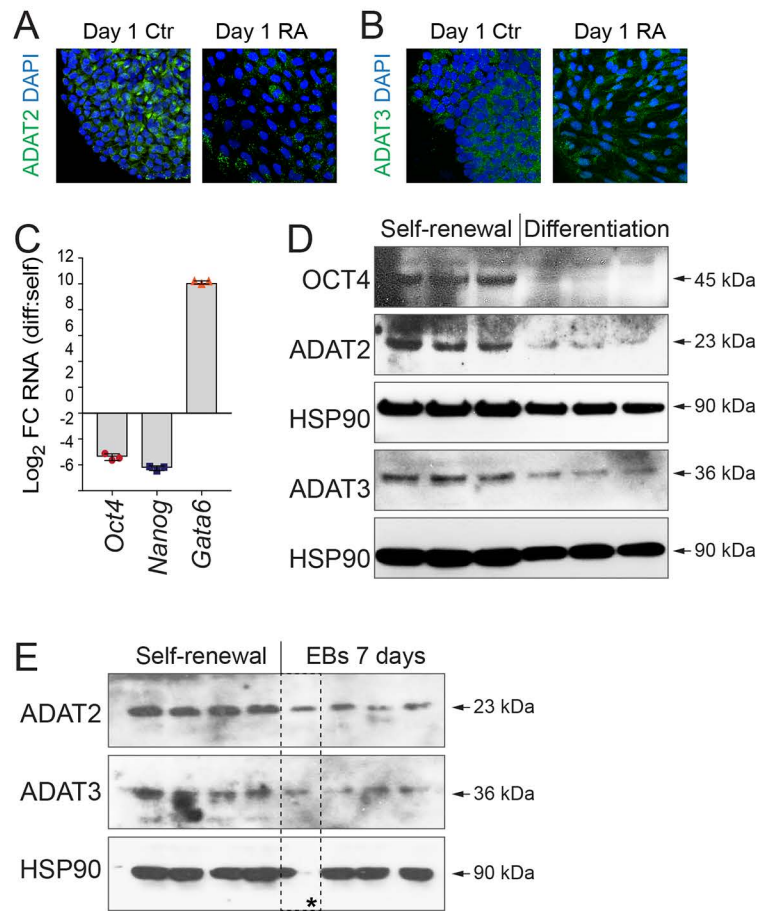
Fig S5

Figure S5. ADAT2 expression in hESCs. (A, B) Immunofluorescence staining for ADAT2 (**A**) and ADAT3 (green) (**B**) in control (ctr) (left hand panels) and retinoic acid (RA)-induced (right hand panels) hESCs for one day. DAPI (blue) counterstains nuclei. (**C**) Log$_2$ fold change in RNA levels of the indicated genes measured by qPCR. (**D**) Western Blot to quantify expression of the indicated proteins (shown in Figure 6E) in self-renewing and differentiated hESCs. (**E**) Western Blot to quantify expression of the indicated proteins (shown in Figure 6F) in self-renewing hESCs and hESCs differentiated into embryoid bodies (EBs). Asterisk: Excluded from the quantification.