Abundant circular RNA expression in terminal myeloid differentiation does not correlate with frequent circRNA translation, or with circRNA-mediated translation regulation

Benoit P. Nicolet[1,2]*, Sjoert B.G. Jansen[3], Esther Heideveld[1], Willem H. Ouwehand[3], Emile van den Akker[1], Marieke von Lindern[1], and Monika C. Wolkers[1,2]*

[1] Department of Hematopoiesis, Sanquin Research and Landsteiner Laboratory, Amsterdam UMC, University of Amsterdam, The Netherlands.
[2] Oncode Institute, Utrecht, The Netherlands
[3] Department of Haematology, University of Cambridge and NHS Blood and Transplant, UK.

***** To whom correspondence should be addressed. M. Wolkers or B. Nicolet, Sanquin Research, Department of Hematopoiesis, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands. Tel: +31-20-5127003; Fax: +31-20-5123474; Emails: m.wolkers@sanquin.nl | b.nicolet@sanquin.nl

ABSTRACT:

Each day, about 10^{12} erythrocytes and platelets are released into the blood stream. This substantial output from hematopoietic stem cells is tightly regulated by transcription factors and epigenetic modifications. Whether and how non-coding RNAs such as circular RNAs (circRNAs) contribute to the differentiation and/or identity of hematopoietic cells is to date not well understood. We recently reported a circRNA expression map of hematopoietic cells, revealing that erythrocytes and platelets contain the highest levels and numbers of circRNAs. Here, we provide the first detailed and comprehensive analysis of circRNA expression during red blood cell and megakaryocyte differentiation. CircRNA expression significantly increased during erythroid precursor differentiation into red blood cells and in differentiating megakaryocytes, in particular upon enucleation. To determine if circRNAs can modulate hematopoietic differentiation, we compared the expression levels of circRNAs and mRNAs with that of ribosomal foot printing read. This multi-omics approach revealed that only 20 out of 748 (2.6%) circRNAs associated with translation regulation of their mRNA counterparts. Furthermore, irrespective of the thousands of identified putative open reading frames in circRNAs, deep ribosome-footprinting sequencing and mass spectrometry analysis provided little evidence for translation of endogenously expressed circRNAs in erythroblasts, megakaryocytes and platelets. In conclusion, circRNAs in platelets and red blood cells are highly abundant and alter their expression profile during differentiation, yet their contribution to regulate cellular processes remains enigmatic.

INTRODUCTION

Each hour, millions of red blood cells (RBC) and platelets are produced in the bone marrow and released into the blood stream. Their production is mediated from erythroid and megakaryoid progenitors. The generation of these progenitors from hematopoietic stem cells is a highly orchestrated process. Tight regulation is obtained by the coordinated expression of transcription factors, long non-coding RNAs and micro-RNAs (Goode et al., 2016; Luo et al., 2015; Petriv et al., 2010). Recently, we showed that also circular RNAs (circRNAs) are abundantly expressed in hematopoietic cells, and that their expression alters during differentiation (Nicolet et al., 2018).

CircRNAs are single stranded, circular RNA molecules that are generated from pre-mRNA transcripts through back-splicing (Chen, 2016; Hansen et al., 2013; Memczak et al., 2013). Back-splicing is driven by the classical spliceosome machinery at canonical splice sites (Ashwal-Fluss et al., 2014), and circularization is mediated by pairing of complementary sequences, typically Alu elements, which are found in flanking introns. The pairing of Alu elements is regulated by RNA-binding proteins such as ADAR, DHX9, and NF90 (Aktaş et al., 2017; Li et al., 2017; Rybak-Wolf et al., 2015). CircRNA can contain both intron and exon segments. While intronic circRNA are mostly retained in the nucleus (Li et al., 2015), exonic circRNA undergo nuclear translocation in a size dependent manner (Huang et al., 2018; Jeck et al., 2013). In the cytoplasm, circRNAs are found to be exonuclease-resistant (Suzuki et al., 2006). In addition, circRNA lack the 3' poly-adenylation tail, which renders them more stable than linear RNA (Jeck et al., 2013).

Many functions have been attributed to circRNAs. CircRNAs were suggested to act as transcriptional activators (Li et al., 2015; Zhang et al., 2016), or to serve as RNA-binding protein cargo (Hentze and Preiss, 2013). CircRNAs can also act as miRNA 'sponge', a feature that is to date limited to about ten circRNAs genome-wide in humans (Guo et al., 2014; Piwecka et al., 2017; Stagsted et al., 2019). In addition, circRNA-mediated regulation of mRNA translation has been proposed in cell lines (Li et al., 2020; Sun et al., 2019; Wu et al., 2019). Lastly, it was shown with over-expression constructs in cell lines (Legnini et al., 2017; Yang et al., 2017), and from exemplary circRNAs in drosophila and human cell lines that circRNA can be translated into protein (Liang et al., 2019; Pamudurti et al., 2017). How these findings

translate to endogenous circRNA expression during human hematopoietic differentiation, is not known.

We and others have observed that RBCs and platelets had the highest numbers and highest circRNA content of all analyzed mature cell types in humans (Alhasan et al., 2016; Maass et al., 2017; Nicolet et al., 2018). However, the origin of these high circRNA levels is not known, and has thus far been attributed to transcriptome degradation (Alhasan et al., 2016). In addition, the function of these circRNA in blood cell development remains unknown.

In the study presented here, we provide a detailed and comprehensive analysis of circRNA expression in erythroid and megakaryoid development. We identified >20,000 circRNAs in erythroid differentiation, and >6,000 in megakaryoid differentiation. The expression of circRNAs alters dramatically upon enucleation and platelet formation, and only partially (~40%) follows mRNA expression in differentiating blood cells. Integration of RNA-seq and ribo-seq data revealed that circRNA-mediated regulation of mRNA translation is not the main mode of action during megakaryocytic maturation. Furthermore, ribosome-footprinting analysis and mass spectrometry analysis demonstrated that translation of endogenous circRNAs may occur in reticulocytes, megakaryocytes, and/or platelets, however this is not a frequent event, Thus, even though circRNAs are highly prevalent in erythroid and megakaryoid development, their mode of action and activity in hematopoietic cells remains to be defined.

RESULTS

CircRNA are abundant during erythropoiesis

We first investigated if and how circRNA expression alters during terminal erythropoiesis. To this end, we examined paired-end RNA sequencing data of erythroblasts isolated from 4 donors that were cultured for 12 days under differentiating conditions, and that were harvested on each day between day 0 to 7, in addition to day 9 and 12 (Figure 1A;(Heshusius et al., 2019)). The average sequencing depth of 30 million reads identified on average 27.3 million mapped reads. To detect, quantify, and to annotate circRNAs from chimeric reads, we used our established pipeline (Nicolet et al., 2018) that is based on the two algorithms DCC (Cheng et al., 2015) and CircExplorer2 (CE2, (Zhang et al., 2014); see methods section)). With a 'low-confidence' cut-off of at least 2 reads in at least one sample at any time of harvest, 20,931 circRNAs were co-detected by DCC and CE (Figure 1B; Table S1). Based on mRNA splicing annotations and the circRNA junction reads, the estimated number of exons/circRNA was 5.58 in differentiating erythroblasts (median= 4 exons/circRNA; Figure 1C). The putative length of circRNAs was 846.14 nt (median= 619 nt; Figure 1D). In line with our previous work (Nicolet et al., 2018), circRNAs preferentially use the second exon of a linear transcript as the first circular exon (Figure 1E), a bias that is not as pronounced for the last circularized exon (Figure 1F). Thus, circRNAs are abundantly expressed during erythropoiesis in comparison to early hematopoiesis (Nicolet et al., 2018).

CircRNA levels increase at enucleation

To define how circRNA expression related to the expression of other transcripts during erythroid differentiation, we quantified the overall expression of protein-coding transcripts (mRNA), non-coding transcripts (ncRNA) and circRNA transcripts (TPM>0.1). We detected 19,449 mRNA-, 10,753 ncRNA-, and 4,526 circRNA-producing genes (Figure 2A). We also quantified the diversity of transcripts expressed during erythrocyte differentiation. At the early time points (i.e. up do day 5), erythroblasts mainly expressed mRNA and ncRNA transcripts (Figure 2B). These numbers steadily decreased over time and were reduced by 82.96% and 76.45%, respectively, at day 12 compared to day 0 (Figure 2B). In contrast, the number of circRNA transcripts substantially increased by a ~3.8-fold from day 6 onwards compared to day 0, and circRNA transcripts even surpassed the number of ncRNA

transcripts at day 8 (Figure 2B). Of note, the steep increase of circRNAs coincided with the enucleation rate (Figure 2C).

To measure the differential expression of circRNAs during erythropoiesis, we focused on 'high confidence' circRNAs, that is, a given circRNA is detected at least twice in all 4 donors at an individual time-point. This filter identified 950 circRNAs that were co-detected by DCC (n=1163) and CE2 (n=958; Figure 2D; Table S1). Pearson's correlation coefficient of all measured time points during erythropoiesis revealed 2 main clusters of circRNAs that were primarily expressed early (day 0-5), or late during differentiation (day 6-12; Figure 2E). In fact, a heatmap representation of high confidence circRNA expression showed a massive increase in circRNA expression from day 6 onwards (Figure S1A). When we compared all days to day 0 (Figure 2F), 136 circRNAs were differentially expressed (p-adjusted<0.05; Table S1). Of these, 16 circRNAs were highly expressed at day 0 and decreased towards day 6, and 119 circRNAs significantly increased their expression levels from day 6 onwards (Figure 2F; Table S1). Thus, circRNA expression substantially increases at enucleation, yet a small group of circRNA show the reverse expression pattern, suggesting that the generation of circRNAs is an active process rather than a simple accumulation in time.

We next investigated how circRNA expression related to the expression of their mRNA counterparts in erythroid differentiation. To this end, we quantified the linear mRNA read counts at the circRNA start and end genomic position, using the build-in linear detection of DCC, as described previously (Nicolet et al., 2018). The circular-over-linear expression ratio (CLR) was calculated for each 'high confidence' circRNA, averaged across the 4 biological replicates for each time-point (Figure S1B; Table S1). Between day 0 and day 5 of erythropoiesis, the mRNA expression was higher compared to circRNA, as no circRNAs showed a CLR above 1. Only from day 6 onwards, 25 circRNAs in total showed a CLR>1, that is, circRNA expression is higher than that of their linear counterpart. Also the overall CLR increased from day 6 onwards and peaked at day 9 with a 7.3-fold increase compared to day 0 (Figure 2G). In conclusion, circRNA expression substantially increases at enucleation and can reach levels that are higher than those of their mRNA counterpart.

CircRNA expression alters from megakaryocytes to platelets

We next investigated how circRNAs are expressed during megakaryopoiesis and in platelets. We performed RNA-seq analysis of three donors on purified immature $CD41a^+$ $CD42b^-$ (CD42-) and mature $CD41a^+$ $CD42b^+$ (CD42+) megakaryocytes that were cultured from cord blood $CD34^+$ progenitors (see methods; Figure 3A). The average sequencing depth of 84.4 million reads per sample resulted in on average 74.6 million mapped reads to the genome. CircRNA detection with DCC and CE2 co-detected 6,751 circRNAs with low-confidence filtering (Figure 3B; Table S2). To define the relation of mRNA, ncRNA and circRNA expression during MK differentiation, we quantified the number of genes expressing these transcripts. We also included our previous analysis of circRNA expression in platelets to this analysis pipeline (Nicolet et al., 2018). Similar to erythropoiesis, the number of expressed mRNA and ncRNA genes decreased in particular in anucleate platelets, where the number of circRNA expressing genes increased by a 2.1-fold compared to MKs (Figure S2A). Furthermore, with 47,654 low confidence circRNAs detected, we also detected the highest diversity of circRNA transcripts in platelets (Figure 3C). The estimated exon usage with a mean of 4.37 exon/circRNA (median= 3) and the putative spliced length with a mean of 645.9 nt/circRNA (median= 461nt), was lower than that of circRNAs in erythropoiesis (Figure S2B, C). Yet, a preference for the 2nd exon usage for circularization was apparent, with no overt preference of the last exon (Figure S2D, E).

To determine the differential circRNA expression in MKs, we focused on the 748 high-confidence circRNAs (Figure S2F). Pearson's sample-correlation coefficient revealed a strong kinship between CD42- and CD42+ MK (Figure 3D). This was also reflected by differential expression analysis of circRNA expression, which identified only 5 out of 748 (0.66%) differentially expressed circRNAs, i.e. FIRRE, MED12L, LTBP1, EFCAB13, and of the key transcription factor for MK differentiation FLI1 (p-adjusted<0.05 and log2 fold change >0.5; Figure 3E; Table S2). Platelets, however, showed an overall distinct expression pattern of circRNAs compared to the two megakaryocytic populations (Figure 3D, F; Table S2). Interestingly, even though the overall expression of circRNAs was similar between CD42- immature and CD42+ mature MK, the number of circRNA with CLR>1 and the overall CLR increased by ~1.8-fold during MK differentiation (Figure 3G, H; Table S2). Thus, circRNA are

expressed in differentiating megakaryocytes, which substantially differ from circRNA expression in platelets.

CircRNA expression can be independent of mRNA expression

We next studied how changes in circRNA expression related to changes in expression of the linear mRNA counterpart. To this end, we calculated the log2 fold change (LFC) in circRNA and mRNA expression in erythroid cells. We isolated five groups based on circRNA and mRNA expression changes (LFC>0.5): *group 1*: decreased in circRNA levels and decreased in mRNA levels; *group 2*: increased circRNA levels and decreased mRNA levels; *group 3*: increased circRNA levels and unchanged mRNA levels; *group 4*: increased circRNA levels and increased mRNA levels; *group 5*: unchanged circRNA levels; *group 6*: decreased circRNA and increased mRNA levels. When we compared the circRNA expression with the mRNA expression before enucleation, i.e. day 0 with day 5, we found that of 36.5% circRNAs, the changes (up or down) coincided with that of the mRNA counterpart (Figure 4A; group 1 (down) and 4 (up)). Of the 34% circRNAs that increased their expression (group 2 and 3), 17.7% of the mRNA counterpart decreased its expression (Figure 4A; group 2), and 16.3% remained unchanged (group 3). We also detected circRNAs whose expression levels remained unchanged throughout the first 5 days of differentiation (group 5; 27%).

We then asked how the circRNA and mRNA expression relates upon enucleation. We compared the LFC of circRNA and mRNA expression of day 0 with that of day 9 of erythrocyte differentiation (Figure 4B). The overall high confidence circRNA expression increased at day 9 compared to day 0 by ~20 times from 32 to 651 identified circRNAs (Figure 4B). Owing to this increase, only 1.4% of circRNA did not alter, and 2.2% decreased their expression levels (Figure 4B; group 5 and 1, respectively). 42.6% of circRNAs showed an increase or decrease of expression in line with their mRNA counterpart (Figure 4B; group 1 and 4). More than half (56.1%) of the circRNAs increased their expression levels, whereas the mRNA expression of the linear counterpart decreased or remained unchanged (Figure 4B; group 2; 38.6%, and group 3; 17.5%, respectively). When comparing day 5 or day 9 to day 0 of differentiation, we detected only 13 and 0 circRNAs, respectively, with decreased circRNA expression levels for which the mRNA levels increased (Figure 4A-B; group

6). Thus, the majority of circRNA do not follow the expression pattern of their linear counterpart in differentiating RBCs prior and after enucleation, indicating that circRNA expression in differentiating erythroid cells can be independent of mRNA expression.

We also compared the changes in circRNA and mRNA expression between CD42- and CD42+ megakaryocytes. The magnitude of the overall expression changes in MK was lower compared to erythroblasts (~6.8 LFC mRNA and ~5.5 LFC circRNA in MK, compared to ~16.7 LFC mRNA and ~13.9 LFC circRNA in erythroblasts from day 0 to 9; Figure 4A-C). 27.1% of circRNA did not alter in expression between CD42- and CD42+ MK (Figure 4C, group 5), and 15.5% of circRNA followed mRNA expression changes (groups 1 and 4). Conversely, in more than half of the circRNAs (57.4%), circRNA levels increased but the mRNA decreased (Figure 4C; group 2; 10.6%) or remained unchanged (Figure 4C; group 3; 46.8%). We found no circRNA with decreased expression levels and increased mRNA levels (Figure 4C, group 6). In conclusion, the expression pattern of circRNA expression only partially follows the mRNA expression, and more than half of the circRNAs change their expression pattern independently of mRNA expression during erythroid and megakaryoid differentiation.

The majority of circRNAs expressed in differentiated cells correlates with that of mRNA expression in progenitors

Several studies reported that circRNAs accumulate in neuronal and muscle cells upon differentiation (Kristensen et al., 2018; Legnini et al., 2017; Westholm et al., 2014). To test whether this accumulation is also seen during blood cell differentiation, we first compared the number of circRNAs identified in *in vitro* differentiated erythroblasts (Figure 2D) with that of our previously published data on mature blood-derived red blood cells (RBC; (Nicolet et al., 2018)). The overlap of circRNA expression of erythroblasts with mature RBCs was only 11.5% (676 out of 5,878 high confidence circRNAs; Figure 4D, left panel). The overlap between circRNA expression in platelets (Nicolet et al., 2018) and MK (Figure S2F) was even lower with 5.6% (602 out of 10,729 high confidence circRNA; Figure 4D, right panel). This limited overlap could stem from the described uptake of transcripts by platelets from exogenous sources (Best et al., 2015), and could possibly incude circRNAs.

Alternatively, the differential circRNA expression could derive from transcription detected in progenitor cells. To address the latter possibility, we compared the mRNAs expressed by differentiating erythrocytes and megakaryocytes with that of circRNAs expressed RBCs and platelets, respectively. To our surprise, the mRNA-expressing genes of in vitro differentiating cells and circRNA-expressing genes of mature cells almost completely overlapped (95.83% and 95.65% for RBC and platelets, respectively, Figure 4E). Thus, the circRNAs detected in enucleated platelets or RBCs correlate better with the mRNA expression than with that of the circRNA expression in their respective progenitors.

CircRNA expression does not correlate with changes in translation efficiency of the linear RNA counterpart

It was recently described that the YAP circRNA can prevent the translation of its mRNA counterpart in a sequence-specific manner (Wu et al., 2019). circRNA overexpression showed that the translation initiation complex was prevented from engaging in active translation of the mRNA (Wu et al., 2019). This finding prompted us to determine the transcriptome-wide correlation of endogenous circRNA expression with the translation efficiency of their mRNA counterpart. We used normalized expression data from the matched RNA-seq and ribo-seq analysis of CD42- and CD42+ MKs. The ribosome footprint (RFP) reads fulfilled all quality controls upon mapping the reads onto the genome, i.e. the read size distribution, periodicity and reads distribution after P-site correction (Figure S3). We determined whether mRNA expression related to changes in RFP abundance by computing the LFC in mRNA abundance and the LFC in RFP abundance for each gene (Figure 5A, Table S3). Most changes in mRNA expression correlated well with changes in RFP counts (Pearson's correlation coefficient: 0.854). In addition, genes that expressed circRNAs (red dot) or not (gray dots) showed very similar patterns (Pearson's correlation coefficient: 0.768). This finding suggests that the ribosomal occupancy of mRNAs was not influenced by the co-expression of circRNA variants. Similarly, when we calculated the translation efficiency per gene *(normalized RFP counts (TPM) / normalized mRNA-seq (TPM))* (Figure 5B), no differences in mRNA translation were observed between genes that do (red dots) or do not (grey dots) express circRNAs.

We next determined whether changes in translation efficiency of mRNAs upon MK maturation related to the expression of circRNA from the same gene. Overall, we observed lower changes in translation efficiency in circRNA-expressing genes than circRNA-less mRNA-genes (span of LFC in translation efficiency = 3.8 and 29.55, respectively; Figure 5B). Only 20 circRNAs (2.67%) showed an LFC>1 in translation efficiency of the mRNA counterpart in CD42+ compared to CD42- MKs (Figure 5C; blue dots). We then examined the mRNA, RFP and circRNA levels in more detail of 2 of these circRNA-expressing genes that showed changes in translation efficiency, i.e. *OMA1*, *DIAPH1* (Figure 5D, E). *OMA1* circRNA expression is increased in CD42+ MKs compared to CD42- MKs, whereas mRNA expression is equal and RFP counts drop by 2.46-fold (Figure 5D). Similarly, DIAPH1 circRNA expression increases for both detected circRNAs upon maturation, whereas mRNA expression is maintained (Figure 5E). However, *DIAPH1* RFP counts moderately increased by 2.18-fold in CD42+ compared to CD42- MK (Figure 5E). Combined, these findings indicate a limited influence of circRNAs on the translation efficiency of their mRNA counterpart during megakaryocyte maturation.

CircRNAs contain putative open reading frames

Most (~80%) circRNAs are reported to fall in the coding region and/or contain the canonical translation start site (Memczak et al., 2013; Stagsted et al., 2019). To define the coding potential of megakaryocytic circRNAs, we used ORF-finder (NCBI) to predict open reading frames (ORFs) in high confidence circRNAs. Because algorithms such as ORF-finder cannot mimic back-splicing in silico, we juxtaposed the circRNA sequence three times (termed 3xCircRNA; Figure S4A). This 3xCircRNA sequence allows the investigation of ORFs spanning over the back-splicing junction that could gain stop codon as a result of frameshifts at the circRNA back-splicing junction. To identify ORFs that are common to both mRNAs and circRNAs, we also used the high confidence circRNA sequence opened-up at the back-spliced junction (1xLinRNA). Open reading frames with a minimal length of 25aa and a canonical AUG start codon in the 1xLinRNA and the 3xCircRNA sequences were included in the analysis.

The majority of 1xLinRNA sequences (729 out of the 748) in MKs contained putative ORFs (n= 2752 unique ORFs; mean ORF length: 75.1 aa; median: 45aa). Similarly, 744 out of 748 3xCircRNA contained at least one predicted ORF (n= 3,948 unique

ORFs; mean ORF length: 112.2aa; median: 54aa; Figure 6A). To isolate circRNA-specific ORFs, we subtracted the ORFs found in 1xLinRNA from those found in 3xCircRNA sequences. Whereas 1731 unique ORFs were shared by both sequences (mean ORF length: 75.4aa; median: 45aa), 1491 unique ORFs were specific for 510 circRNAs (mean ORF length: 188.7aa; median: 111aa; Figure 6B). CircRNA-specific ORFs were longer than those shared by 1xLinRNA and 3xCircRNA (p-adjusted= 4.024e-13; Figure S2G). Thus, 68.2% of circRNAs in MKs have specific putative open reading frames. CircRNA-specific putative ORFs were also detected in platelets, with 19,995 ORFs (6,904 out of 10,729 circRNAs, 64.3% of circRNA), and in differentiating erythroblasts, with 1,864 ORFs (639 out of 950 circRNAs, 67.3% of circRNA; Figure S4B-C).

Some circRNAs display a high ribosome density

To determine whether the circRNA containing putative ORFs showed any evidence of translation into proteins, we re-analyzed previously published RFP sequencing data from erythroblasts (Mills et al., 2016). We determined the ribosome density on the coding region of mRNA sequence, and on the circRNA sequence (Figure 6C). We mapped the RFP read on the whole mRNA coding sequence and on the 'high confidence' circRNA sequence that was opened-up at the back-spliced junction. We then calculated the RFP density per kilobase on the mRNA and the circRNA sequences. To measure differences in RPF reads between circRNA and mRNA, we calculated the ratio of circRNA-over-mRNA RFP reads density (ribo-CLR). Of note, a ribo-CLR > 1 would be indicative of a higher RFP density over the circRNA sequence than on the mRNA sequence. Erythroblasts contained only 57 circRNA-mRNA ribo-seq pairs that had reliable expression levels (>50RFP per kilobase; Figure 6D, left panel; Table S3). Out of these 57 circRNAs, 32 (56% of pairs) had a ribo-CLR >2. It is thus conceivable that these 32 circRNA are translated in erythroblasts.

In MK ribo-seq data, the vast majority (95.9%) of the circRNA-expressing genes had a ribo-CLR below 2, which indicates similar mRNA and circRNA RFP read densities (Figure 6D, middle panel; Table S3). Indeed, of the 748 circRNAs, only 31 (4.14%) showed a higher RFP density than their corresponding mRNA (Table S3). Lastly, we re-analyzed published ribo-seq data on platelets (Mills et al., 2016). 465 circRNAs

showed a ribo-CLR>2 (Figure 6D, right panel; Table S3). Nevertheless, with 4.34% of all 10,729 circRNAs, the overall percentage of circRNAs that could potentially contribute to the platelet proteome resembled that of MK. Thus, a small proportion of circRNAs in erythrocytes, megakaryocytes and platelets showed a higher ribosome density than their mRNA counterpart.

CircRNAs in platelets show hundreds of ribosome footprints on the back-spliced junction

Ribosome occupancy should also be found at the back-spliced junction of the circRNA, if this region contributes to the putative circRNA-specific coding region. We therefore re-screened all RFP reads from erythrocytes, megakaryocytes, and platelets and with an adaptation of the ribo-seq reads alignment to allow for chimeric reads detection and subsequent back-spliced junction detection. Overlap of the chimeric reads to the genome was set with a minimum anchor point of 8nt to 16nt on either side of the read. We also allowed 1 mismatch. Of the 114 million reads in erythroblasts, we found only 2 RFP reads for anchors of 8, 9, or 10nt (Figure 6E, left panel; Table S3). The 2 RFPs corresponded to 2 circRNAs *circ-ARHGEF12* and *circ-SPECC1*. As the median ribo-seq read density in erythroblasts reached 1.984 $x10^{-3}$ reads per nt, one would already expect by chance about 53 reads on the 950 'high confidence' circRNA (0.055 read/circRNA junction).

We also screened the 757 million ribo-seq reads in megakaryocytes for ribosome foot prints at the back-spliced junction. As the median ribo-seq read density for all translated genes in megakaryocytes was 0.23 reads per nt, one would expect by chance about 4972 reads on the 748 'high confidence' circRNA if they were translated (6.64 reads/circRNA junction). However, only 3 RFPs reads matching circRNAs in megakaryocytes were detected (Figure 6E, middle panel; Table S3). These 3 RFPs reads corresponded to 2 circRNAs; i.e. *circ-PRDM2 and circ-YEATS2*.

Of the ~178million RFP reads in platelets, one would expect by chance about 3378 RFPs on the 10,729 'high confidence' circRNA of platelets (median RFP density on mRNA 0.011read/nucleotide, 0.314 read/circRNA junction). Yet, we only detected 180 RFPs, 168 RFPs and 182 RFPs, over the circRNA back-splice junction in platelets with an anchor size of 8, 9, and 10nt, respectively (Figure 6E, right panel; Table S3). The circRNA-RFPs corresponded to 56 different circRNAs. Of note, 13 of

these 56 circRNAs (HIPK3, TMEM135, CORO1C, DNAJC6, GSAP, ASH2L, FAM120A, MCU, FARSA, WDR78, ZC3H6, NCOA2, APOOL) also displayed a ribo-CLR>2 (Figure S4D), indicating ribosome reads on both the back-splice junction and the full circRNA sequence. Overall, the RFP analysis showed some but limited evidence of translation, in particular in MK, and erythroblasts. Platelets provide the best indication of putative circRNA translation, with - based on the sequencing depth of the riboseq data sets- ~58 times more circRNA-specific ribo-seq reads than erythroblast and ~258 times more than megakaryocytes.

Mass spectrometry fails to identify circRNA-specific peptides in platelets

Because the RFP analysis in platelets showed the highest probability of translation, we further searched for evidence of translation in this cell type. We generated a reference peptide library from the 3 translation frames over the 508 circRNA back-spliced junction of the high-confidence circRNAs with ribo-CLR above 2 or detected circ-RFP. Translation frame(s) that contained a STOP codon before the back-spliced junction were excluded from this reference peptide library. This left 1566 unique putative circRNA peptides spanning the back-spliced junction with a minimum of 5 amino acids. Of these 1566 putative circRNA peptides, 12 (0.77%) did not contain lysine nor arginine, 97 (6.2%) and 123 (7.9%) contained only arginine, or lysine, respectively. Thereby the majority (99.2%) of putative circRNA peptides are cleavable by trypsine treatment and should be detectable by mass spectrometry. To also identify circRNA-specific peptides outside the circRNA-junction, we included the full-length circRNA-specific putative ORFs of platelets in the peptide library we identified in Figure S4B. Lastly, we also included the reference proteome (UniProt) to prevent possible ‘background’ hits to match known canonical protein sequences. Published MS data of platelets (Van Oorschot et al., 2019) were used to search for circRNA-specific peptides. Of the 192,743 identified peptides, not one peptide matched a circRNA junction (Figure S4D). Also no peptides were found from circRNA-specific ORFs that were not explained by the mRNA-derived proteome.

Thus, even though ribo-seq data suggest some possible translation from circRNAs in platelets, the detection of peptides from back-spliced circRNA junctions, or from circRNA-specific ORFs fall- if present at all- below the detection limit of data-dependent Mass spectrometry analysis.

DISCUSSION:

In this study, we present a comprehensive analysis of circRNAs in terminal differentiation of erythroblasts and platelets. CircRNA expression is in particular increased at enucleation in erythroid differentiation. This could reflect an accumulation of transcript degradation left-overs upon enucleation, as previously suggested (Alhasan et al., 2016). While this is an attractive hypothesis, not all circRNA follow this overall trend. In fact, more than 50% of circRNAs were regulated independently from changes in linear mRNA expression. This was true when we compared differences in expression levels prior to and post enucleation. Furthermore, a cluster of circRNAs is expressed by erythroid progenitors, but downregulated when differentiation progresses. We also observed a clear increase in number and expression levels of specific circRNAs at enucleation, which were not detected in progenitors. This increase of circRNA number and amount is unlikely to stem from a technical reason, as all samples had sufficient sequencing depth (~30 million reads) to allow for a reliable detection of lowly expressed circRNA. Thus, the majority of measured circRNAs in terminally differentiated erythrocytes cannot be explained by mere accumulation from progenitors alone. This was also true for circRNA expression in MK compared to platelets.

Intriguingly, the circRNAs in differentiated RBCs and platelets strongly overlapped with the mRNAs that were expressed in their progenitors. It has been proposed that splicing occurs in the cytoplasm of platelets and MK (Denis et al., 2005). It is therefore tempting to speculate that back-splicing and formation of circRNA could also occur in the cytoplasm. In line with this, several splicing factors were detected in platelets by MS analysis (Burkhart et al., 2012; Van Oorschot et al., 2019)), including DHX9. Therefore, at the period of enucleation of erythrocytes and during platelet formation, mRNA could be circularized, and thus contribute to increased numbers of circRNAs during enucleation and therefore explain the substantial overlap with the mRNAs in progenitors. This hypothesis, however, requires experimental confirmation.

About 4% of circRNAs in platelets and erythrocytes did not overlap with the mRNA levels in MK and differentiating erythroblasts. These circRNAs could be acquired from other cell types in the circulation, as observed for linear RNA in the context of

tumor ‘educated’ platelets (Best et al., 2015), or endogenous vascular-derived RNA (Clancy et al., 2017). A study in platelets from healthy and diseased individuals could shed light on this. If true, circRNAs taken up by platelets or erythrocytes could in fact serve as novel biomarkers, which could be in particular valuable for diagnostics due to their high stability.

The function of circRNAs still remains elusive. CircRNA could serve to reinforce transcription, and lead to differential exon usage as previously proposed (Kelly et al., 2015). Differential exon usage was also recently reported to be supported by pre-mRNA exonic sequences (Fiszbein et al., 2019). Whether this is indeed occurring in primary blood cells is yet to be determined.

CircRNAs are also reported to regulate mRNA translation in a sequence-specific manner (Li et al., 2020; Sun et al., 2019; Wu et al., 2019). We investigated here the possible effect of circRNA on its mRNA counterpart. We only identified 20 circRNAs in MKs that show limited indications for translation regulation. Whether our findings in MKs are also applicable to enucleated erythrocytes or platelets is not known. However, it is also conceivable that circRNA affect translation of other mRNA (Li et al., 2020; Sun et al., 2019). CircRNA-mediated regulation of mRNA translation is an attractive hypothesis in erythrocytes and platelets, because the majority of translation is shut down in mature cells (Mills et al., 2016), and circRNAs could support translation shutdown

CircRNAs were also shown to be translated (Legnini et al., 2017; Liang et al., 2019; Pamudurti et al., 2017; Yang et al., 2017), in particular from exogenously expressed circRNAs. This finding suggests that the highly stable circRNA transcripts could help maintain the proteome in long-lived red blood cells and platelets. However, we found very limited indications that translation from endogenous circRNAs occurs. Similar to a recent study (Stagsted et al., 2019), we detected only few circRNA-specific RFP. The low number of detected circRNA-specific RFP reads could not be explained by low sequencing depth, as the ribo-seq datasets entailed 114, 757, 178 million reads in erythroblasts, MKs, and platelets, respectively. Yet, only platelets had 180 RFP reads on the circRNA back-spliced junction, which points to some translation occurring. MS analysis however, did not provide supporting evidence of translation at

circRNA back-spliced junctions. Yet, circRNA-derived proteins may be generated on low levels and result in production that is below the detection limit of MS. Immunoprecipitation enrichment of specific targets could help to overcome this detection limit for specific circRNA (Liang et al., 2019; Pamudurti et al., 2017).
mRNA translation also may mask the circRNA translation and its contribution to the proteome, when the circRNA-derived peptides are identical to the mRNA-derived peptides. Nevertheless, because only 117 out of the 465 circRNAs (25.16%) with high ribo-CLR show protein expression in MS, the relative contribution of circRNA-derived translation is arguably limited in light of the mRNA-derived translatome. Compiled, our study thus only supports a limited possible contribution of circRNA-translation to the proteome in platelets.

In conclusion, we here provide the landscape of circRNA expression during in erythroid differentiation and megakaryocyte maturation. This study can serve as blueprint for future studies in these cell types. Our results also challenge the current views on the ubiquitous functions proposed for circRNAs in the biological context of terminal hematopoiesis.

MATERIALS AND METHODS

CircRNA identification and analysis

Data on erythroblast differentiation were retrieved from our previous study (Heshusius et al., 2019). circRNA identification was performed with previously established circRNA detection pipeline (Nicolet et al., 2018). Briefly, the quality of RNA-seq reads was assessed using FastQC version 0.11.8 (Simon Andrews, Babraham institute). Data were aligned to the human genome (GRCh37 / hg19-release75) using STAR version 2.5.2b (Dobin and Gingeras, 2015) allowing for chimeric detection, with an overhang 'anchor' of 15 nt on either side of the read. The chimeric output file of STAR was analyzed with DCC 0.4.6 (Cheng et al., 2015) and CircExplorer2 (CE2) 2.3.2 (Zhang et al., 2014) to detect, filter, and annotate circRNA. DCC was used for detection of linear reads at the circRNA coordinates (using the option -G). CircRNA expression was considered low confidence when at least 2 junction reads were found in at least 1 sample by both DCC and CE, and high confidence detection when at least 2 junction reads were found in all biological replicates of one specific cell type, by both tools. Supporting circRNA read counts were normalized to reads per million mapped reads (RPM). The maximal circRNA spliced length was calculated with the exon length information provided by the CE annotations. These annotations were then used to calculate the first and last circularized exon. For differential expression analysis of high confidence circRNA, we used DESeq2 (Love et al., 2014) with p-adjusted <0.05. The 'minmu' option was set to 0.1. The log2 Fold Change (LFC) was calculated using the formula: $LFC(AB) = log2(B) - log2(A)$. Data analysis was performed in R (3.5.1) and R-Studio (1.1.453). All scripts are available upon reasonable request.

RNA-seq analysis

After quality control with FastQC, raw RNA-seq reads were aligned with Salmon version 0.13.1 (Patro et al., 2017) on the coding and non-coding transcriptome (ENSEMBL, GRCh38 release 92) to obtain Transcripts Per kilobase per Million (TPM) normalized counts. To get mRNA and ncRNA expression per gene, TPM counts were summed up per gene name, according to ENSEMBL BioMart annotations (Kinsella et al., 2011).

In-vitro megakaryocyte differentiation

Megakaryocytes were cultured from human cord blood-derived CD34+ cells. Cord blood was diluted in PBS, loaded on ficoll-pague (GE Healthcare) and centrifuged 15 minutes at 400g. The white blood cell layer was collected from the ficoll gradient, washed twice in PBS+ 1mM EDTA + 0.2% HSA. CD34+ cells were purified with anti-CD34 Microbeads (Miltenyi biotec 130-046-702) and MACS LS column (Miltenyi, 130-042-401) according to manufacturer's guidelines. Purity for CD34+ cells was >90%, as determined by flow-cytometry with anti-CD34 antibody (#581-PE; Beckman Coulter, measured on Beckman Coulter FC500). CD34+ cells were seeded at a density of 50.000 cells/well in a 12 well plate in CellgroSGM (CellGenix) supplemented with 100ng/mL human recombinant TPO (CellGenix, Cat.1417) and 10ng/mL IL1β (Miltenyi Biotec, Cat.130-093-895). At day 3, 1:1 Fresh media was added with an additional 10ng/mL TPO and 1ng/mL IL1β. On day 5, cells were split at a density of 100.000 cells in 6 wells plates, and fresh media with 10ng/mL TPO and 1ng/mL IL1β was added. On day 10, cells were harvested and supplemented with 100µg/mL Cycloheximide (Sigma, Cat. 1810-1G) in culture medium prior to cell sorting.

Megakaryocyte selection

Cultures cells were harvested when >85% of cells were CD41a+ (CD41a-APC; clone HIP8; BD Pharmingen; Cat.559777), as determined by flow-cytometry. Cells were pelleted at 120g for 1 minute at 4ºC and washed once in ice cold buffer (PBS, 1mM EDTA, 0.2% HSA + 100µg/mL cycloheximide). Two-step selection was performed on ice with EasySep® Human PE-Positive Selection Kit (18551), using anti-human CD42b-PE (clone HIP1, BD Pharmingen, 555473) according to manufacturer's guidelines. CD41a+ CD42b- and CD41a+ CD42b+ cell fractions were selected, washed using PBS + 100µg/mL cycloheximide, and kept on ice until further processing. Viability of the cells was >90% (7-AAD, Biolegend), and purity for CD41a+ CD42b- was >85% and CD41a+ CD42b+ was >95%, as determined by flow-cytometry. Cells were immediately used for processing.

RNA-sequencing

Cell lysis and RNA isolation for RNA sequencing and ribosomal footprinting was performed according to Ignolia et al. (Ingolia et al., 2009) with some adaptations.

Briefly, CD42- and CD42+ selected megakaryocytes were pelleted. Cell pellets were dissolved in 150µL ice cold lysis buffer (20mM Tris pH 7.4, 250mM NaCl2, 5mM MgCl2, 0.5% Triton X-100, 1mM Dithiotreitol (DTT), 0.024U/µL Turbo DNAse (Ambion), 0.01U/µL RNAseOUT (Invitrogen) and 100µg/mL Cycloheximide). The lysate was immediately split in two aliquots of 50 and 100 µL, for total RNA-sequencing and ribosomal footprint analysis, respectively. 2.5µL SuperaseIN RNAse inhibitor (Invitrogen) was immediately added to the total RNA sample, and total RNA was extracted using the miRNAeasy kit (Qiagen) according to manufacturer's protocol. RNA quality was assessed on Agilent bioanalyzer RNA chip (all samples with RIN >9.0), and quantified using Qubit 2.0 (Thermofisher). Ribosomal RNA was removed using Ribo Zero Gold (Epicentre/Illumina; 20020596) using the manufacturer's protocol, and the cDNA library was constructed with the KAPA stranded RNA-seq kit (KAPA biosystems KK8400) according to manufacturer's protocol with Illumina Truseq forked adapters. Library quality was assessed on Agilent Bioanalyzer High sensitivity DNA chip and quantified using qPCR with the KAPA library quantification kit (KK4824). Libraries were pooled to a 2nM concentration and 125bp paired-end reads were sequenced on HiSeq 2000 (Illumina).

Ribosomal footprinting

100 µL of cell-lysates was used for ribosome foot-printing preparation. Samples were kept on ice for 10 minutes and regularly vortexed to fully lyse the outer membranes and release the ribosomes. Lysates were spun at 20.000g for 10min at 4ºC and supernatant was harvested to remove the remaining pelleted cell debris and nuclei. Lysates were treated with 2µL of RNAse 1 (Ambion) and incubated at 37ºC for 45 minutes. The reaction was stopped by the addition of 5µL SUPERAseIN RNAse inhibitor. Lysates were loaded on a 1M 0.150mL Sucrose cushion in 8x34 mm polycarbonate thick wall tubes and centrifuged at 120.000g for 30 minutes in a TLA-120.1 rotor in a pre-cooled 4ºC table top ultracentrifuge. The pellet was resuspended using 100µL resuspension buffer (10mM Tris pH 7.0, 1% SDS and 0.01U/µL proteinase K (NEB)) for 30 minutes at 37ºC. Footprints were subsequently extracted with miRNAeasy kit according to manufacturer's guidelines for small RNA. rRNA contamination was removed using the Ribo-Zero gold rRNA removal kit

(epicenter/Illumina; 20020596). Footprint RNA quality was assessed using the Agilent Bioanalyzer using the RNA chip. Footprints were selected for 26-34nt with custom markers (see (Ingolia et al., 2009)) on Elchrom Scientific ORIGINS system and short fragment gels, using gel electro-elution into Spectra/Por 3 3.5KD MWCO dialysis membranes. Libraries were produced using the Art-Seq kit (Epicentre/Illumina) according to the manufacturer's protocol. The quality of libraries was assessed on Agilent Bioanalyzer with high sensitivity DNA chip and quantified using qPCR with the KAPA library quantification kit (KK4824). Libraries were pooled to a 2nM concentration and sequenced on HiSeq 2000 (Illumina).

Ribo-seq analysis

The quality of raw ribo-seq reads was assessed using FastQC. Reads were trimmed for sequencing adapters using Trimmomatic version 0.39 (Bolger et al., 2014), excluding reads <25 and >35nt long. Trimmed reads were then mapped with Salmon on the human transcriptome (GRCh38 release 92) to get TPM normalized counts. Using a k-mer of 11 performed best for ribo-seq mapping to build the Salmon index. The BAM files resulting from genome (GRCh37 release 75) mapping with STAR were further processed in Ribo-seQC (Calviello et al., 2019) to assess the data quality. Translation efficiency was obtained by dividing the normalized ribo-seq counts in TPM by the normalized RNA-seq counts in TPM. To estimate the ribosome footprint read density, we used the linearized sequence of high confidence circRNA and the coding transcriptome. After aligning the ribo-seq reads with Salmon, ribo-seq reads density per kilobase were obtained by dividing the estimated counts (not TPM) by circRNA length (for circRNA) and CDS length (for mRNA). The circular-over-linear ribo-seq reads density per kilobase (ribo-CLR) was obtained by dividing the circRNA ribo-seq density per kilobase, by the ribo-seq density per kilobase of mRNA. The ratio was calculated separately for each sample and averaged per population. Ribo-seq data for erythroblast and platelets were retrieved from a previous study (Mills et al., 2016).

CircRNA-specific detection from ribo-seq reads

To identify circRNA-specific ribo-seq reads, we used an approach similar to the 'classical' circRNA detection described above. Genome alignment on GRCh37

(release 75) with STAR was used for chimeric reads detection with incremental anchors of 8 to 16 nt on either side of a chimeric read, allowing for 1 mismatch. CE2 was then used to detect, quantify and annotate the circRNAs, and data were further processed in R.

ORF prediction

The linearized, spliced 'high confidence' circRNA sequences were obtained using BEDtools version 2.18 (Quinlan and Hall, 2010). The sequence was juxtaposed 3 time to mimic a circRNA sequence. Open reading frames (ORFs) were predicted for both linearized circRNA sequence (1xLinRNA) and tripled (3xCircRNA) sequences using ORF-finder (version 0.4.3; NCBI). Resulting ORF sequences were analyzed and filtered in R.

Detection of circRNA peptides by mass spectrometry

The high confidence circRNA sequences including all exons between first and last exon were obtained using BEDtools. The last 77nt of the circRNA sequences were joined to the first 77nt to generate a junction sequence library. This library was translated in 3 frames. Open reading frames (ORFs) that presented a length <30 amino acids (that is, not spanning the circRNA junction) were removed. All ORFs were transformed into a fasta library. A proteome reference library containing circRNA junction ORFs, as well as circRNA-specific ORFs (from 3xCircRNA sequences of platelets, filtered for unique circ-specific ORFs), and uniport Human proteome reference (downloaded 13-02-2019), was constructed and used for peptide matching with Proteome Discoverer (version 2.2; Thermo-Fisher) of our previously published mass spectrometry dataset on platelets from healthy donor (Van Oorschot et al., 2019). The precursor mass tolerance was set to 10ppm with a fragment mass tolerance of 0.6Da and a target FDR of 0.01. Only peptides with high confidence detection in MaxQuant were considered.

Plots and graphs

Heatmaps were generated in R using corrplot (version 0.84; T. Wei and V. Simko, 2017, https://github.com/taiyun/corrplot) or pheatmap 1.0.8 (Kolde, 2012). Plots and

graphs were generated with ggplot2 (Wickham, 2009) in R, or Graphpad PRISM version 7.0.

Data retrieval and deposition

For *ex vivo* RBC and platelet analysis, we used our previously published data sets. Data were retrieved from National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and Sequence Repository Archive (SRA) accession numbers for platelets (project: PRJEB4522): ERR335311, ERR335312 and ERR335313 (Kissopoulou et al., 2013); for RBCs: (GEO: GSE63703) SRR2124299, SRR2124300, SRR2124301 and (GEO: GSE69192) SRR2038798 (Alhasan et al., 2016; Doss et al., 2015) ; for erythroid differentiation: accession number GEO: GSE124363 (Heshusius et al., 2019). Ribosome footprinting datasets of platelets and erythroblasts were retrieved from GEO with accession number: GSE85864 (Mills et al., 2016). Raw mass spectrometry datasets of platelets were obtained from the PRoteomics IDEntifications Database (PRIDE) repository under the accession number PXD009020 (Van Oorschot et al., 2019).

DATA AVAILABILITY

Ribosome footprinting and RNA-seq dataset of megakaryocytes were deposited on NCBI's Gene Expression Omnibus (GEO) with accession number GSE159579.

FINANCIAL SUPPORT

This research was supported by the European Research council (ERC, consolidator grant), by the Dutch Science Foundation (Apasia grant), and by Oncode (all to M.C.W.).

ACKNOWLEDGMENTS

We also thank S. Heshusius for helpful discussions, and M. Hansen, P.J. van Alphen, and M. van den Biggelaar for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS:

B.P.N. analyzed data, B.P.N., S.J, and E.H. performed experiments. M.C.W., W.H.O, E.v.d.A, and M.v.L supervised, B.P.N. and M.C.W wrote the manuscript.

FIGURE LEGENDS:

Figure1: CircRNAs are abundant during *in vitro* erythroid differentiation

(**A**) Erythroblast progenitors were cultured for 12 days under differentiation conditions as in (Heshusius et al., 2019). Samples were taken at indicated days of differentiation, and RNA-seq was performed (n=4 donors). (**B**) CircRNA detection using DCC and CircExplorer2 (see methods) to identify 'low confidence' (circRNA with at least 2 reads in at least 1 sample) in all compiled samples from (A). The intersection of the circles represents the co-detected circRNAs with both tools. (**C**-**D**) Characterization of low confidence circRNA. (**C**) Putative exon usage and (**D**) putative spliced length, using splicing annotations from linear isoforms. (**E-F**) Characterization of (**E**) start and (**F**) end exon usage by 'low confidence' circRNAs, based on the circRNA junction positions and canonical mRNA splicing annotation.

Figure 2: CircRNA expression increases at enucleation of erythroid cells

(**A**) Number of genes encoding mRNA, non-coding RNA (ncRNA) and circRNA, and (**B**) the number of transcripts detected during erythroid differentiation (cut-off: >0.1 TPM for mRNA and ncRNA; low confidence circRNA). (**C**) Percentage of enucleated cells at indicated time point (n=4 donors; adapted from (Heshusius et al., 2019)). (**D**) CircRNA detection using DCC and CircExplorer2 (see methods) from RNA-seq data generated from Figure 1. The intersection of the circle represents the co-detected circRNA referred to as 'high confidence' (circRNA with at least 2 reads in all 4 biological replicates of 1 time-point). (**E**) Pearson's sample-correlation coefficient map of 'high confidence' circRNA expression. (**F**) Heatmap of the differentially expressed 'high confidence' circRNA (n=135; p-adjusted <0.05) corrected for sequencing depth (RPM). RPM: Reads per million mapped (linear) reads. (**G**) Circular over linear ratio (CLR) was calculated for 'high confidence' circRNAs during erythropoiesis, median is indicated in red. Differences were assessed by two-sided t test with Benjamini-Hochberg p-value adjustment, all compared to day 0. (***: padj<0.001, ****: padj<0.0001).

Figure3: CircRNA expression during megakaryocyte differentiation

(**A**) Diagram representing *in vitro* differentiation from CD42- and CD42+ megakaryocytes (n=3 per population) to platelets (Nicolet et al., 2018). (**B**) CircRNA detection in CD42- and CD42+ megakaryocytes from RNA-seq data, using DCC and

CircExplorer2. The intersection of the circle represents the co-detected 'low confidence' circRNAs. Of note, platelet data were not included for circRNA detection in megakaryocytes. (**C**) Number of mRNA, non-coding RNA (ncRNA) and circRNA transcripts detected during megakaryoid differentiation (>0.1 TPM for mRNA and ncRNA; low confidence circRNA). (**D**) Pearson's sample-correlation coefficient map between CD42- and CD42+ megakaryocytes, and platelets. (**E**) Volcano plot showing the differential expression of high confidence circRNA between CD42- (left) and CD42+ megakaryocytes (right; n=5; p-adjusted<0.05 and LFC>0.5). (**F**) Heatmap of high confidence circRNAs detected in megakaryocytes and platelets. (**G-H**) Circular-over-linear ratio (CLR) was calculated for CD42- and CD42+ megakaryocytes and plotted against (**G**) circRNA expression or (**H**) compared between the two populations. Differences were assessed in G by two-sided t test p-value adjustment using the Benjamini-Hochberg procedure; ****: padj<0.0001.

Figure 4: CircRNA expression can be independent of mRNA expression

(**A-C**) Log2 Fold Change (LFC) for circRNA and corresponding mRNA was calculated for (**A**) day 0 to 5 and (**B**) day 0 to 9 of erythroid differentiation, and for (**C**) CD42- to CD42+ megakaryocytes differentiation. CircRNA-mRNA pairs are represented in: decreased circRNA / decreased mRNA (group 1, pink), increased circRNA / decreased mRNA (group 2, red), increased circRNA / equal mRNA (group 3, purple), increased circRNA / increased mRNA (group 4, orange), equal circRNA (group 5, grey), decreased circRNA / increased mRNA (group 6, grey). (**D**) Overlap in circRNA expression between *in vitro* RBC and *ex vivo* RBC (data from (Doss et al., 2015; Heshusius et al., 2019; Nicolet et al., 2018); left panel; Figure 1), and *in vitro* megakaryocytes and *ex vivo* platelets ((Kissopoulou et al., 2013; Nicolet et al., 2018); right panel). (**E**) Overlap between circRNA in differentiated cells and mRNA in progenitor cells, for erythroid (left panel) and megakaryoid cells (right panel).

Figure 5: CircRNA-mediated regulation of mRNA translation is limited in megakaryocytes

(**A**) Log2 Fold Change (LFC) in ribosomal occupancy (ribo-seq) and mRNA expression between CD42- and CD42+ megakaryocytes. Gene products that consist of both mRNA and one or more circRNA are plotted in red. (**B**) LFC in translation efficiency (see methods) and linear mRNA expression between CD42- and CD42+

megakaryocytes. (**C**) LFC of translation efficiency and circRNA expression between CD42- and CD42+ megakaryocytes. Genes are plotted in blue when LFC in translation efficiency was more than two-fold (LFC>1). (**D-E**) CircRNA, mRNA and ribo-seq expression of (**D**) *OMA1* and (**E**) *DIAPH1* (n=6). (A-C) The linear regression line is represented in blue. RPM: reads per million mapped (linear) reads, TPM: transcripts per kilobase per million.

Figure 6: Putative ORF detection in circRNAs, with limited evidence for translation

(**A**) Open reading frames (ORFs) were computed (see methods) for linearized high confidence circRNA sequences (1xLinRNA), and for tripled juxtaposed linearized sequence of high confidence circRNA to include circularized junction areas (3xCircRNA, see Figure S4A). Bar graph depicts the length and frequency of all identified ORFs found in 1xLinRNA (pink) or 3xCircRNA (blue). (**B**) Bar graph depicting the length and frequency of the ORFs found in both 1xLinRNA and 3xCircRNA sequences (purple) or specifically in circRNA (green). (**C-D**) Ribosome footprinting sequencing (Ribo-seq) data were used to determine the ribosome density per kilobase on mRNA sequence and circRNA sequence. The circular-over-linear ribosome density ratio (ribo-CLR) was calculated. (**C**) Schematic representation of circRNA (back-spliced junction in blue) and mRNA sequence, and ribo-seq density resulting in high or low ribo-CLR. (**D**) Calculated ribo-CLR for erythroblasts (left panel), megakaryocytes (middle panel), and platelets (right panel) and plotted against the mRNA ribosome foot print (RFP) density (RFP per kilobase). (**E**) Ribo-seq data of erythroblasts (left panel; data from (Mills et al., 2016)), CD42- or CD42+ megakaryocytes (middle panel), and platelets (middle panel; data from (Mills et al., 2016)) were screened for RFP reads on the circRNA junction (circ-RFP) using CE2. The minimum chimeric alignment anchor size for alignment (8 to 16 nt) is indicated. Circ-RFP that were not found expressed in low confidence circRNA (based on RNA-seq detection) were excluded. (**F**) Venn diagram showing the overlap in platelets of circRNAs with ribo-CLR >2, circRNAs with circ-RFP and circRNA-specific peptide detected by mass spectrometry (no peptide detected; MS; data from (Van Oorschot et al., 2019)).

REFERENCES:

Aktaş, T., Ilik, I.A., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., Manke, T., Backofen, R., and Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. Nature *544*, 115–119.

Alhasan, A.A., Izuogu, O.G., Al-Balool, H.H., Steyn, J.S., Evans, A., Colzani, M., Ghevaert, C., Mountford, J.C., Marenah, L., Elliott, D.J., et al. (2016). Circular RNA enrichment in platelets is a signature of transcriptome degradation. Blood *127*, e1–e11.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). CircRNA Biogenesis competes with Pre-mRNA splicing. Mol. Cell *56*, 55–66.

Best, M.G., Sol, N., Kooi, I., Tannous, J., Westerman, B.A., Rustenburg, F., Schellen, P., Verschueren, H., Post, E., Koster, J., et al. (2015). RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. Cancer Cell *28*, 666–676.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Burkhart, J.M., Vaudel, M., Gambaryan, S., Radau, S., Walter, U., Martens, L., Geiger, J., Sickmann, A., and Zahedi, R.P. (2012). The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. Blood *120*, 73–82.

Calviello, L., Sydow, D., Harnett, D., and Ohler, U. (2019). Ribo-seQC: comprehensive analysis of cytoplasmic and organellar ribosome profiling data. BioRxiv 601468.

Chen, L.L. (2016). The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol. *17*, 205–211.

Cheng, J., Metge, F., and Dieterich, C. (2015). DCC – specific identification and quantification of circular RNAs from sequencing data. Bioinformatics *32*, 1–13.

Clancy, L., Beaulieu, L.M., Tanriverdi, K., and Freedman, J.E. (2017). The role of RNA uptake in platelet heterogeneity. Thromb. Haemost. *117*, 948–961.

Denis, M.M., Tolley, N.D., Bunting, M., Schwertz, H., Jiang, H., Lindemann, S., Yost, C.C., Rubner, F.J., Albertine, K.H., Swoboda, K.J., et al. (2005). Escaping the nuclear confines: Signal-dependent pre-mRNA splicing in anucleate platelets. Cell *122*, 379–391.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. In Current Protocols in Bioinformatics, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. 11.14.1-11.14.19.

Doss, J.F., Corcoran, D.L., Jima, D.D., Telen, M.J., Dave, S.S., and Chi, J.T. (2015). A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. BMC Genomics *16*, 1–16.

Fiszbein, A., Krick, K.S., Begg, B.E., and Burge, C.B. (2019). Exon-Mediated Activation of Transcription Starts. Cell *179*, 1551–1565.e17.

Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. Dev. Cell *36*, 572–587.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. Genome Biol *15*, 409.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hentze, M.W., and Preiss, T. (2013). Circular RNAs: Splicing's enigma variations. EMBO J. *32*, 923–925.

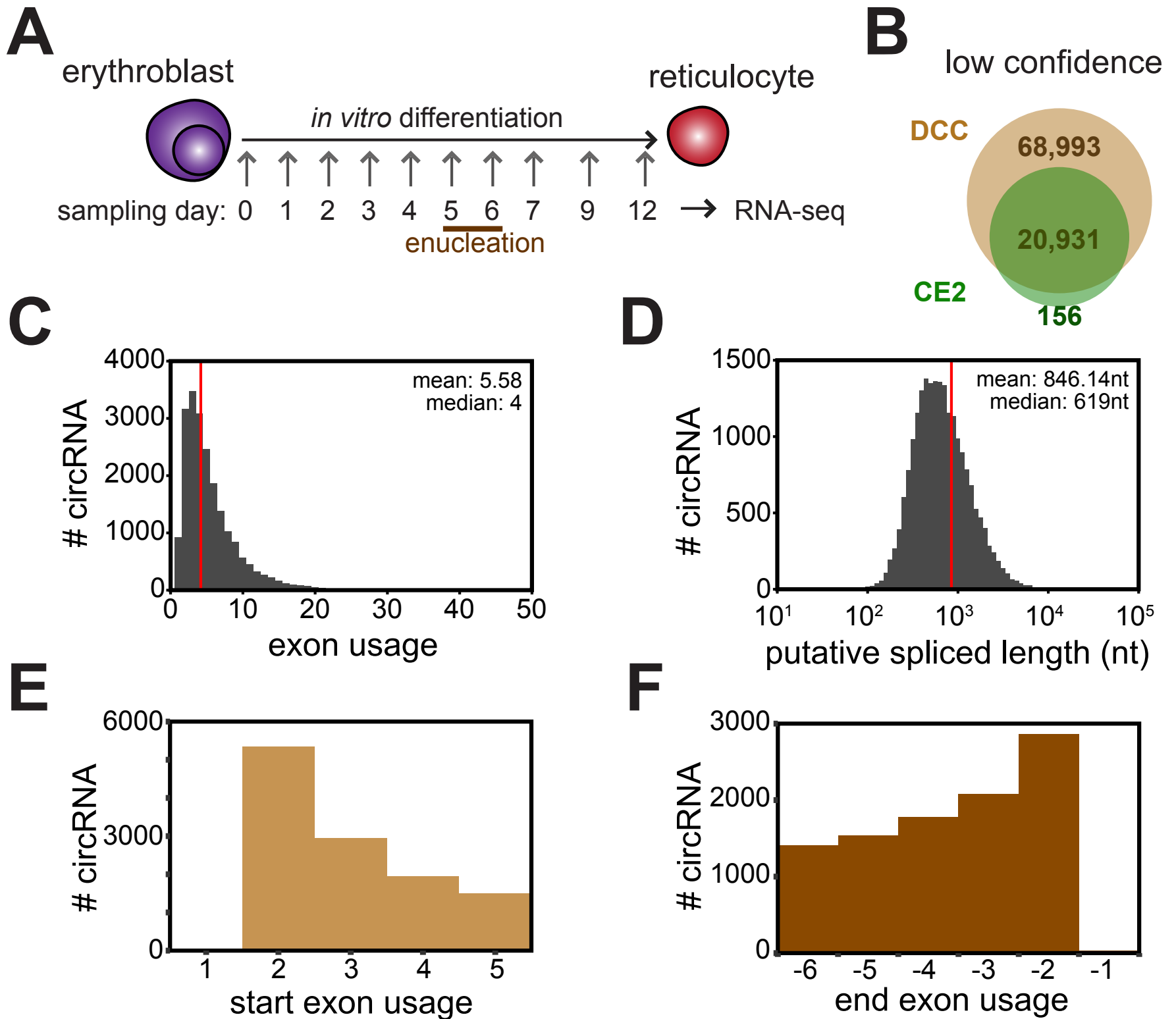
Heshusius, S., Heideveld, E., Burger, P., Thiel-Valkhof, M., Sellink, E., Varga, E., Ovchynnikova, E., Visser, A., Martens, J.H.A., von Lindern, M., et al. (2019). Large-scale in-vitro production of red blood cells from human peripheral blood mononuclear cells. BioRxiv *31*, 3337–3350.

Huang, C., Liang, D., Tatomer, D.C., and Wilusz, J.E. (2018). A length-dependent

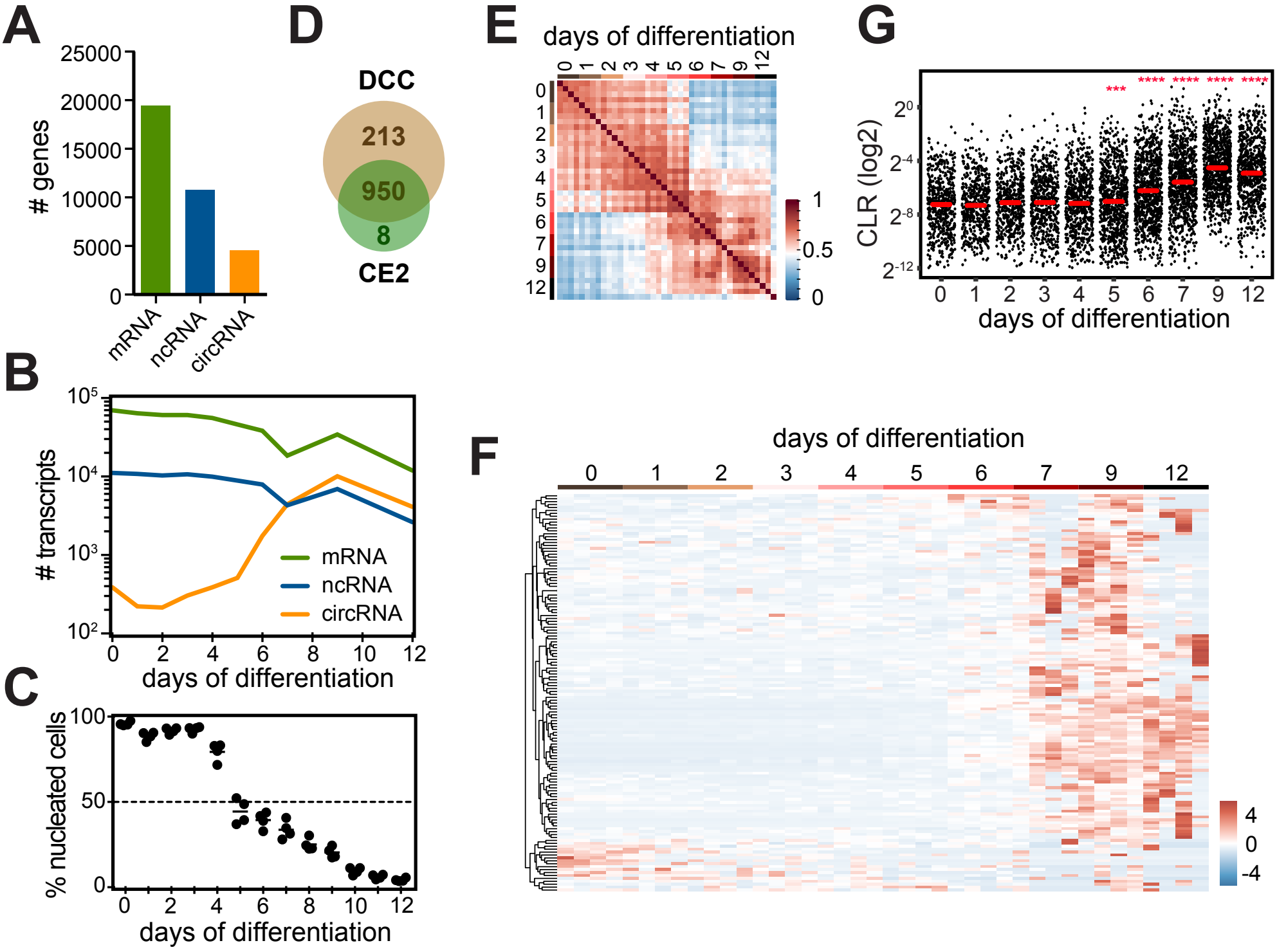
evolutionarily conserved pathway controls nuclear export of circular RNAs. Genes Dev. *32*, 639–644.
Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (80-.). *324*, 218–223.
Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. Rna *19*, 141–157.
Kelly, S., Greenman, C., Cook, P.R., and Papantonis, A. (2015). Exon Skipping Is Correlated with Exon Circularization. J. Mol. Biol. *427*, 2414–2417.
Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database *2011*, 1–9.
Kissopoulou, A., Jonasson, J., Lindahl, T.L., and Osman, A. (2013). Next generation sequencing analysis of human platelet polyA+ mRNAs and rRNA-depleted total RNA. PLoS One *8*.
Kolde, R. (2012). Pheatmap: pretty heatmaps.
Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018). Circular RNAs in cancer: Opportunities and challenges in the field. Oncogene *37*, 555–565.
Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol. Cell *66*, 22–37.e9.
Li, X., Liu, C., Xue, W., Zhang, Y., Jiang, S., Yin, Q., Wei, J., Yao, R., Yang, L., and Chen, L.-L. (2017). Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. Mol. Cell *67*, 214–227.e7.
Li, X., Xiao, L., Chung, H.K., Ma, X.-X., Liu, X., Song, J.-L., Jin, C., Rao, J.N., Gorospe, M., and Wang, J. (2020). Interaction between HuR and circPABPN1 modulates autophagy in the intestinal epithelium by altering ATG16L1 translation. Mol. Cell. Biol.
Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. Nat. Struct. Mol. Biol. *22*, 256–264.
Liang, W.C., Wong, C.W., Liang, P.P., Shi, M., Cao, Y., Rao, S.T., Tsui, S.K.W., Waye, M.M.Y., Zhang, Q., Fu, W.M., et al. (2019). Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. Genome Biol. *20*, 1–12.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
Luo, M., Jeong, M., Sun, D., Park, H.J., Rodriguez, B.A.T., Xia, Z., Yang, L., Zhang, X., Sheng, K., Darlington, G.J., et al. (2015). Long non-coding RNAs control hematopoietic stem cell function. Cell Stem Cell *16*, 426–438.
Maass, P.G., Gla, P., Memczak, S., Dittmar, G., Hollfinger, I., Schreyer, L., Sauer, A. V, and Toka, O. (2017). A map of human circular RNAs in clinically relevant tissues.
Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.
Mills, E.W., Wangen, J., Green, R., and Ingolia, N.T. (2016). Dynamic Regulation of a Ribosome Rescue Pathway in Erythroid Cells and Platelets. Cell Rep. *17*, 1–10.
Nicolet, B.P., Engels, S., Aglialoro, F., van den Akker, E., von Lindern, M., and Wolkers, M.C. (2018). Circular RNA expression in human hematopoietic cells is widespread and cell-type specific. Nucleic Acids Res. *46*, 8168–8180.
Van Oorschot, R., Hansen, M., Koornneef, J.M., Marneth, A.E., Bergevoet, S.M., Van Bergen, M.G.J.M., Van Alphen, F.P.J., Van Der Zwaan, C., Martens, J.H.A., Vermeulen, M., et al. (2019). Molecular mechanisms of bleeding disorderassociated GFI1BQ287* mutation and its affected pathways in megakaryocytes and platelets. Haematologica *104*, 1460–1472.
Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan,

M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol. Cell *66*, 9–21.e7.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.
Petriv, O.I., Kuchenbauer, F., Delaney, A.D., Lecault, V., White, A., Kent, D., Marmolejo, L., Heuser, M., Berg, T., Copley, M., et al. (2010). Comprehensive microRNA expression profiling of the hematopoietic hierarchy. Proc. Natl. Acad. Sci. *107*, 15443–15448.
Piwecka, M., Glažar, P., Hernandez-Miranda, L.R., Memczak, S., Wolf, S.A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Cerda Jara, C.A., Fenske, P., et al. (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. Science (80-.). *8526*, eaam8526.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., et al. (2015). Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. Mol. Cell *58*, 870–885.
Stagsted, L.V., Nielsen, K.M., Daugaard, I., and Hansen, T.B. (2019). Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles. Life Sci. Alliance *2*, e201900398.
Sun, Y.M., Wang, W.T., Zeng, Z.C., Chen, T.Q., Han, C., Pan, Q., Huang, W., Fang, K., Sun, L.Y., Zhou, Y.F., et al. (2019). CircMYBL2, a circRNA from MYBL2, regulates FLT3 translation by recruiting PTBP1 to promote FLT3-ITD AML progression. Blood *134*, 1533–1546.
Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res. *34*.
Westholm, J.O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S.E., Graveley, B.R., and Lai, E.C. (2014). Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. Cell Rep. *9*, 1966–1980.
Wickham, H. (2009). ggplot2□: Elegant Graphics for Data Analysis (New York, NY: Springer New York).
Wu, N., Yuan, Z., Du, K.Y., Fang, L., Lyu, J., Zhang, C., He, A., Eshaghi, E., Zeng, K., Ma, J., et al. (2019). Translation of yes-associated protein (YAP) was antagonized by its circular RNA via suppressing the assembly of the translation initiation machinery. Cell Death Differ.
Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L.-L., Wang, Y., et al. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. Cell Res. *27*, 626–641.
Zhang, X.O., Wang, H. Bin, Zhang, Y., Lu, X., Chen, L.L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. Cell *159*, 134–147.
Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.L., Yang, L., and Chen, L.L. (2016). The Biogenesis of Nascent Circular RNAs. Cell Rep. *15*, 611–624.

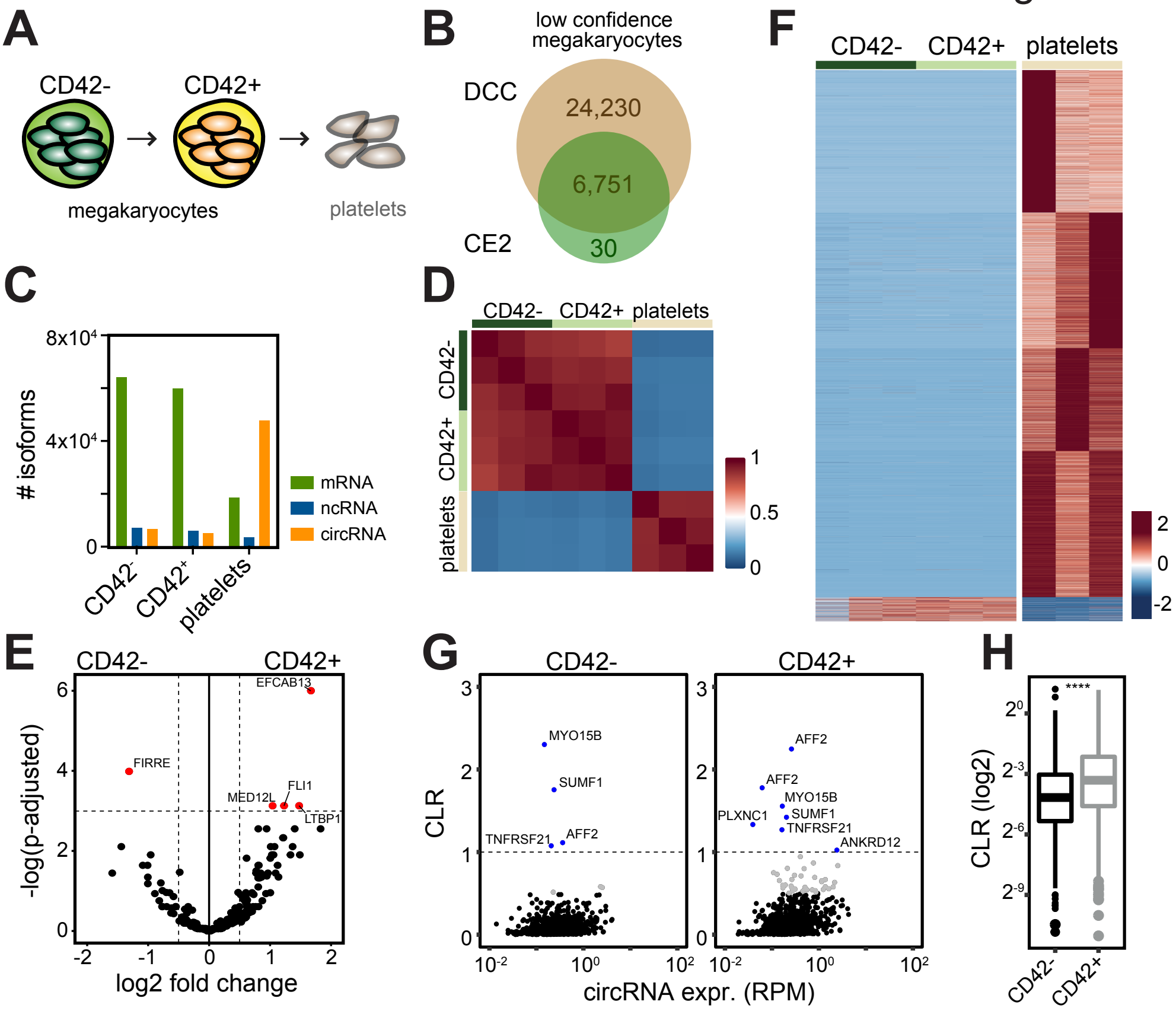
Nicolet et al. Figure 1



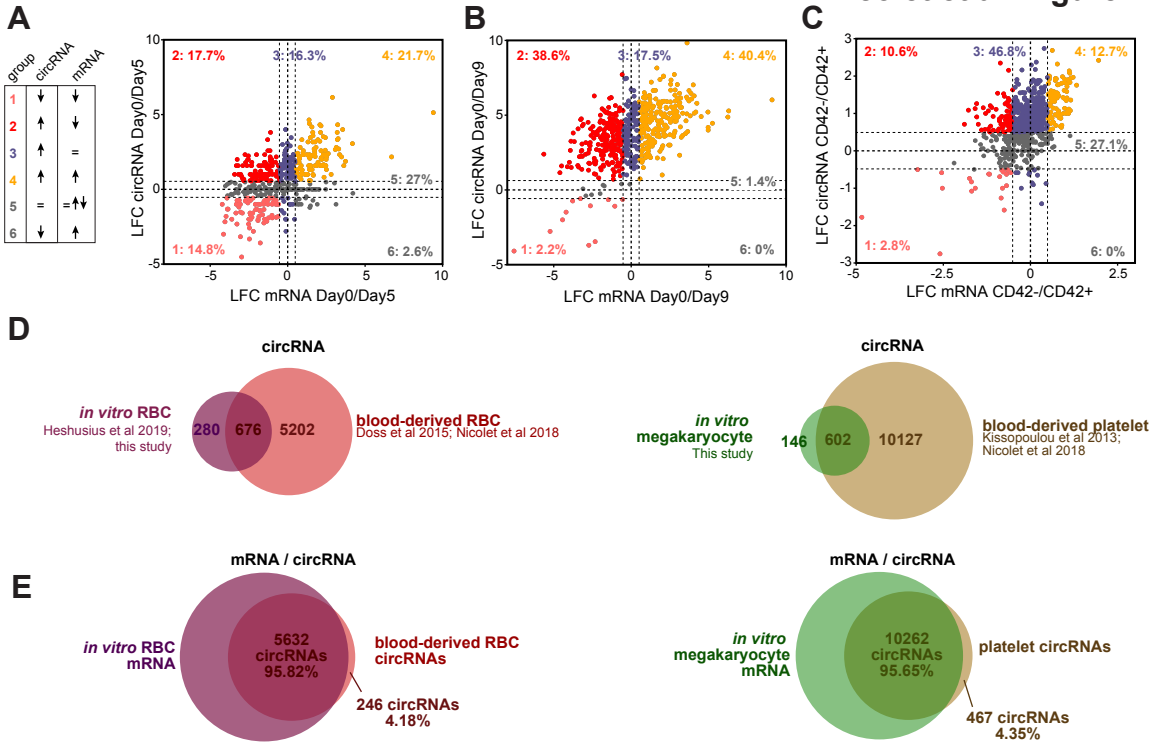
Nicolet et al. Figure 2



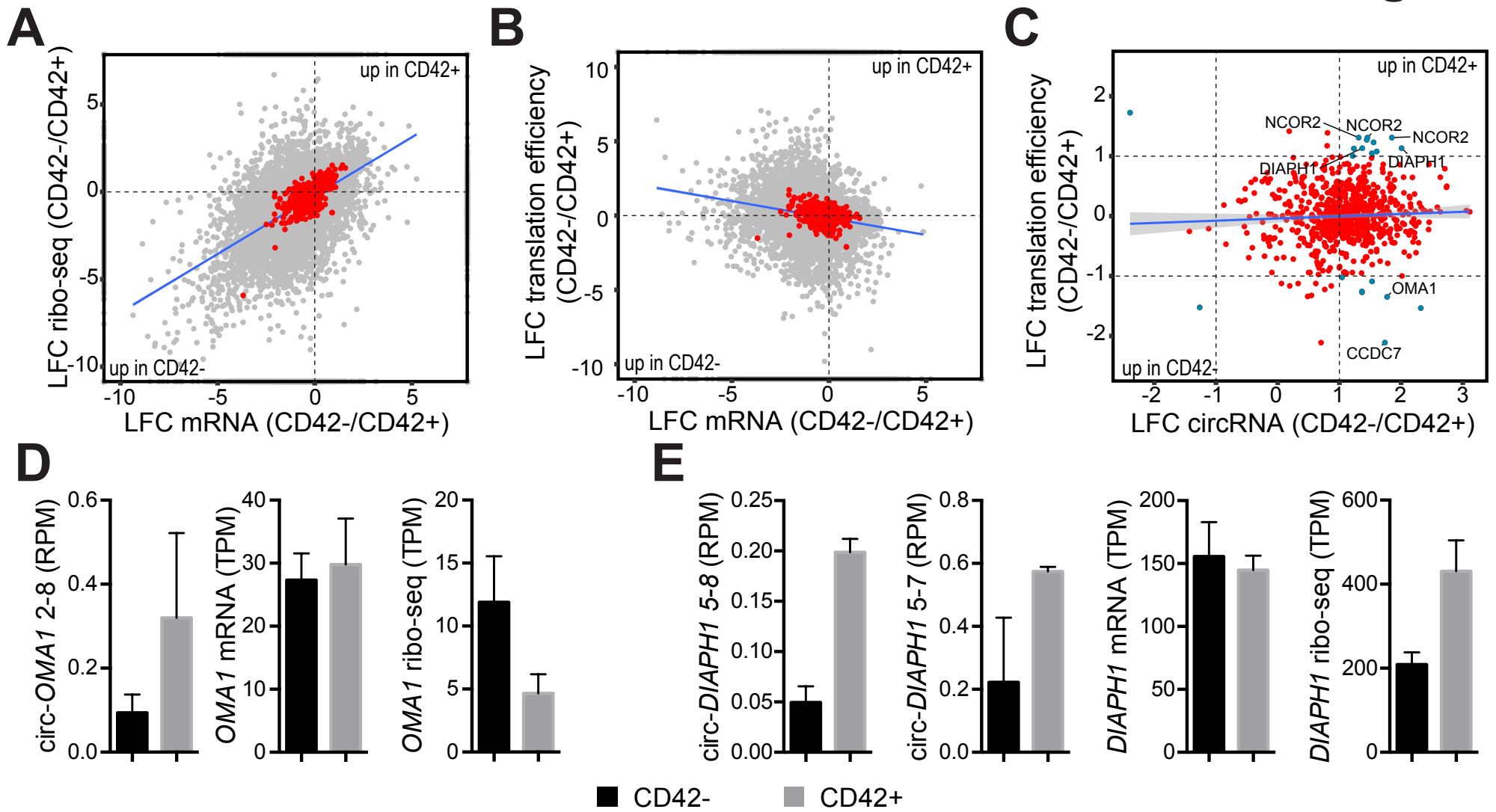
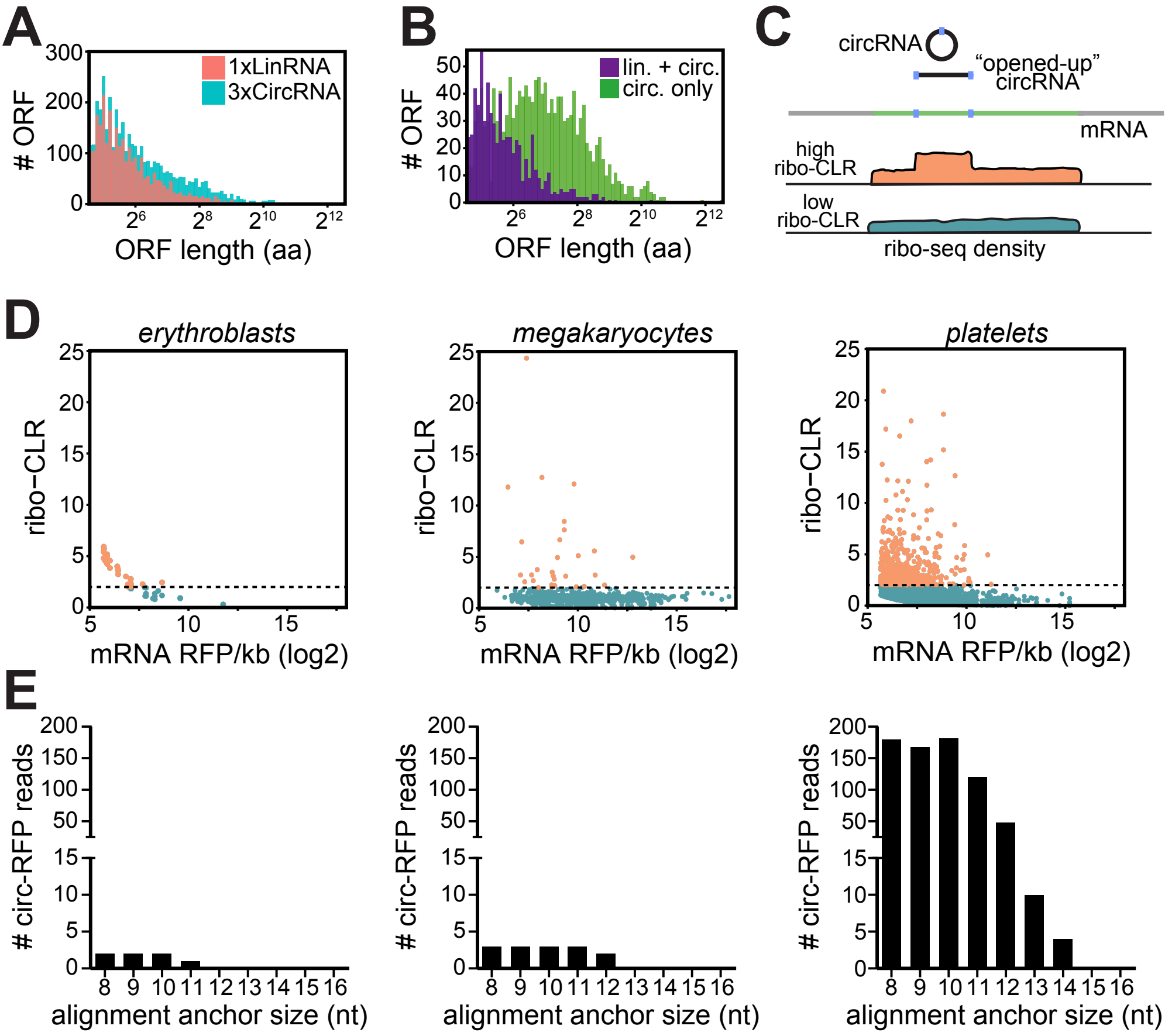
Nicolet et al. Figure 3



Nicolet et al. Figure 4



Nicolet et al. Fig. 5



Nicolet et al. Fig. 6