

PLATELETS AND THROMBOPOIESIS

Circular RNA enrichment in platelets is a signature of transcriptome degradation

Abd A. Alhasan,^{1,*} Osagie G. Izuogu,^{1,*} Haya H. Al-Balool,² Jannetta S. Steyn,^{1,3} Amanda Evans,⁴ Maria Colzani,⁴ Cedric Ghevaert,⁴ Joanne C. Mountford,^{5,6} Lamin Marenah,⁶ David J. Elliott,¹ Mauro Santibanez-Koref,¹ and Michael S. Jackson¹

¹Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom; ²Kuwait Medical Genetics Centre, Al-Sabah Medical Area, Kuwait City, Kuwait; ³Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; ⁴National Health Service Blood and Transplant, Department of Haematology, University of Cambridge, Cambridge, United Kingdom; ⁵Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom; and ⁶Tissue and Cellular Therapies, Research Development & Innovation, Scottish National Blood Transfusion Service, Glasgow, United Kingdom

Key Points

- Circular RNAs are hugely enriched in platelets compared with nucleated cell types.
- Lack of enrichment in megakaryocyte progenitors implicates degradation of platelet linear RNAs.

In platelets, splicing and translation occur in the absence of a nucleus. However, the integrity and stability of mRNAs derived from megakaryocyte progenitor cells remain poorly quantified on a transcriptome-wide level. As circular RNAs (circRNAs) are resistant to degradation by exonucleases, their abundance relative to linear RNAs can be used as a surrogate marker for mRNA stability in the absence of transcription. Here we show that circRNAs are enriched in human platelets 17- to 188-fold relative to nucleated tissues and 14- to 26-fold relative to samples digested with RNase R to selectively remove linear RNA. We compare RNAseq read depths inside and outside circRNAs to provide in silico evidence of transcript circularity, show that exons within circRNAs are enriched on average 12.7 times in platelets relative to nucleated tissues and identify 3162 genes

significantly enriched for circRNAs, including some where all RNAseq reads appear to be derived from circular molecules. We also confirm that this is a feature of other anucleate cells through transcriptome sequencing of mature erythrocytes, demonstrate that circRNAs are not enriched in cultured megakaryocytes, and demonstrate that linear RNAs decay more rapidly than circRNAs in platelet preparations. Collectively, these results suggest that circulating platelets have lost >90% of their progenitor mRNAs and that translation in platelets occurs against the backdrop of a highly degraded transcriptome. Finally, we find that transcripts previously classified as products of reverse transcriptase template switching are both enriched in platelets and resistant to decay, countering the recent suggestion that up to 50% of rearranged RNAs are artifacts. (*Blood*. 2016;127(9):e1-e11)

Introduction

Platelets are small, anucleate, circulating blood cells derived from polyploid megakaryocyte progenitors¹ and are vital for functions such as hemostasis, wound healing, and angiogenesis.² Unlike anucleate erythrocytes, which have a lifespan of ~4 months and have greatly reduced ribosome numbers,³ platelets are short lived (~8-11 days⁴) and translationally competent,⁵ and a number of genes have been shown to undergo cytoplasmic splicing on platelet activation.^{6,7} Consistent with these observations, recent RNAseq analyses have defined a complex platelet transcriptome.⁸⁻¹¹ However, some transcripts believed to contribute to platelet function are present at very low levels in RNAseq data,¹¹ and attempts to integrate transcriptome and proteome data have given conflicting results.¹²⁻¹⁵ In the absence of nuclear transcription, RNAs within platelets are assumed to degrade over time. Although the abundance of individual mRNAs has been shown to fall in platelets at ambient temperatures,¹⁶ the stability and integrity of mRNAs within circulating platelets is poorly characterized at the transcriptome level.

However, this is highly relevant to functional inferences drawn from RNAseq data.^{9,11,15,17,18}

Transcripts with rearranged exon order are now known to be common in eukaryotic organisms and are characterized by back-splice exon junctions inconsistent with the underlying genomic DNA.¹⁹⁻²⁴ Evidence for linear and polyadenylated structures has been presented,^{20,25,26} but it is now clear that most are circular (circRNAs) and cytoplasmic.²¹⁻²³ Recent experiments have established that intronic repeats flanking rearranged exons can promote circularization and that circRNAs primarily use canonical splice sites²⁷⁻³⁰ and can be regulated by RNA binding proteins.^{29,31} Although 2 noncoding circRNAs act as miRNA sponges,^{22,26} and circRNAs with retained introns (exon-intron or ElcircRNAs) can enhance transcription of their parental genes through interaction with U1 snRNP and RNA PolIII,³² the vast majority of circRNAs consist of exons from protein encoding genes and have no known function.^{22,33}

Submitted June 3, 2015; accepted December 1, 2015. Prepublished online as *Blood* First Edition paper, December 10, 2015; DOI 10.1182/blood-2015-06-649434.

*A.A.A. and O.G.I. contributed equally to this work.

This article contains a data supplement.

There is an Inside *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2016 by The American Society of Hematology

circRNAs are resistant to attack by exonucleases involved in the regulation of mRNA,³⁴ and resistance to one, RNase R, has been used to define them.^{23,35} Furthermore, blocking of transcription can increase circRNA levels relative to linear.²³ This differential stability raises the possibility that abundance of circRNAs relative to linear forms could provide insight into the platelet transcriptome. Here, we analyze the circRNA population of platelets and compare it to populations in nucleated cells, anucleate erythrocytes, and cells where linear RNA has been digested with RNase R. We show that circRNAs are highly enriched in both platelets and erythrocytes relative to nucleated tissues and confirm this experimentally using quantitative polymerase chain reaction (qPCR). We identify >3000 genes where circRNAs are significantly enriched relative to nucleated cell types, including some genes where only circRNAs producing exons are represented within the RNAseq data. We also use qPCR to establish that circRNAs are not enriched in cultured megakaryocytes and that, over time, the levels of linear RNAs fall relative to circRNAs in platelet preparations. Our results suggest that >90% of mRNAs have been lost from circulating platelets and have potentially important implications for both platelet biology and our understanding of circRNAs.

Materials and methods

Datasets used

In silico analyses were performed on RNAseq data generated during this study and on 36 publicly available datasets: 24 ribosomal RNA depleted total RNA samples sequenced on the Illumina platform,^{11,23,27,36,37} 2 polyA⁺ selected samples sequenced on the Illumina platform,^{11,38} and 10 nonribosomal RNA depleted total RNA samples sequenced on the SOLiD platform.¹⁵ For details, see supplemental Methods and supplemental Table S1 available on the *Blood* Web site.

Erythrocyte RNAseq

Ribosome-depleted RNAseq of total RNA from red blood cells (RBCs) was performed by AROS Applied Biotechnology (Aarhus, Denmark) using the TruSeq RiboZero Stranded mRNA LT kit (Illumina). The library was sequenced using a paired-end 100-bp protocol according to the manufacturer's recommendations (supplemental Methods).

circRNA identification

Identification of back-splice exon junctions was performed using PTESFinder Pipeline v. 1 (<http://sourceforge.net/projects/ptesfinder-v1/>). This identifies RNAseq reads that map in inverted order but the same orientation to the same RefSeq entry, generates sequence models of back-splice junctions, and remaps all reads to the models. It then removes reads that do not use known exon junctions or that map equally well or better to the genomic or transcriptome reference (see supplemental Methods and supplemental Figure S1 for details).

Biological samples: origin and preparation

Tissue RNAs were obtained from Biochain (Amsbio, UK). RBCs were obtained from the Scottish National Blood Transfusion Service in accordance with the terms of the standard donor consent form and information and with approval of the Scottish National Blood Transfusion Service Sample Governance Committee. Cord blood was obtained after informed consent under a protocol approved by the National Research Ethics Service (Cambridgeshire 4 Research Ethics Committee ref. no. 07/MRES/44). Whole blood samples were collected from healthy volunteers with approval from Newcastle University's Faculty of Medical Sciences Ethics Committee. Platelets and platelet-rich plasma (PRP) were isolated according to a protocol that uses a platelet inhibitory cocktail.³⁹ Isolation of peripheral blood mononuclear cells (PBMCs) and RBCs was performed using Histopaque-1.077 density gradients (Sigma-Aldrich).

Megakaryocytes were cultured and isolated according to the protocol of Tjissen et al.⁴⁰ For further details, see supplemental Methods.

RNA isolation and cDNA generation

RNA was isolated using Trizol (Lifetechnologies) and treated with DNase I (Promega) according to manufacturer's instructions. RNA quantitation was performed using the NanoDrop ND-1000 spectrophotometer (Thermo Scientific), and quality was assessed using the Agilent 2100 Bioanalyser (Agilent Technologies). cDNA was synthesized using high-capacity cDNA kits (Applied Biosystem), random hexamers, and Moloney murine leukemia virus (MMLV) unless stated otherwise.

Real-time PCR

Real-time PCR was performed using Taqman mastermix (Life Technologies) as described previously.²⁰ Primers and assay efficiencies are given in supplemental Table S2. Transcript expression was normalized using the Δ CT method relative to the geometric mean of 4 housekeeping genes (*GAPDH*, *PPIA*, *TUBB*, and *GUSB*) analyzed using TaqMan gene expression assays (Applied Biosystems). Circular transcripts were also normalized against the linear transcript from the same gene where appropriate. The Wilcoxon rank sum test was used to compare levels of circRNA vs linear RNA across multiple structures (analysis of paired $-\Delta$ Ct values) and to compare linear vs circRNA expression ratios between conditions or sample types (analysis of $-\Delta\Delta$ Ct values across all structures).

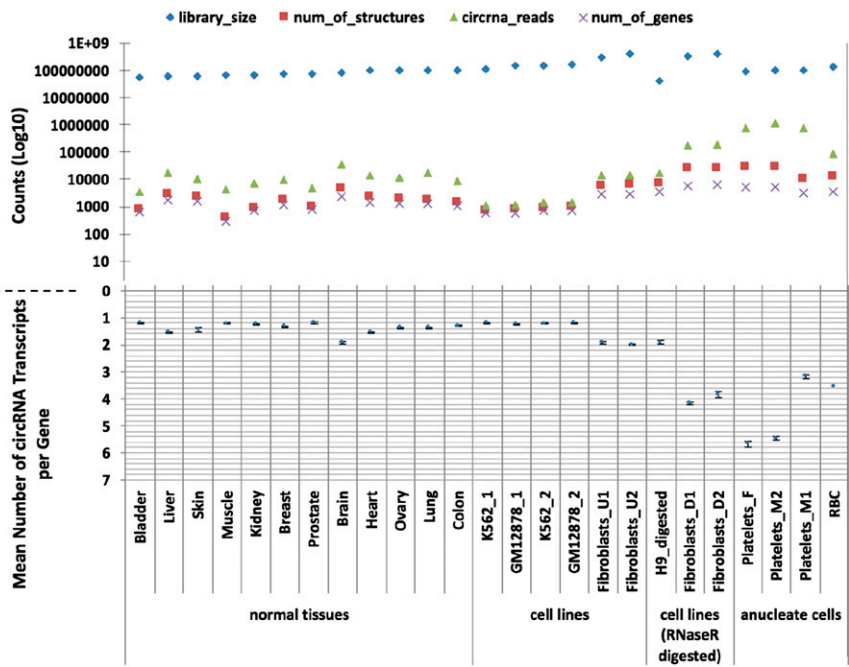
Results

Platelets are highly enriched for circRNAs

We first identified 3 publicly available RNAseq datasets from human platelets,^{9,11,15} one of which had a sequence read length long enough for the identification of back-splice junctions within circRNAs and included samples processed to retain unpolyadenylated RNAs (Array express ID E-MTAB-1846; 100-bp Illumina ribosome-depleted total RNAseq from 3 individuals, and polyA⁺ RNAseq from 1 individual¹¹). Analysis of the 3 total RNAseq samples using a back-splice exon junction discovery pipeline (PTESFinder; see "Materials and methods") identified 33 829 distinct structures consistent with circRNAs (supplemental Figure S1; supplemental Table S3), ~4 to 5 times more than reported for fibroblasts²³ and ~15 times more than reported for leukocytes and HEK293 cells²² in libraries of comparable size. Analysis with independent software previously used for identification of both genic and nongenic circRNAs²² also identified an excess in platelets relative to these datasets (supplemental Figure S1). Furthermore, ~63% of the 33 829 structures have previously been identified within RNAseq data generated following RNase R digestion (to selectively remove linear transcripts^{23,27}), consistent with them being circRNAs. Despite the large number identified, the majority are not platelet specific, as 24 632 (~73%) were also present in ≥ 1 of 15 ENCODE (Encyclopedia of DNA Elements) RNAseq datasets previously mined by Salzman et al⁴¹ (supplemental Figure S1).

Having identified an apparent abundance of circRNAs in platelets, we then used PTESFinder to identify circRNAs within other publicly available datasets generated using ribosome depleted total RNAseq and sequenced using the Illumina platform. These comprised 12 human tissues,³⁷ ENCODE cell lines,⁴² and cell lines treated with RNase R to selectively remove linear RNAs.^{23,27} To enable comparison with a distinct anucleate cell type, we also generated >120 million 100-bp paired-end reads from total RNA extracted from leukocyte depleted RBCs (GEO record no. GSE69192; see "Materials and methods"). Analyses of these data, summarized in Figure 1 and presented in supplemental Table S3, show that 17 to 188 times more circRNA

Figure 1. CircRNAs in human tissues. The number of circRNAs (structure counts), junction reads (back-splice read counts), circRNA producing genes (gene counts), mean number of circRNA transcripts per circRNA producing genes, and RNAseq library sizes (read counts) are shown. Samples are presented as 4 groups: nucleated tissues, nucleated cell lines, nucleated cell lines digested with RNase R, and anucleated cell types. The number of back-splice reads is significantly higher in anucleate samples (platelets and RBCs) than all others ($P = 2.7 \times 10^4$, Wilcoxon rank sum test; figures adjusted for library size). All data were processed using PTES-Finder, and all were publicly available with the exception of the data from RBCs (see “Materials and methods” and supplemental Table S1 for details of datasets, and supplemental Table S3 for details of structures).



(back-splice) reads are present in platelets than in human nucleated tissues (7579-11 138 circRNA junction reads per million reads vs 59-447), and 14 to 26 times more than in samples treated with RNase R (427-463 junction reads per million reads). In addition, the mean numbers of circRNA structures per circRNA producing gene are larger in anucleate and RNase R digested samples than in others, with 2 of the 3 platelet samples averaging >5 structures per gene compared with 1 to 2 for nucleated tissues and cell lines.

We also analyzed the platelet polyA⁺ sample from E-MTAB-1846 and identified 841 circRNA structures from 453 genes. However, their nucleotide composition was found to be significantly enriched for A residues relative to structures present only in platelet total RNA ($\chi^2 = 12\,317$; $P < 2.2e^{-16}$; supplemental Table S4). As this suggests these structures are nonpolyadenylated circRNA transcripts rich in A nucleotides, they were not analyzed further.

More than 600 platelet genes generate more than 20 distinct circRNAs

To illustrate the variation in circRNA distributions, the numbers of circRNA structures identified per gene are displayed for 4 samples in Figure 2A. Strikingly, both platelets and fibroblasts digested with RNase R (which increases circRNA levels by degrading linear RNAs) have large numbers of genes generating >15 distinct circRNA structures. For example, the platelet sample has >600 genes each with >20 different circRNAs. Furthermore, as shown in Table 1, individual structures in anucleate samples are supported by much larger back-splice junction read counts than in nucleated tissues ($P = 2.7 \times 10^{-4}$ anucleate vs nucleated counts, Wilcoxon rank sum test). All 3 platelet samples have >1000 structures supported by >100 reads, in sharp contrast to undigested nucleated samples. As an example of the diversity of structures and read support within platelets, the 49 circRNA structures identified within the *XPO1* gene are shown in Figure 2B, with supporting read counts ranging from 1 (E8-E5 and E16-E7) to 4045 (E4-E2). It is noteworthy that all 21 internal exons of this gene are encompassed by ≥ 1 circRNA (illustrating that most exons within some genes can contribute to circRNAs). Only 11 of these structures were identified previously in RNase R digested H9 RNA.²⁷

Real-time PCR confirmation of high circRNA transcript levels in anucleate cells

To experimentally confirm enrichment of circRNAs in both anucleate cell types, we used qPCR to assess relative levels of linear and circRNAs. We first assayed the CD45 antigen (*PTPRC*), which is highly expressed in PBMCs, and performed cell counts to confirm that PBMCs were being efficiently removed during platelet isolation (supplemental Figure S2; supplemental Methods). As a precaution against reverse transcriptase (RT)-specific artifacts,⁴³ we also checked that all circRNAs being analyzed could be amplified using templates generated with both avian myeloblastosis virus (AMV) and MMLV RTs (supplemental Figure S3). We then analyzed the abundance of 2 known circRNA structures from *MANIA2* (E4-E2 and E5-E2²²) relative to the canonical *MANIA2* transcript, 2 structured from *PHC3* (a gene expressed in platelets for which assays were already available²⁰), and an additional 7 abundant circRNAs with their associated linear RNAs (supplemental Table S2). For each, paired assays with a common probe were used (Figure 3A).

Both of the *MANIA2* circRNA junctions are much more abundant in platelets and RBCs than the linear mRNA (Figure 3B). However, they are less abundant than the linear transcripts in virtually all nucleated tissues analyzed. This is consistent with the RNAseq data and previous analyses of nucleated tissues.^{20,22} The other 9 circRNAs are also expressed at higher levels in platelets and RBCs than their corresponding linear structures (Figure 3C), with circRNAs from *SMARCA5*, *UBXN7*, and *PNN* registering ~6 to 10 cycles before their corresponding linear RNAs, suggesting they are ~50- to 1000-fold more abundant. In contrast, all are expressed in nucleated cells at an equivalent or lower level than their linear counterparts. Collectively, the difference in relative expression levels of linear and circRNAs between anucleate and nucleated tissues is highly significant ($P = 2.7 \times 10^{-12}$, Wilcoxon rank sum test).

Finally, the impact of RNase R digestion on expression was also assessed (Figure 3D). Digestion increases the levels of all back-splice junctions relative to the canonical linear junctions in each cell type analyzed ($P = 7.8 \times 10^{-3}$, Wilcoxon paired rank sum test; mock digest vs RNase R digested), although the *MANIA2* E5-E2 back-splice

Table 1. circRNA read support for structures and their frequency in all samples

circRNA reads	Bladder	Liver	Skin	Muscle	Kidney	Breast	Prostate	Brain	Heart	Ovary	Lung	Colon	K562_1	GM12878_1	K562_2	GM12878_2	Fibroblasts_U1	Fibroblasts_U2	H9_D	Fibroblasts_D1	Fibroblasts_D2	Fibroblasts_F	Platelet_M1	Platelet_M2	RBCs
>10	55	308	164	136	196	186	71	670	235	263	521	190	3	9	11	6	125	141	227	2797	3052	7281	9605	8459	1272
>100	0	11	3	2	2	3	1	19	5	3	6	5	0	0	0	0	4	2	2	200	224	1197	1686	1325	96
>1000	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	2	5	83	147	95
K/M	66	295	170	69	104	126	59	447	148	125	172	84	9	7	9	9	45	36	427	546	464	8469	11138	7579	597

For each sample, the number of distinct circRNA structures supported by >10, 100, and 1000 circRNA (back-splice) reads are shown, together with the number of circRNA reads per million reads (K/M). D, digested with RNase R; U, undigested, matched sample.

shows only a marginal increase in the platelet sample. This confirms that these structures include circRNAs. However, the impact of RNase R digestion is more pronounced in PBMCs and in HEK293 cells than in either anucleate sample ($P < .02$, Wilcoxon paired rank sum test for all pairwise nucleated vs anucleated comparisons; Figure 3D).

For some genes, only reads from circRNA producing exons are detected in platelets

The abundance of individual circRNAs relative to mRNAs from the same loci has previously been estimated in silico by comparing back-splice junction read counts to flanking canonical junction counts.^{20,21,44} However, because most circRNA producing genes generate multiple overlapping structures (Figure 2B; supplemental Table S3), the canonical junctions flanking one circRNA are often present within others. We therefore performed a transcriptome-wide analysis where we calculated for each gene the reads per kilobase of transcript per million mapped reads (RPKM) across all exons internal to all known circRNA structures (RPKM_I), and the RPKM across exons external to all known circRNAs (RPKM_E). The former is a measure of both linear and circRNA expression, the latter is linear only. Box and whisker plots showing RPKM_I/RPKM_E ratios for all circRNA producing genes in all samples are presented in Figure 4A. The ratios from individual genes extend over 6 to 8 orders of magnitude for each sample, but what is most striking is that the median RPKM_I/RPKM_E ratios of the anucleate samples are ~12 times higher than nucleated tissues and ~5 times higher than samples digested with RNase R (Figure 4A). The distribution of RNAseq reads therefore reflects the pattern of circRNA structures inferred from exon-junction read counts alone (Figures 1 and 2) and provides further evidence that back-splice junctions are predominantly derived from circRNAs.

To identify genes significantly enriched for circRNAs in platelets, we then estimated the contribution of circRNA producing exons to total transcription for each gene by calculating RPKM_I/(RPKM_I + RPKM_E) ratios and compared ratios in the platelet samples to ratios in the 12 nucleated tissues using the Wilcoxon rank test ($P = .05$, Benjamin and Hochberg false discovery rate = 0.01). Of 8041 circRNA producing genes analyzed, 3162 showed significantly enriched transcription from circRNA-producing exons in platelets (supplemental Table S5). When the ratios from all genes with an average RPKM > 1 in platelets are plotted (Figure 4B), it can be seen that for the vast majority of enriched genes the contribution of circRNA exons to total transcription is >0.6 in nucleated tissues but >0.8 in platelets. Furthermore, for 457 genes, transcription from circRNA-producing exons exceeds an average of 99% in the platelet samples, and for 15 genes, it is 100% in all 3 samples (supplemental Table S5). As this assay is not dependent on exon junction counts, which require long sequence reads for accurate detection, we applied it to one of the SOLiD platelet RNAseq datasets¹⁵ and confirmed the presence of similar patterns of circRNA exon enrichment (supplemental Figure S4; supplemental Data File 1).

Reads from circRNA producing exons are enriched in platelets by up to 3000-fold

To estimate the magnitude of circRNA exon enrichment for each gene in platelets relative to other tissues, we further analyzed genes with an average RPKM_I >1 by normalizing their mean platelet RPKM_I/RPKM_E ratio against the mean RPKM_I/RPKM_E ratio from all 12 nucleated samples. The average enrichment was 12.7 times for all genes and 22 times for the 3162 significantly enriched genes, with the highest enrichment being 3590 times for *TMEM181* (Figure 4C; supplemental Table S5). To illustrate these extreme levels of enrichment, RNAseq read counts and circRNA junction read counts for 4 enriched

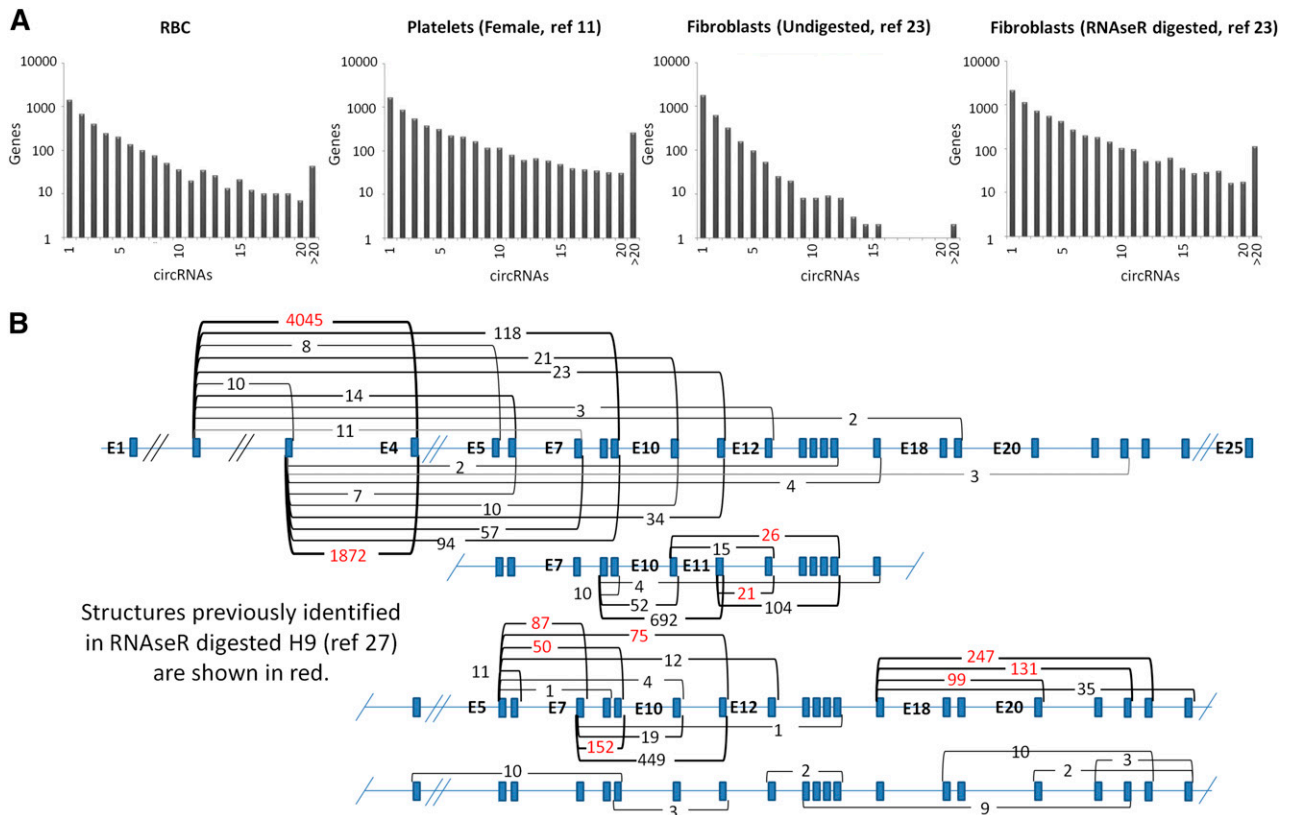


Figure 2. circRNA structures per gene. (A) Histograms showing circRNA numbers per gene established using PTSEFinder for RBCs (this study), platelets,¹¹ fibroblasts, and fibroblasts digested with RNase R.²³ (B) Schematic diagram showing *XPO1* intron/exon organization with read counts and inferred structure of all circRNAs identified within the 3 platelet samples. Inferred structures assume that all internal Refseq exons are present within the structure defined by each back-splice. Read counts for the 11 structures identified previously within RNase R digested H9 ES cells²⁷ are highlighted in red.

genes from a single platelet sample are shown in Figure 5A-D. The vast majority of platelet RNAseq reads from *PNN* (enrichment 2189 times) map to exons 7 and 8, which correspond to the most abundant circRNA (Figure 5A); the majority of reads from *SLC20A2* (enrichment 152 times) map to the abundant circRNAs involving exons 3 to 5 (Figure 5B); all reads from *POMT1* (enrichment 127 times) map to exons within its only circRNA (E4-E3; Figure 5C), whereas for *PHC3* (enrichment 240 times), reads map largely to the abundant E6-E5 and E7-E5 structures (Figure 5D).

Although these in silico analyses point to enrichment specifically in anucleate cells, no total RNAseq data have been generated from the direct progenitors of these cells. For all genes with assays available, we therefore also analyzed the relative levels of linear and circRNAs in megakaryocytes cultured from cord blood. These were harvested after 10 days, when >70% of cells were positive for CD41 and CD42 (supplemental Methods). The results, presented in Figure 5E-F, establish that relative levels of linear and circRNA in megakaryocytes are similar to other nucleated cell types (Figure 2B); all circRNAs are expressed at a much lower level relative to linear RNAs than in platelets ($P = 1.4 \times 10^{-6}$, Wilcoxon paired rank sum test across all structures). This is consistent with circRNA enrichment after platelet formation.

Confirmation of differential decay of linear and circRNAs in platelet preparations

Human mRNAs are subject to a wide variety of decay pathways⁴⁵ and exhibit extensive variation in decay rates.⁴⁶⁻⁴⁸ To investigate decay in platelets, we first correlated the change in transcript

abundance between megakaryocyte PolyA⁺ RNA and platelet PolyA⁺ RNA with published estimates of mRNA half-life.⁴⁶ This identified a weak but highly significant correlation, despite the difference in cell types between datasets, with short half-life genes showing a more pronounced reduction in transcript abundance ($r = 0.17$, $P = 3.35e^{-44}$; Figure 6A). This is consistent with a transcriptome-wide role for mRNA decay in platelet expression levels.

To experimentally investigate differential decay, we then used real-time PCR to analyze changes in linear and circRNA abundance in PRP over time (see “Materials and methods”). Consistent with previous analyses,¹⁶ initial experiments suggested that all transcript levels remained stable at 4°C but that degradation of mRNA could be detected when incubation temperature was elevated (data not shown). To assess decay at a physiologically relevant temperature, we assayed the relative abundance of linear and circRNAs in PRP from 3 individuals, before and after incubation at 37°C for 72 and 96 hours (Figure 6B). The levels of all 4 housekeeping genes were reduced by incubation, with CT values increasing by 4 to 6 cycles (Figure 6B, top left panel). Reduction was also observed when genes which are highly abundant in platelets were analyzed (supplemental Figure S5). Normalization of linear transcripts from circRNA producing genes against the housekeeping pool (Figure 6B, top right panel) suggests that there are differences in decay rates between genes, with *UBXN7*, *PNN*, and *MANIA2* transcripts being enriched relative to control housekeeping RNAs. Strikingly, however, circRNAs show significant enrichment (2- to 6-fold) relative to linear RNAs after incubation (Figure 6B, lower panel; $P < 2 \times 10^{-3}$ for both day 0 vs 72 hours and day 0 v 96 hours, Wilcoxon rank sum test across all structures). This confirms that linear structures decay more rapidly than circRNAs in platelets.

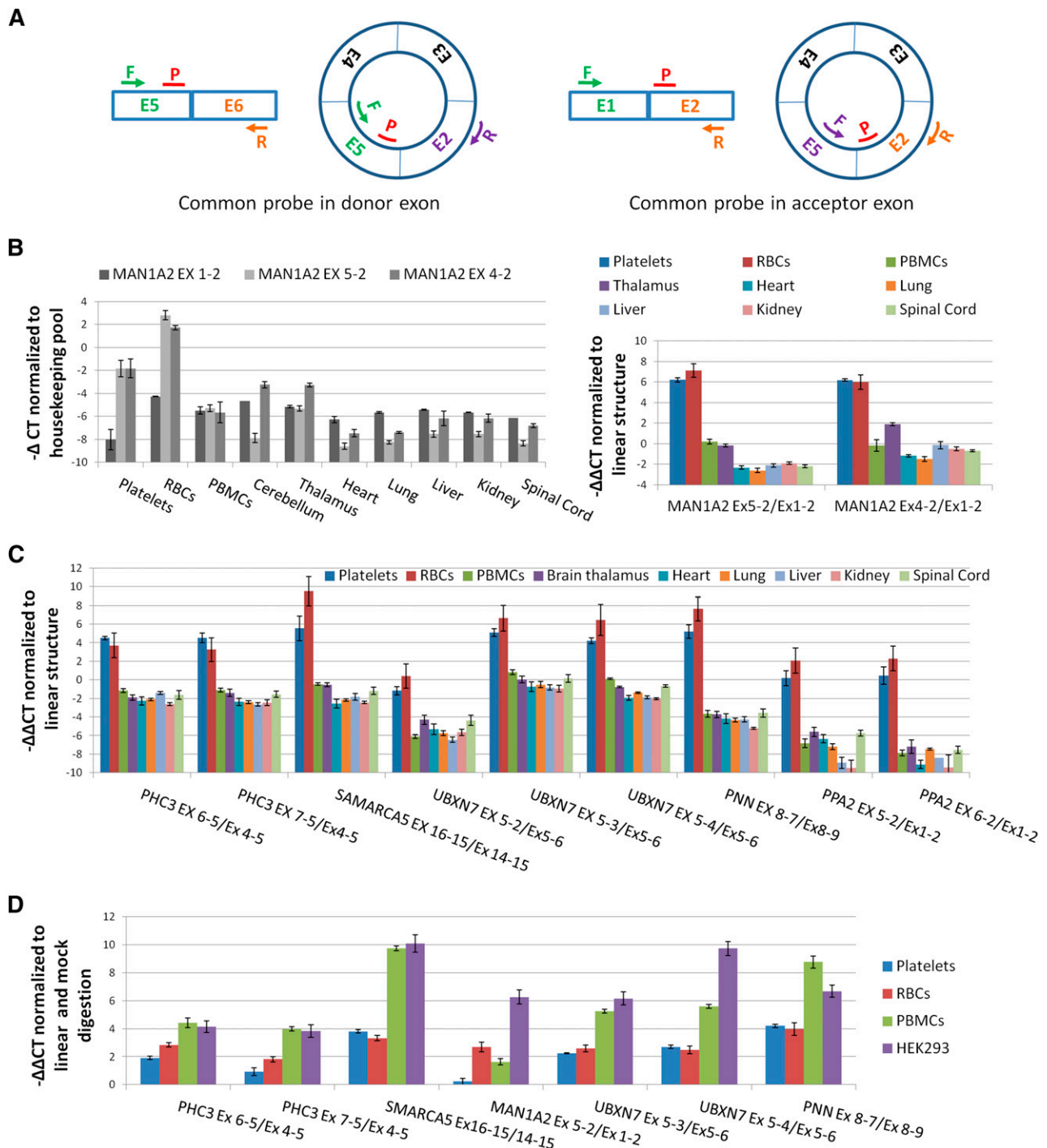


Figure 3. Confirmation of circRNA abundance and resistance to RNase R. (A) Schema of qPCR assays using an E5-E2 circRNA as an example. All assays use a common reporter probe and use either an exon downstream of the circRNA to assay linear expression (probe in donor exon) or an exon upstream (probe in acceptor exon). (B) (Left) Expression levels ($-\Delta\text{CT}$ values) of linear (Ex1-2) and circular (Ex5-2 and Ex4-2) *MAN1A2* transcripts relative to housekeeping pool. (Right) Expression levels of circRNAs relative to linear RNAs from the same loci normalized to housekeeping pool ($-\Delta\text{CT}$ values). (C) Expression of 9 circRNAs relative to linear forms from the same loci, normalized to housekeeping pool ($-\Delta\text{CT}$ values). (D) Change in CT values of circRNAs relative to linear forms from the same loci in RNase R digested RNAs, normalized to mock digested RNAs ($-\Delta\text{CT}$ values). Templates, circRNAs, and linear forms assayed are indicated.

Discussion

Platelets are anucleate yet retain machinery required for splicing and translation, and there is extensive evidence that genes are translated in activated platelets.^{5-7,49-51} As a result, recent platelet transcriptome analyses have generated considerable interest.^{8,17,18} However, here we

showed that the platelet transcriptome is significantly enriched for circRNAs as assayed using back-splice junction counts, real-time PCR, and an RNAseq comparative read depth approach. Back-splice reads are enriched 17- to 188-fold relative to nucleated tissues, and >3000 genes have a statistically significant enrichment of reads from circRNA-producing exons. Furthermore, for some genes, only RNAseq reads derived from circRNA-producing exons are detected within platelet

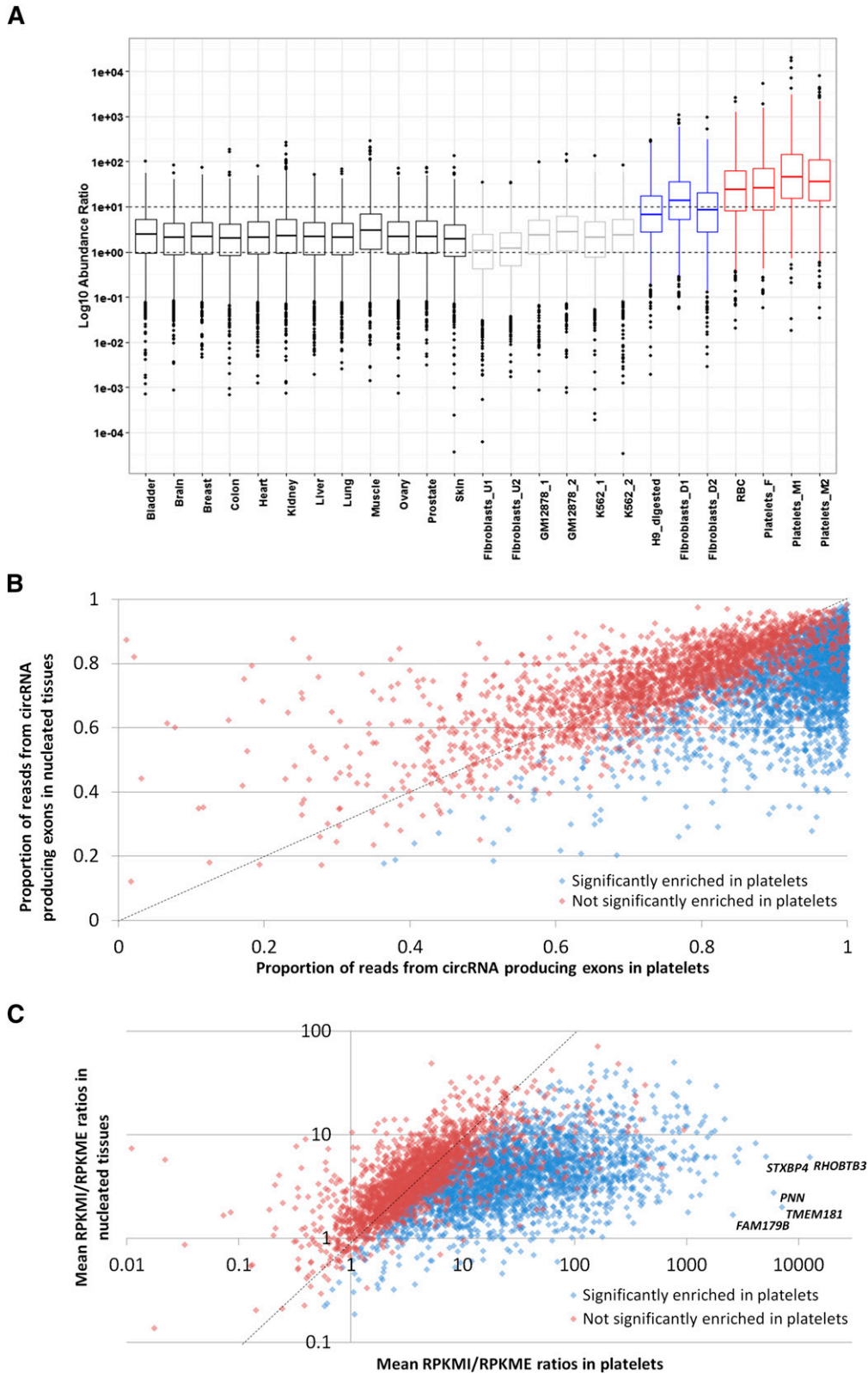


Figure 4. Differential read depth defines genes with significant circRNA enrichment in platelets. (A) Box and whisker plots showing the ratio of RPKM from circRNA producing exons (RPKM_I) to RPKM from exons that do not produce circRNAs (RPKM_E) for all genes in each sample. The median and upper and lower quartiles are shown, with outliers as solid circles. (B) The proportion of reads from circRNA producing exons averaged across all nucleated samples (y-axis) and platelets (x-axis). (C) Fold enrichment of reads from circRNA producing exons in platelets relative to nucleated tissues. All genes with an average RPKM_I > 1 in platelets and expressed in 8 or more nucleated tissues are shown. Blue, genes significantly enriched in platelets; red, genes not significantly enriched in platelets. The data points corresponding to the 5 most enriched genes are indicated. The slope x = y is shown as a dashed line.

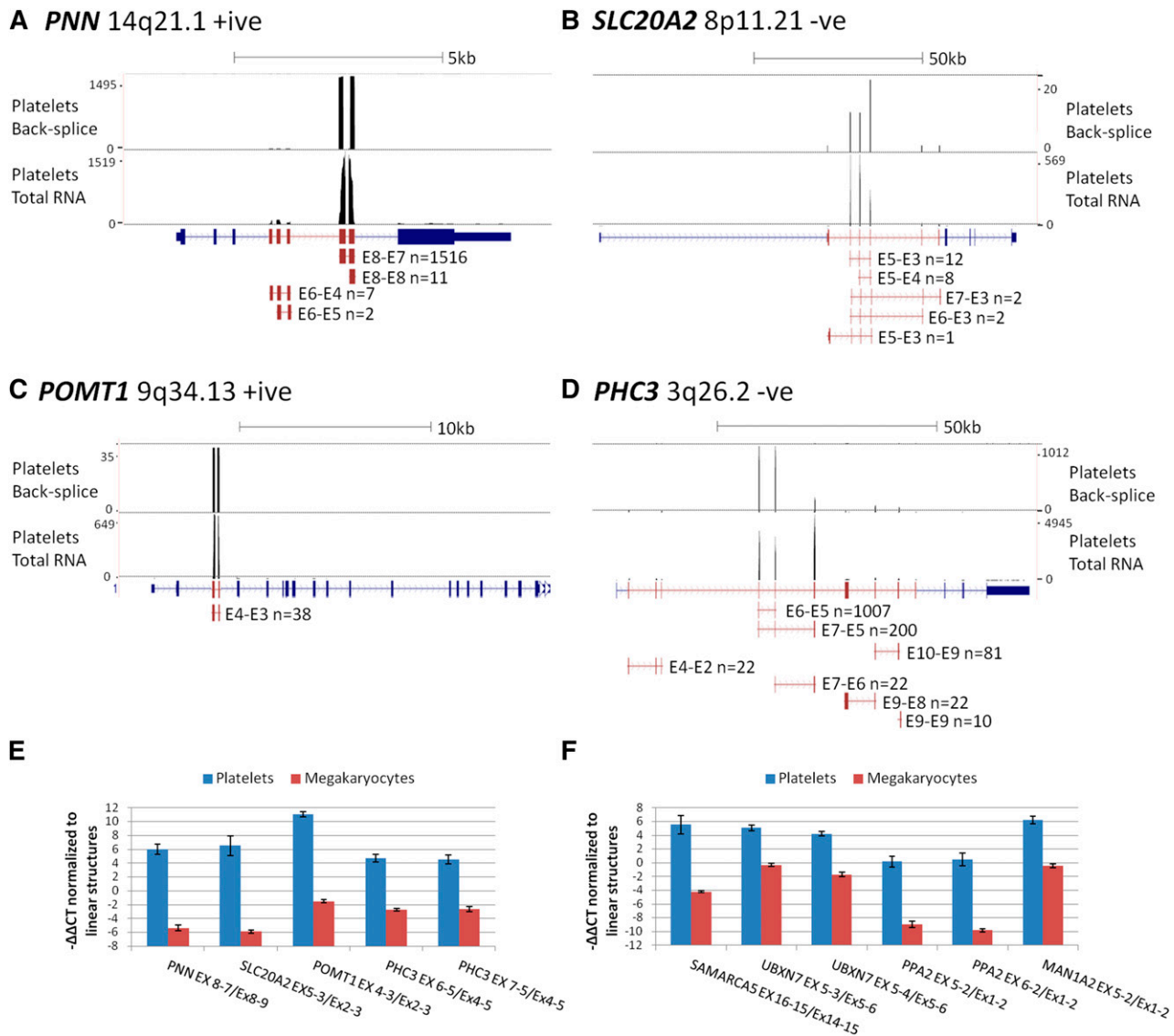


Figure 5. circRNA enrichment occurs in platelets but not megakaryocytes. (A-D) Adapted University of California Santa Cruz screen shots for 4 genes with extreme circRNA enrichment in platelets. Each panel shows RNAseq read abundance in a single platelet total RNA sample (F1), together with the position and abundance of back-splice exon junctions in the same sample. The intron-exon structure is shown below, with exons known to contribute to circRNAs shown in red. CircRNA structures present in the sample together with back-splice frequencies are also shown (for *PHC3*, only structures with ≥ 10 back-splice reads are shown). Above each panel, the gene names, cytogenetic locations, coding strand, and scale in kilobases are indicated. (E-F) Expression levels of circRNAs relative to linear RNAs from the same loci in platelets and cultured megakaryocytes, normalized to the housekeeping pool ($-\Delta\Delta CT$ values).

samples, highlighting the impact this enrichment will have on RNAseq-based estimates of expression levels. We also established that circRNAs are not enriched in cultured megakaryocytes compared with other nucleated cell types.

Although we cannot formally rule out the possibility that some circRNA enrichment occurs at the transcriptional level late in megakaryocyte maturation, collectively our results suggest that enrichment occurs specifically in platelets, after the loss of nuclear influence. The similarly high levels of enrichment observed in nucleated RBCs are consistent with this, whereas the observed reduction of linear RNAs relative to circRNAs in platelets incubated at 37°C confirms that differential decay occurs. The overall level of enrichment of RNAseq reads from circRNA-producing exons is consistent with platelet mRNA levels being $\geq 90\%$ lower than within nucleated tissues (average enrichment, 12.7 times for genes with $RPKM_I > 1$). Although this is a rough and indirect estimate, it is likely to be conservative as it assumes

no degradation of circRNAs. Furthermore, the fraction of translatable RNAs is likely to be even lower, as our analysis has only considered RNA abundance and not structural integrity. This is consistent with the suggestion that platelet lifespan may be determined by translational competence.⁵²

Degradation has been assumed to be a feature of the platelet transcriptome based purely on their anucleate nature.⁵² Mitochondrially encoded genes are among the most highly expressed,⁵³ levels of individual mRNAs have been shown to fall when platelet preparations are incubated at ambient temperature,¹⁶ and there is some evidence that translation is reduced in aged platelets relative to the whole population.⁵² Consideration of mRNA half lives alone would suggest that mRNA degradation is likely to be substantial, unless platelet-specific mechanisms exist to counteract it. If we assume the average age of circulating platelets is 4 days, then mRNA half-lives⁴⁶ would predict gene specific reductions in expression ranging from 1.2 to 720 times,

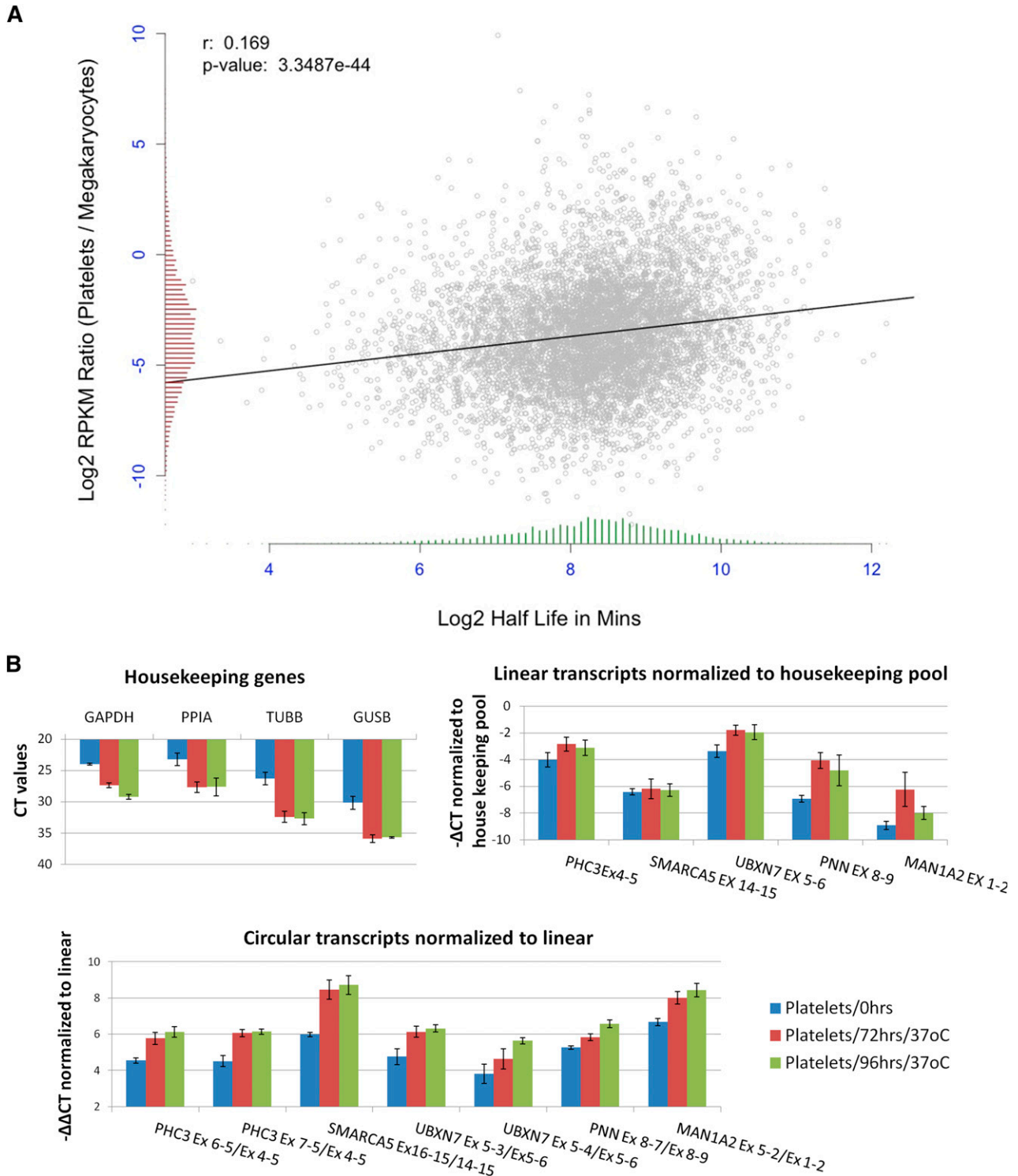


Figure 6. Degradation of platelet RNA. (A) Correlation of the change in gene expression between megakaryocyte polyA⁺ RNA and platelet polyA⁺ RNA (y-axis), and mRNA half-life estimates (x-axis). Distributions of both are shown as histograms on each axis. (B) qPCR analysis of differential decay of linear and circRNAs in platelets following incubation at 37°C for 0, 72, and 96 hours. Data from 3 biological replicates are shown. (Top left) Expression levels of housekeeping genes (CT values). (Top right) Expression levels of linear structures from 5 circRNA-producing genes relative to the housekeeping pool (–ΔCT values). (Bottom) Expression levels of 7 circRNAs relative to linear transcripts from the same loci, both normalized to the housekeeping pool (–ΔΔCT values).

with a mean reduction of 14 times. Our estimate of 12.7 times based on RNAseq read counts is broadly consistent with this, despite the fact that $RPKM_I / (RPKM_I + RPKM_E)$ ratios will vary between genes due to a variety of factors that are unrelated to circularity, including differential isoform abundance, overlapping untranslated regions between

neighboring genes, and the reduction in RNAseq read coverage close to transcript termini.

Extensive mRNA degradation provides a simple explanation for conflicting reports concerning the correlation between the platelet transcriptome and proteome^{12,13,15,54} and for the presence of proteins in

platelets without corresponding mRNAs.^{15,51} It can also explain why many genes shown to be spliced and/or translated in platelets, such as tissue factor (*F3*),^{7,55} *IL1B*,⁵⁶ *SLC23A2*,⁵⁷ *FGA*, *FGB*, *FGG*, *ALB*,⁵⁸ and *BCL3*,⁵⁹ are present at very low levels or not at all in some RNAseq datasets.^{10,11} Collectively, our results indicate that translation in platelets occurs against the backdrop of a rapidly degrading transcriptome, that platelets will differ in their translational potential in an age-dependent manner, and that RNAseq analyses targeted toward intact species that are capped^{15,18} or polysome bound will be required to analyze translational potential in these cells effectively.

circRNAs have been identified from RNAseq data by the back-splice exon junctions they contain, and such junctions have been taken prima facie as evidence of circularity.^{24,33,41,60} Our exon-partitioned analysis of read-depth provides additional in silico evidence of circularity, as *trans*-splicing of independent linear pre-mRNA^{20,43} would affect exons both up- and downstream of the back-spliced exons. The identification of higher numbers of circRNAs in platelets than in total RNA digested with RNase R is, however, surprising and indicates that cessation of transcription in platelets leads to more extreme enrichment of circular RNA through differential decay in vivo than enzymatic digestion of linear species in vitro. This could be due both to limitations of RNase R specificity and the lability of RNA after extraction. The significant enrichment of A nucleotides within back-splice transcripts present within platelet polyA⁺ RNA is also noteworthy as it means their presence within oligo-dT selected RNA fractions does not necessarily imply polyadenylation, as has been suggested previously,^{20,43} and may adversely affect the accuracy of RNAseq-based mRNA expression estimates of the genes involved.

It has recently been suggested that up to 50% of RNAs with back-splice exon junctions are RT artifacts. This was extrapolated from an analysis of 13 back-splice RNAs where the only 6 RNAs that could be independently validated were those that could be amplified using both AMV and MMLV RTs.⁴³ However, among the 7 RNAs that could not be validated, 4 were intragenic back-splice structures, and all are present within platelets: The back-splice counts of 3 (*PHC3* E6-E5, *CNTLN* E5-E3, *RERE* E3-E3) are enriched relative to other tissues (~1000, ~100, and ~4 times), and read counts in exons between each back-splice are also enriched (240, ~30, and 500 times) indicative of circRNAs. Furthermore, the *PHC3* E6-E5 structure degrades more slowly than linear RNA, is resistant to RNase R digestion,²³ and in our hands can be amplified from cDNA generated using either MMLV or AMV (supplemental Figure S3). These diverse lines of evidence suggest that there may not be a straightforward relationship between RT specificity and artifacts as previously suggested.⁴³

Finally, cellular levels of circRNAs are known to be low in proliferating and neoplastic human cells^{44,60} but to increase during differentiation of ES cells²⁶ and during *Drosophila* development,⁶¹ suggesting that they may be a biomarker for aging. Among nucleated tissues, they are highest in both human and *Drosophila* neural tissues^{44,61} and are particularly enriched in synapses.⁴⁴ They are also enriched

in exosomes.⁶² These observations, together with the presence of microRNA binding sites within circRNAs⁶¹ and the dynamic expression patterns of some,⁴⁴ implies that they may be of functional importance. circRNA levels can be modulated by intronic sequence identity^{23,27,28} and splicing factors such as Quaking³¹ and Muscleblind.²⁹ However, our qPCR analysis of transcripts in megakaryocytes suggests that the extreme enrichment in platelets reported here is not due to modulation at the pre-mRNA level within progenitor cells. With no transcription in platelets it can, therefore, only be accounted for by differential degradation/decay. The very high levels of enrichment reported here suggests that, even if the relative levels of linear and circRNAs from each gene remain constant at the level of transcription, subtle changes in expression level, rates of mRNA export, changes in the cell cycle,⁶⁰ changes in nuclear and cytoplasmic volumes, and distance from the nucleus, could all alter relative abundance over time. Many reported differences in linear and circRNA levels could, therefore, be due to differences in stability alone.

Acknowledgments

The financial support from the Leverhulme Trust (grant RPG-2012-795 to M.S.J., M.S.-K., and D.J.E.), Biotechnology and Biological Sciences Research Council (studentship BB/J014516/1 to O.G.I.), and the Wellcome Trust (grants WT089225MA and WT089225/Z/09/Z to D.J.E.) is gratefully acknowledged.

Authorship

Contributions: O.G.I. performed the bioinformatic analyses of RNAseq data, circRNA abundance, circRNA exon enrichment, and RNA decay; A.A.A. performed all cell purification and isolation, RNA extraction, real-time PCR analyses, and platelet time course analyses; H.H.A.-B. developed all real-time PCR assays; J.S.S. provided assistance with the analysis of PolyA⁺ transcripts; J.C.M. and L.M. provided the erythrocytes used to generate RNAseq data; A.E., M.C., and C.G. isolated human haematopoietic progenitors from cord blood, followed by differentiation to megakaryocytes and sorting to purity; M.S.J., M.S.-K., and D.J.E. conceived the project; and M.S.J. drafted the manuscript with input from all authors.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Michael S. Jackson, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom; e-mail: michael.jackson@newcastle.ac.uk.

References

- Deutsch VR, Tomer A. Megakaryocyte development and platelet production. *Br J Haematol*. 2006;134(5):453-466.
- Garraud O, Cognasse F. Are Platelets Cells? And if Yes, are They Immune Cells? *Front Immunol*. 2015;6:70.
- Migliaccio AR. Erythroblast enucleation. *Haematologica*. 2010;95(12):1985-1988.
- Harker LA, Roskos LK, Marzec UM, et al. Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. *Blood*. 2000;95(8):2514-2522.
- Weyrich AS, Schwertz H, Kraiss LW, Zimmerman GA. Protein synthesis by platelets: historical and new perspectives. *J Thromb Haemost*. 2009;7(2):241-246.
- Denis MM, Tolley ND, Bunting M, et al. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell*. 2005;122(3):379-391.
- Schwertz H, Tolley ND, Foulks JM, et al. Signal-dependent splicing of tissue factor pre-mRNA modulates the thrombogenicity of human platelets. *J Exp Med*. 2006;203(11):2433-2440.
- Freedman JE. A platelet transcriptome revolution. *Blood*. 2011;118(14):3760-3761.
- Rowley JW, Oler AJ, Tolley ND, et al. Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. *Blood*. 2011;118(14):e101-e111.

10. Bray PF, McKenzie SE, Edelstein LC, et al. The complex transcriptional landscape of the anucleate human platelet. *BMC Genomics*. 2013; 14:1.
11. Kissopoulou A, Jonasson J, Lindahl TL, Osman A. Next generation sequencing analysis of human platelet PolyA+ mRNAs and rRNA-depleted total RNA. *PLoS One*. 2013;8(12):e81809.
12. Burkhart JM, Vaudel M, Gambaryan S, et al. The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood*. 2012;120(15):e73-e82.
13. Rowley JW, Weyrich AS. Coordinate expression of transcripts and proteins in platelets. *Blood*. 2013;121(26):5255-5256.
14. Geiger J, Burkhart JM, Gambaryan S, Walter U, Sickmann A, Zahedi RP. Response: platelet transcriptome and proteome—relation rather than correlation. *Blood*. 2013;121(26):5257-5258.
15. Londin ER, Hatzimichael E, Loher P, et al. The human platelet: strong transcriptome correlations among individuals associate weakly with the platelet proteome. *Biol Direct*. 2014;9:3.
16. Stiegler G, Fischer G, Ramanathan G, Bencur P, Weigel G, Mannhalter C. P-selectin mRNA is maintained in platelet concentrates stored at 4 degrees C. *Transfusion*. 2009;49(5):921-927.
17. Rowley JW, Schwertz H, Weyrich AS. Platelet mRNA: the meaning behind the message. *Curr Opin Hematol*. 2012;19(5):385-391.
18. Schubert S, Weyrich AS, Rowley JW. A tour through the transcriptional landscape of platelets. *Blood*. 2014;124(4):493-502.
19. Dixon RJ, Eperon IC, Hall L, Samani NJ. A genome-wide survey demonstrates widespread non-linear mRNA in expressed sequences from multiple species. *Nucleic Acids Res*. 2005;33(18):5904-5913.
20. Al-Balool HH, Weber D, Liu Y, et al. Post-transcriptional exon shuffling events in humans can be evolutionarily conserved and abundant. *Genome Res*. 2011;21(11):1788-1799.
21. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One*. 2012;7(2):e30733.
22. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333-338.
23. Jeck WR, Sorrentino JA, Wang K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*. 2013;19(2):141-157.
24. Wang PL, Bao Y, Yee MC, et al. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One*. 2014;9(6):e90859.
25. Caudevilla C, Serra D, Milari A, et al. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc Natl Acad Sci USA*. 1998; 95(21):12185-12190.
26. Wu CS, Yu CY, Chuang CY, et al. Integrative transcriptome sequencing identifies trans-splicing events with important roles in human embryonic stem cell pluripotency. *Genome Res*. 2014;24(1):25-36.
27. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159(1):134-147.
28. Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev*. 2014;28(20):2233-2247.
29. Ashwal-Fluss R, Meyer M, Pamudurti NR, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell*. 2014;56(1):55-66.
30. Starke S, Jost I, Rossbach O, et al. Exon circularization requires canonical splice signals. *Cell Reports*. 2015;10(1):103-111.
31. Conn SJ, Pillman KA, Toubia J, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell*. 2015;160(6):1125-1134.
32. Li Z, Huang C, Bao C, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol*. 2015;22(3):256-264.
33. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol*. 2014;15(7):409.
34. Mukherjee C, Patil DP, Kennedy BA, Bakthavachalu B, Bundschuh R, Schoenberg DR. Identification of cytoplasmic capping targets reveals a role for cap homeostasis in translation and mRNA stability. *Cell Reports*. 2012;2(3):674-684.
35. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol*. 2014;32(5):453-461.
36. Birney E, Stamatoyannopoulos JA, Dutta A, et al; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799-816.
37. Nielsen MM, Tehler D, Vang S, et al. Identification of expressed and conserved human noncoding RNAs. *RNA*. 2014;20(2):236-251.
38. Nürnberg ST, Rendon A, Smethurst PA, et al; HaemGen Consortium. A GWAS sequence variant for platelet volume marks an alternative DNMT3 promoter in megakaryocytes near a MEIS1 binding site. *Blood*. 2012;120(24):4859-4868.
39. Amisten S. A rapid and efficient platelet purification protocol for platelet gene expression studies. *Methods Mol Biol*. 2012;788:155-172.
40. Tijssen MR, Cvejic A, Joshi A, et al. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell*. 2011;20(5):597-609.
41. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet*. 2013;9(9):e1003777.
42. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57-74.
43. Yu CY, Liu HJ, Hung LY, Kuo HC, Chuang TJ. Is an observed non-co-linear RNA product spliced in trans, in cis or just in vitro? *Nucleic Acids Res*. 2014;42(14):9410-9423.
44. Rybak-Wolf A, Stottmeister C, Glažar P, et al. Circular RNAs in the mammalian brain are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell*. 2015;58(5):870-885.
45. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet*. 2012;13(4):246-259.
46. Friedel CC, Dölken L, Ruzsics Z, Koszinowski UH, Zimmer R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res*. 2009;37(17):e115.
47. Rabani M, Levin JZ, Fan L, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol*. 2011;29(5):436-442.
48. Tani H, Akimitsu N. Genome-wide technology for determining RNA stability in mammalian cells: historical perspective and recent advantages based on modified nucleotide labeling. *RNA Biol*. 2012;9(10):1233-1238.
49. Lindemann S, Tolley ND, Dixon DA, et al. Activated platelets mediate inflammatory signaling by regulated interleukin 1beta synthesis. *J Cell Biol*. 2001;154(3):485-490.
50. Zimmerman GA, Weyrich AS. Signal-dependent protein synthesis by activated platelets: new pathways to altered phenotype and function. *Arterioscler Thromb Vasc Biol*. 2008;28(3):s17-s24.
51. Cecchetti L, Tolley ND, Michetti N, Bury L, Weyrich AS, Greslele P. Megakaryocytes differentially sort mRNAs for matrix metalloproteinases and their inhibitors into platelets: a mechanism for regulating synthetic events. *Blood*. 2011;118(7):1903-1911.
52. Steiner M, Baldini M. Protein synthesis in aging blood platelets. *Blood*. 1969;33(4):628-633.
53. Gnatenko DV, Dunn JJ, McCorkle SR, Weissmann D, Perrotta PL, Bahou WF. Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood*. 2003; 101(6):2285-2293.
54. McRedmond JP, Park SD, Reilly DF, et al. Integration of proteomics and genomics in platelets: a profile of platelet proteins and platelet-specific genes. *Mol Cell Proteomics*. 2004;3(2):133-144.
55. Rondina MT, Schwertz H, Harris ES, et al. The septic milieu triggers expression of spliced tissue factor mRNA in human platelets. *J Thromb Haemost*. 2011;9(4):748-758.
56. Hawrylowicz CM, Santoro SA, Platt FM, Unanue ER. Activated platelets express IL-1 activity. *J Immunol*. 1989;143(12):4015-4018.
57. Savini I, Catani MV, Arnone R, et al. Translational control of the ascorbic acid transporter SVCT2 in human platelets. *Free Radic Biol Med*. 2007; 42(5):608-616.
58. Kieffer N, Guichard J, Farcet JP, Vainchenker W, Breton-Gorius J. Biosynthesis of major platelet proteins in human blood platelets. *Eur J Biochem*. 1987;164(1):189-195.
59. Weyrich AS, Dixon DA, Pabla R, et al. Signal-dependent translation of a regulatory protein, Bcl-3, in activated human platelets. *Proc Natl Acad Sci USA*. 1998;95(10):5556-5561.
60. Bachmayr-Heyda A, Reiner AT, Auer K, et al. Correlation of circular RNA abundance with proliferation—exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep*. 2015;5:8057.
61. Westholm JO, Miura P, Olson S, et al. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Reports*. 2014;9(5):1966-1980.
62. Li Y, Zheng Q, Bao C, et al. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res*. 2015; 25(8):981-984.