Exploring transcriptomic landscapes in red blood cells, in their extracellular vesicles and on a single-cell level

Erja Kerkelä

Finnish Red Cross Blood Service

Jenni Lahtela

University of Helsinki, Institute for Molecular Medicine Finland (FIMM)

Antti Larjo

Finnish Red Cross Blood Service

Ulla Impola

Finnish Red Cross Blood Service

Laura Mäenpää

University of Helsinki, Research Programs Unit, Molecular Neurology

Pirkko Mattila

University of Helsinki, Institute of Molecular Medicine Finland (FIMM)

Research article

Keywords: erythrocyte, MALAT1, RNA sequencing, transcriptomics, extracellular vesicle

---

Abstract

Circulating human red blood cells (RBCs) consist of mature erythrocytes and immature reticulocytes. Being anucleated, RBCs lack typical transcriptomes, but are known to contain small amounts of diverse long transcripts and microRNAs. However, the exact role and importance of these RNAs is lacking. Results To study this further, we explored the transcriptomes of RBCs and extracellular vesicles (EVs) of RBCs using next-generation sequencing. Furthermore, to understand the dynamics of the RBC transcriptome, we performed single-cell RNA sequencing on RBCs. An analysis of the single-cell transcriptomes revealed that approximately 10% of the cells contained detectable levels of mRNA and fell into three subpopulations based on their transcriptomes. Decrease in the mRNA quantity was observed across the populations. Qualitative changes included the differences in the globin transcripts and changes in the expression of ribosomal genes. A specific short splice form of a long non-coding RNA, Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1), was the most enriched marker in one subpopulation of RBCs, co-expressing with ribosomal structural transcripts. MALAT1 expression was confirmed by qPCR in CD71-enriched reticulocytes, which were also characterized with imaging flow cytometry at single cell level. Conclusions Analysis of the RBC transcriptome shows enrichment of pathways and functional categories required for the maturation of reticulocytes and erythrocyte functions. The RBC transcriptome was detected in their EVs, proposing vesiculation as a mechanism to remove the cellular contents from RBCs throughout their life cycle. Our experiments on single cell level revealed that lncRNA MALAT1 is the marker for one of the three RBC populations co-expressing with a group of ribosomal protein transcripts.

Background

Human erythrocytes are specialized in providing gas transport throughout the body. Complex genetic networks control hematopoietic stem cell differentiation into progenitors that give rise to billions of erythrocytes daily. Erythroblasts undergo consecutive maturation steps, culminating in enucleation to yield reticulocytes and subsequently, biconcave erythrocytes (Fig. 1A) (1). The overall process through which the reticulocyte is converted to an erythrocyte is not completely understood. Early reticulocyte maturation is characterized by extensive membrane remodeling and the selective removal of unnecessary plasma membrane proteins (e.g., CD71, CD98) through vesiculation via the endosome-exosome pathway (2,3). In contrast, late maturation is characterized by the degradation and elimination of the final cell organelles through autophagy and mitophagy. These processes enclose unwanted materials into autophagosomes and remove them from the cell (4–6).

During terminal erythroid differentiation, globin mRNAs accumulate and comprise over 95% of total cellular mRNA (7). Retained ribosomes and pre-synthesized mRNAs permit continued translation activities in reticulocytes after enucleation, especially that of hemoglobin (Hb) (8–11). Apart from Hb mRNA, reticulocytes have been shown to contain surprisingly diverse transcriptomes. These transcriptomes are hypothesized to encode information regarding terminal erythroid differentiation and maturation (9,11,12). However, the role and regulation of these transcripts in these final stages of

reticulocyte maturation is mostly unknown, even though non-coding RNAs are known to have an impact on erythroid differentiation in general (13).

The remarkable enrichment of globin mRNAs relies on their exceptionally long half-lives and possibly also selective degradation of non-globin mRNAs (8,14). Anucleated red blood cells (RBCs) are unable to transcribe new mRNA. Thus, a progressive decrease in all cellular RNA is observed during their maturation. The disappearance of RNA on a large scale occurs simultaneously with the loss of cellular structures such as the mitochondria and ribosomes (15). Multiple mechanisms of mRNA stability and degradation have been described for α-globin as well as other mRNAs (16,17).

Extracellular vesicles (EVs) are a heterogeneous population of phospholipid membrane particles that may contain functional molecules, such as RNA and have a well-established role in intercellular signaling (18,19). RBCs, together with platelets, are the main EV-secreting cells in the blood (20). In addition to having an important role in reticulocyte maturation, vesiculation is proposed as one mechanism that allows erythrocytes to clear away damaged or harmful molecules and thus prevent their early removal from circulation. Similar vesiculation occurs during storage-related aging (21). The RNA content of EVs is of special interest, especially from the point of view of biomarker discovery (22). To our knowledge, no one has analyzed the transcriptomes of EVs originating from RBCs.

We studied the transcriptomes of single-cell RBCs from both fresh blood and RBC unit. In parallel, we used RNA sequencing to study bulk transcriptomes of RBCs and EVs isolated from RBC units. Our study shows that although the overall RNA content in RBCs is low, subpopulations of RBCs have significant differences in their transcriptomes, which are especially defined by globin transcripts and expression of a long non-coding RNA (lncRNA), namely, Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT1).

Results

Four leukocyte-free RBC units were used to analyze transcriptomes in RBCs and RBC-derived EVs (EV characterization is described in Supplementary Methods and Supplementary Fig. S3). In parallel, RBC transcriptomes from one RBC unit and from one fresh blood sample were analyzed on a single-cell level to explore subpopulations of RBCs (Fig. 1B).

RBC transcriptome is scarce but relatively rich and is transferred to RBC-derived EVs

The transcriptomes of RBCs and RBC-derived EVs were analyzed using NGS. For transcriptomic analysis, the mean amounts of uniquely aligned reads produced for the four RBC samples and four EV samples were 39.6×10^6 and 21.7×10^6, respectively (Supplementary Table S1). The RNAseq data quality was good as judged by body read coverage and principal component analyses (Supplementary Fig. S1A and S1B).

In general, according to the correlation plots, diversity of transcriptomes among the EV group was higher than among the RBC group (Supplementary Fig. S2).

An abundant repertoire of expressed genes was found in all RBC samples (Supplementary Table S2 and S4). 'RBC transcriptome' used for subsequent annotations, as well as functional category enrichment and pathway analyses, consists of 2070 (40%) expressed genes that were shared among all 4 samples (Fig. 2A). Corresponding EV samples contained fewer expressed genes (Supplementary Tables S3 and S5) and were more diverse in the RNA content, as only 686 (15%) of the expressed genes were shared in all samples (Fig. 2B). Most of the transcripts (555 out of 686) in EVs, were also found in RBCs. In addition, most highly expressed genes were shared between RBC and EVs, such as well-known erythrocyte biology-related gene hemoglobin subunit beta (HBB, vast majority of transcripts), ferritin (FTL), hemogen (HEMGN), and 5'-aminolevulinate synthase 2 (ALAS2). The remaining 131 transcripts in EVs (highlighted in yellow in Supplementary Table S3) that were not found in RBCs comprise a probable contamination of RNA from EVs of other cells, mostly platelets, possibly also leukocytes and endothelial cells due to a small residue of plasma in the blood unit production process (maximum 20 ml/RBC unit). In western blot analysis, however, we did not detect platelet marker CD61 (Supplementary Fig. S3). Instead of contamination markers, RBC marker CD235a and Hb were clearly visible in Western blot indicating that most of the EVs are of RBC origin. The lack of leukocyte or platelet transcripts in the RBC data confirmed that the RBC transcriptome is of RBC origin.

Functional category and canonical pathway analyses reveal processes related to RBC maturation and function

Among the most enriched canonical pathways in RBCs were eukaryotic initiation factor 2 (eIF2) signaling and its regulation, mitochondrial dysfunction, protein ubiquitination pathway, oxidative phosphorylation, and mechanistic target of rapamycin (mTOR) signaling (Fig. 3A and Table 1, Supplementary Table S6). The same pathways were also enriched in EVs, of which eIF2 signaling and mitochondria -related pathways had even lower p-values than RBCs (Fig. 3B, Table 1, Supplementary Table S7). Different enrichment in pathways was also evident between RBCs and EVs, such as heme biosynthesis II, NRF2-mediated oxidative stress response- and hypoxia signaling -pathways in RBCs with very significant p-values, while two latter were not statistically significantly enriched in EVs.

The top shared molecular and cellular functions in RBC and EV transcriptomes were related to gene expression, protein synthesis, RNA modification, and cell death and survival (Table 1, Supplementary Table S8 and S9). Under ´physiological system development and function´ –category, ´hematological system development´ and ´hematopoiesis´ were among the top five most enriched categories as expected, with several erythroid cell -related subcategories being most significant (Table 1, Supplementary Table S8 and S9). Other obvious RBC-functions included synthesis and homeostasis of different RBC–related molecules, such as porphyrin, heme and iron (category ´small molecule biochemistry´, p-values 7.61E-08 - 7.3E-05 in RBCs) and free radical scavenging, which was clearly more

significant category in EVs compared to RBCs (Table 1, Supplementary Table S8 and S9). Furthermore, the most significant subcategory under ´cellular function and maintenance´ was autophagy (p-value 1.26E-11 in RBCs), strongly related to reticulocyte maturation.

Suspected contamination present in the EV transcriptome is reflected by the presence of ´inflammatory response´ and ´immune cell trafficking´ -categories and different leukocyte–related subcategories in, e.g., ´hematological system development´ (Supplementary Table S9). Due to contamination the differential expression analysis between RBC and EV transcriptomes was not performed. However, it was still evident that more mitochondrial transcripts were present in the EV transcriptome, as judged by expression intensity and diversity (Supplementary Table S3). That is most likely why mitochondria-related canonical pathways had considerably more significant p-values in EV transcriptome compared to that of RBCs (Table 1).

Sequencing on the single-cell level reveals separate RBC subpopulations defined by differential expression of hemoglobin and the long non-coding RNA, MALAT1

To study the cellular heterogeneity among RBCs, we analyzed fresh and blood unit RBCs using the single-cell transcriptomics method from 10x Genomics (23) aiming at a capture of 2000 cells per sample. A total of 396 RBCs in blood unit and 173 RBCs in fresh blood samples were initially judged as cells based on the total number of detected transcripts. Total number of detected genes per sample was 1142 for blood units and 992 for fresh RBCs. As expected, the number of captured transcripts, detected as unique molecular identifiers (UMIs), varied greatly between individual cells (4-4201 for blood unit and 12-2258 for fresh RBCs, mRNA capture efficiency of our single-cell instrument being 10-15%), and the median number of genes detected per cell was low (5 for blood unit and 15 for fresh RBCs).

A total of 554 RBCs from blood unit and fresh blood was used for final analysis. The genes with 5 or more total transcripts across the cells (158 in total, Supplementary Table S10) included 122 (77%) shared genes with bulk RBC transcriptome (Supplementary Fig. S5). Among these were genes known to be associated with RBC biology such as hemoglobin subunit beta (HBB), hemoglobin subunit alpha 1 and 2 (HBA1/HBA2), ferritin light chain (FTL) and glycophorin C (GYPC). This finding demonstrated the consistency between the RNA sequencing analysis methods used.

The RBCs were further analyzed to identify subpopulations. The cells from blood unit and fresh blood samples clustered in an overlapping manner, supporting our decision to use combined analysis with merged data (Fig. 4A). Clustering of the RBCs based on their transcriptome profiles resulted in three distinct cell clusters (Fig. 4B). Cells in cluster 1 had the highest amount of detected total transcripts (AvgUMI=110), while in clusters 0 and 2 the total transcript counts were clearly lower (AvgUMI=60 and 65, respectively) (Fig. 4C and 4D). This could mean that cells in different clusters may represent RBCs, more specifically reticulocytes, at various maturation stages. The variation measured by the number of genes detected was highest in cluster 1 and 0, while the percentage of mitochondrial genes was highest in

cluster 0 (Fig. 4D). Cells in cluster 1 were characterized by a high expression of RBC-specific genes such as HBB, HBA1, HBA2, FTL and GYPC (p<0.0001 for all) (Fig. 5A-C, Supplementary Table S11), suggesting that these cells most likely represent the population of young reticulocytes, while the cells in clusters 0 and 2 could represent more mature reticulocytes or early erythrocytes.

The cells in cluster 0 were significantly enriched with long non-coding RNA MALAT1 transcripts (p<0.0001) (Fig. 5A-C, Supplementary Table S11). Cluster 0 was also characterized by the low or absent expression of HBB, HBA1 and HBA2, and it was enriched in the expression of mitochondrial and ribosomal genes such as mitochondrially encoded cytochrome c oxidase I and II (MT-CO1 and MT-CO2), ribosomal protein S19 (RPS19) and ribosomal protein S18 (RPS18) (p<0.0001 for all, Fig. 5A, Supplementary Table S11). Additionally, few genes typically expressed in platelets, including beta-2-microglobulin (B2M), thymosin beta 4, X-linked (TMSB4X) and thymosin beta 10 (TMSB10), were detected in cluster 0 cells.

Cells in cluster 2 were characterized by very low total UMI count and expression of HBB, HBA2, and FTL genes (p<0.0001 for all, Fig. 5A, Supplementary Table S11). The single cell analysis also provided further evidence for the purity of the RBC population. In case of contamination with leukocytes, they would have formed a separate cluster characterized by relevant markers.

All reads for MALAT1 obtained from single-cell sequencing map to short transcript forms

According to current annotations in Ensembl Genome Browser, 16 shorter transcripts of MALAT1 are predicted in addition to the well-annotated full length (8708 bp) transcript MALAT1-202: http://www.ensembl.org/Homo_sapiens/Gene/Splice?db=core;g=ENSG00000251562;r=11:65497762-65506516

We used IGV with Ensembl gene annotations to visualize short reads mapping to the MALAT1 gene. Based on the alignment, the sequencing reads did not map to the 3´ end of the full length transcript MALAT1-202. Instead, they mapped to the three short transcripts MALAT1-203, MALAT1-211 and MALAT1-217 (Fig. 6). This finding agreed with the bulk RNAseq data. According to our knowledge this is the first report of the predicted splice site shared by the transcripts MALAT1-203, MALAT1-211 and MALAT1-217 as an exon 1 start site.

Characterization of CD71-enriched reticulocytes

With imaging flow cytometry we wanted to study the dynamics of RNA content in enriched reticulocytes. Double positive events for CD71 and RNA-binding dye thiazole orange (TO) were gated as CD71^{high}, CD71^{medium} and CD71$^{low-negative}$ populations based on intensity values on each channel and images. Of those cells, 35% expressed CD71 at high level, 33 % at medium level and 17% had low or negative

expression (Fig. 7A). Although high CD71 expression generally correlated with high TO intensity, especially in CD71^{high} population, there were also cells with high CD71 expression and low TO intensity, and vice versa (Fig. 7B and C) indicating high diversity in the amount of RNA in reticulocyte subpopulations. CD71-enriched cells had varying morphologies and concave shaped cells having high CD71 expression and TO intensity were also observed (Fig. 7C). Also RBC marker CD235a expression was detected with varying levels of intensity (Fig. 7C), which, however, did not correlate with CD71 expression or TO intensity (data not shown).

The bulk expression of MALAT1 lncRNA in enriched reticulocytes was confirmed with quantitative real time PCR (qPCR). As expected, the expression of HBB was high in CD71-enriched cells with the Cq values ranging from 15.7 to 20.2 (mean 17.5, SD 1.9), while MALAT1 had lower but clearly detectable expression with Cq values (mean 32.9, SD 2.2) (Fig. 7D).

We did not detect any contamination from other blood cell types, i.e., the expression of CD41a indicating platelets and CD45 for leukocytes was not detected (data not shown), which confirms that MALAT1 expression originated from reticulocytes.

Discussion

To understand the final stages of RBC maturation, we studied the transcriptomic landscapes of RBCs and their EVs using traditional RNAseq and single-cell RNAseq. Based on our single-cell analysis, there are significant differences in the transcriptomes of three RBC subpopulations. Two distinct subpopulations are defined by the expression of globin genes, and one is defined by the expression of the lncRNA, MALAT1, also expressed in bulk RNAseq data and verified by qPCR.

Indeed, lncRNAs in general and MALAT1 are dynamically expressed and regulated during erythropoiesis (13,24). Their depletion severely impaired erythrocyte maturation by inhibiting enucleation (25). Only some cells in the MALAT1 population contained globin transcripts. Instead, this population was enriched in transcripts coding for ribosomal and mitochondrial structural proteins. We assume that these subpopulations reflect the different stages of reticulocyte maturation and that other transcripts in the MALAT1 -population can only be recovered in the late stage of reticulocyte differentiation when not masked by the bulk globin transcripts and not being bound to organelles due to ceased biogenesis of ribosomes and mitochondria.

During their terminal differentiation in peripheral blood, anucleated reticulocytes adjust their protein expression profiles post-transcriptionally. Although not yet studied in RBCs, there is rapidly accumulating evidence for translational regulation executed by ribosome-associated lncRNAs (26,27). Our data support this, as we found co-expression of MALAT1 transcripts and a plethora of mRNAs coding for ribosomal structural proteins (Fig. 5A) in the same subpopulation of RBCs. Recent footprinting studies have shown that the ribosome is the primary destination for the majority of cytoplasmic lncRNAs (28). In addition, MALAT1 has also been shown to act as a miRNA sponge and as a potent autophagy inducer (29).

Reticulocytes are known to employ miRNA-based mechanisms that could have a role in both protein translation control and degradation of mRNA during maturation (30–32). Intriguingly, MALAT1 hits from our single cell data all mapped to the same region, showing predictions of a short, cytosolic previously unknown transcript forms of MALAT1, instead of nuclear-retained, known long form of MALAT1. Since it is poorly polyadenylated (33), the full length MALAT1 transcript is unlikely captured in our poly-A-based single-cell assay in RBCs.

Obvious decrease in the mRNA quantity was observed across the subpopulations in single cell RNAseq data. To support these findings, we employed imaging cell cytometry to correlate cellular RNA content to the expression of reticulocyte marker CD71. Decreasing trend in CD71 expression and intensity of TO staining was observed in CD71-enriched reticulocytes, which is known to indicate reticulocyte maturation (34,35). However, TO intensity and thus cellular RNA amount, was surprisingly heterogeneous, especially within CD71^{medium} and CD71^{low} populations. Furthermore, we observed varying morphologies across the populations, i.e., also concaved cells having high CD71 and TO intensity. This is partially in contrast to the study by Malleret et $al.$ who concluded that CD71^{high} and CD71^{medium} reticulocytes share a similar gross morphology (large and multilobular) when compared to the smaller and increasingly concave reticulocytes seen in the CD71^{low} and CD71^{neg} populations (35), however, their imaging method was different from ours.

In our bulk RBC transcriptomes, both the number of expressed genes and the selection of highly expressed genes match well to the previously published ones (9,11,12) despite the different experimental set up used. The RBC and EV transcriptomes reflect the processes needed both in the reticulocyte maturation as well as for mature erythrocyte functions. During the two days in circulation, reticulocytes need to balance the production of Hb and final termination of protein translation processes. The enrichment of both protein synthesis -category and the mTOR pathway in our data reflect the uniquely high demand for Hb protein translation, supported by the finding that inhibition of mTOR resulted in reduced maturation of reticulocytes (36). Additionally, the enrichment of eIF2 signaling indicates active protein translation. In contrast, the enriched ubiquitin proteasome pathway and ubiquitination -category, also found to be enriched by Doss et $al.$ (12), points to the targeted degradation of unnecessary proteins in reticulocytes (37) and ubiquitination is an essential step both in enucleation (38) and in mitochondrial autophagy (39). Interestingly, a recent proteomics study by Chu et $al.$ comparing CD71-positive and -negative RBCs revealed the exactly same set of canonical pathways upregulated during reticulocyte maturation than our transcriptomic level data (40).

Autophagy was highly enriched process in RBC transcriptome and is crucial for the removal of mitochondria and ribosomes at the final maturation stage (41). Mitophagy can occur in reticulocytes through the autophagy-related gene (ATG) protein pathway (41) and through ATG-independent mechanism involving BCL2 interacting protein 3 like (BNIP3L)(42,43), which was one of the highly expressed transcripts in RBCs (bulk and single-cell data) and EVs, supporting the finding of Doss et $al.$ (12). Players from both of these pathways were present in our data, except for the lysosomal compartment. According to the theory of Griffits and co-workers, autophagosomes could be combined

with glycophorin A-coated vesicles, instead of lysosomes, and subsequently be removed by exocytosis (44). Our observations support this theory of organelle removal in EVs, as the RBC transcriptome was observed in EVs and especially enriched in mitochondrial transcripts and functions (oxidative phosphorylation, mitochondrial dysfunction), as well as transcripts for ribosomal proteins and translational machinery (eIF2 signaling). Whether RNA molecules are passively captured to EVs along with removal of, e.g., unwanted proteins and organelles or if there might be some active selection, remains to be explored.

EV transcriptome contained a group of expressed genes not found in RBCs. They most likely originate from EVs of residual plasma in the RBC product resulting from the manufacturing process. Consequently, there is inherent biological variation between EV transcriptomes arising from small but variable amounts of plasma EVs. However, the overall EV transcriptome is similar to that of RBCs with RBC function -related transcripts expressed at high level compared to contaminating ones indicating that vesiculation is truly a way to remove cellular components also from mature RBCs.

The most highly expressed transcript found in EVs but not in bulk RBC transcriptome, TMSB4X, was also found in some cells of MALAT1 -subpopulation and is previously associated with platelets (45). However, it is unlikely that we have platelet contamination in single cell data, since the same cells occasionally contained Hb and other RBC markers. In addition, cells were collected and purified in the same way for the bulk data, where platelet contamination was not detected. These transcripts could either have a very low expression in RBCs, hidden in the bulk data and only become detectable in single-cell data in the cells that have lower Hb expression. If TMSB4X and TMSB10 transcripts are truly of platelet origin, one hypothesis is that platelet-EVs may have adhered to or even fused with (46) RBCs, thereby becoming a part of the RBC transcriptome. The transfer of transcripts via EVs is a well-known phenomenon (18).

Conclusion

Our study shows that RBCs still have a versatile transcriptome that can be utilized to produce not only Hb but also proteins needed in the final stages of reticulocyte maturation and for mature erythrocyte functions. We show for the first time that the transcriptome of RBCs is transferred to EVs, which makes these transcripts available for intercellular communication in the blood. In addition, RBC-EV transcriptome provides important background information for any plasma EV RNAseq studies and possible biomarker discovery.

Remarkably, we were able to determine the RBC transcriptome at the single-cell level, uncovering entirely new information on the final stages of RBC maturation. Single-cell sequencing of RBCs showed that there is heterogeneity among RBCs clustering in three different populations according to their transcriptomes. This data fits well to the accumulating understanding about the dynamic nature of reticulocyte maturation, making it unnecessary to strictly classify reticulocytes as one cell population. Our study revealed that the recently discovered lncRNA MALAT1 is the marker for one of the RBC populations co-expressing with a group of ribosomal protein transcripts. Our findings suggest a ribosomal location and a

novel role for MALAT1 in translational control in reticulocytes. Moreover, we show that the relevant transcripts are not the full-length isoform MALAT1-202 but short, alternative transcripts currently only known as predictions. Functional studies are needed to verify the exact role of MALAT1 in maturing reticulocytes.

Methods

Isolation of RBCs, extracellular vesicles and RNA

Red blood cell (RBC) samples were collected at day 8 for bulk sequencing (n=4) from carefully mixed leukoreduced RBC units (B+), stored in the saline-adenine-glucose-mannitol (SAGM) solution with citrate-phosphate-dextrose (CPD) anticoagulant at 2–6 °C under standard blood bank conditions (obtained from the Finnish Red Cross Blood Service, FRCBS) (experimental setup described in Fig 1 Two of the donors were female (aged 49 and 28) and two were male (aged 49 and 43). To remove any residual leukocytes, 10 ml of RBC unit was diluted with 10 ml of phosphate-buffered saline (PBS) and run through density gradient centrifugation (Ficoll-Paque plus, GE Healthcare).

EVs were collected from the same RBC units. RBC concentrate sample (15 ml) was diluted with 15 ml of PBS and centrifuged first 20 min at 805 x g and subsequently 20 min 3000 x g without brake to remove the cells (Eppendorf centrifuge 5810R). The resulting supernatant was ultracentrifuged for 1 h at 100 000 x g (+4∘C, fixed angle MLA-50 rotor, Beckman Coulter) to obtain EVs. Next, EVs were washed with PBS, ultracentrifuged for 1 h at 100 000 x g (+4∘C, swing-out MLS-50 rotor, Beckman Coulter). Characterization of EVs obtained with this method has been described previously (47) and is shown in Supplementary methods and Supplementary Fig. S2.

For single-cell sequencing, one fresh blood sample (donor: female, 43 years) and one blood unit sample (donor male: 39 years), as comparable to fresh blood sample as possible (collected at day 2) were processed. RBCs from fresh blood sample were collected similarly as from blood unit sample and additionally purified with an Acrodisc WBC filter (Pall Life Sciences).

Total RNA from RBCs (200µl, ~ 1x10^9 cells) and EVs was isolated by using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's protocol.

Preparation of RNAseq -libraries and sequencing

To prepare RNAseq libraries, 10 ng (RBCs) or 0.1-1 ng (EVs) of total input RNA was used with the SMART-seq v4 Ultra low input RNA kit (Clontech). This method relies on poly-dT –based cDNA synthesis, which enabled us to avoid rRNA inserts in the libraries. RNAseq libraries were sequenced using the Illumina HiSeq 2000 system with v4 chemistry and 100 bp paired-end reads. Targeted sequencing depth was 20 x 10^6 and 40 x 10^6 paired end reads for EV and RBC samples, respectively.

Analysis of RNAseq data

R 3.4.0 was used to identify patterns in the RNAseq -data with Principal component analysis (PCA) and Scatter plot function of Microsoft Excel was used to calculate correlation coefficients for data pairs originating from RBC and EV samples.

For data analysis, fastq data was trimmed with the Trimmomatic tool to remove Illumina platform-specific adapters. STAR –aligner was used to map the trimmed reads to the GRCh37r72 version of human genome. Only reads uniquely mapped to genome and not mapped to ribosomal RNA species were utilized for HTseq tool to count reads mapping to genes. The RSeQC-tool package was used to determine the quality of RNAseq data.

Estimates for transcript abundances, reported as transcripts per million (TPM), were produced using the Kallisto algorithm (48) and further analyzed using the Sleuth software (49). The final gene lists (minimum 5 counts/gene) from individual RBC and EV samples were compared using the Venny 2.1.0 tool (50) with Ensembl gene IDs.

Gene and transcript annotations, pathway analysis, and functional category enrichment

Expression data was annotated, and the core expression analysis was performed using QIAGEN's Ingenuity® Pathway Analysis (IPA, QIAGEN, release Summer, 2018, Ensembl Human: GRCh38.86). Statistically significant enrichment of genes was examined on canonical pathways, molecular and cellular functions and physiological system development and function -categories. Disease -related categories were omitted, except for the canonical pathway enrichment. On canonical pathways, all pathways having p-value <0.05 were listed. On functional categories, default settings were used, which resulted in the lists of categories with p values £0.00455 for EVs and £ 0.000168 for RBCs. The significance values were calculated by Fisher's exact test right-tailed. Ensembl IDs that could not be annotated with IPA were annotated using the Ensembl genome browser (release 88-Mar 2017).

Single RBC analysis using the 10 x Genomics platform

For single-cell analysis, RBCs were washed with 0.04% bovine serum albumin and 1 x PBS (Thermo Fisher Scientific), filtered using a Flowmi Cell Strainer (40 µm, Bel-Art), counted with a LUNA Automated Cell Counter (Logos Biosystems) and loaded into a Chromium Single-Cell Chip v2 (10x Genomics) for Chromium Controller gel beads in emulsion (GEMs) generation. The GEM generation, cDNA purification and amplification were performed according to the manufacturer's instructions using a Chromium Single-Cell 3' v2 Reagent Kit with a 2000 target cell capture per sample. Library construction was performed

according to the manufacturer's instructions while scaling down the reaction. Sample libraries were sequenced in Illumina HiSeq 2500 using HiSeq Rapid SBS Kit v2 and HiSeq PE Rapid Cluster Kit v2 (Illumina) according to the instructions by 10XGenomics.

The Cell Ranger v1.3 mkfastq and count analysis pipelines (10x Genomics) were used to demultiplex and convert Chromium single-cell 3' RNA-sequencing barcodes and read data to FASTQ files and to generate aligned reads and gene-cell matrices. Blood unit and fresh RBC samples were analyzed as one sample. The Cell Ranger v1.3 aggr pipeline was used to produce a combined single gene-cell barcode matrix using normalization to the same sequencing depth. The gene-barcode matrices were analyzed using a Seurat 2.0.1 package (51,52) with R 3.4.0. First, the data were normalized to 10,000 molecules, and log-scaling was performed as described (51). Genes detected in only one cell were excluded from the analysis. No further filtering was performed based on total unique molecular identifiers (UMIs) or the percentage of mitochondrial genes (Supplementary Fig. S4A). For the remaining cells, principal component analysis (PCA) was performed using variable genes. First, ten principal components were used for clustering (resolution 0.3) and for t-distributed stochastic neighbor embedding (t-SNE), allowing duplicates. Positive cluster marker genes were identified by comparing clusters using genes that were differentially expressed in at least 5% of the cluster's cells and that showed at least a 0.10 log-scale fold difference on average between clusters. Genes with a p-value lower than 0.001 were considered significant.

The final gene list of the combined single gene-cell barcode matrix generated from fresh blood and blood unit was compared to bulk RNAseq data using Ensembl gene IDs and the Venny 2.1.0 tool. The comparison was restricted to genes with 5 or more total UMIs across the cells.

Visualization of single-cell RNAseq data for MALAT1

An Integrative Genomics Viewer (IGV(53)) was used to visualize the aligned sequencing data (.bam files) with Ensembl Release 80 annotations, http://software.broadinstitute.org/software/igv/.

Enrichment of reticulocytes

In order to obtain enough RNA for qPCR, reticulocytes were enriched using CD71-conjugated microbeads (Miltenyi Biotec) according to the manufacturer's protocol. First, RBCs (n=2 from blood unit, n=3 from fresh blood) were purified as described above using density gradient centrifugation and Acrodisc WBC filter (for fresh blood). Approximately 1×10^9 RBCs were suspended in 800 µl of buffer (2 mM EDTA-PBS containing 0.5 % BSA) and labelled with 150 µl CD71 microbeads. After 15 min incubation at room temperature, 10 ml buffer was added and samples were centrifuged at 600 x g. Samples were suspended in 800 µl of buffer and magnetic separation was done using MACS LS columns. After washing and elution, cells were counted using Bürker chamber.

Analyses on the ImageStreamX MkII

To characterize the CD71-enriched samples (n = 5), the cells were labelled with CD71-PE and CD235a-Pacific blue antibodies (Biolegend), as well as thiazole orange (TO) for RNA (Sigma, 0.01 ng/ml in PBS, stock diluted in DMSO) for 45 min at RT. Additionally, to check for platelet or leukocyte contamination, two samples were labelled with CD41a-FITC (BD Pharmingen) and CD45-PerCp-Cyanine 5.5 (eBioscience), respectively. At least 1000 events for each sample were acquired using a 12 channel Amnis ImageStream^X Mark II (EMD Millipore) imaging flow cytometer. Samples were acquired at 60 × magnification with low flow rate/high sensitivity. The integrated software INSPIRE (EMD Millipore) was used for data collection. The instrument and INSPIRE software were set up as follows: Excitation lasers 405, 488, 642 and 785 and channels Ch1 and Ch9 (bright field, BF), Ch6 (scattering channel), plus fluorescence channels Ch3, Ch5, Ch7 and Ch11 were activated for signal detection. Single color controls were used for compensation. Compensated data files were analyzed using image-based algorithms available in the IDEAS statistical analysis software package. Single cells were separated from debris and doublets using a bivariate plot of aspect ratio vs area of the BF image. Double positive events for CD71-PE and TO were gated as CD71^{high}, CD71^{medium} and CD71$^{low-negative}$ populations based on intensity values on each channel and images. In parallel, all samples were run with BD FACSAria IIU (BD Biosciences) flow cytometer using FACSDiva version 8.0.1 software (BD Biosciences) and analyzed with FlowJo version 10.0.7 software.

Quantitative real-time PCR

Quantitative real time PCR (qPCR) was performed on CD71-enriched cells (n = 5). For cDNA synthesis, approximately 10^5 cells were lysed and DNase I -treated using Taqman Cells-to-CT kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Maximum amount of cell lysate (9 µl) per reaction was directly used for real-time qPCR with MALAT1 and HBB Taqman assays. MALAT1 Taqman assay was custom made against the sequence region found to be expressed in RNAseq analyses (Fig. 6). Each reaction was performed in duplicate and the signal was detected by ABI 7900HT real-time PCR machine. No template and −RT controls were included in the experiment. RNA could not be quantified and thus, MALAT1 and HBB expression levels are shown as quantitation cycle (Cq) values.

List Of Abbreviations

Cq: quantitation cycle

Hb: hemoglobin

EV: extracellular vesicle

lncRNA: long non coding RNA

MALAT1: Metastasis Associated Lung Adenocarcinoma Transcript 1

RBC: red blood cell

TMP: transcripts per million

TO: thiazole orange

t-SNE: t-distributed stochastic neighbor embedding

UMI: unique molecular identifier

Declarations

Ethical considerations

Only surplus RBC units, not needed for clinical use due to storage excess were used in the study. Blood donors are informed about the research use of excess blood units. Blood samples were treated anonymously and obtained from healthy volunteers who had given their informed consent. The research was in accordance with the rules of the Finnish Supervisory Authority for Welfare and Health (Valvira, Helsinki, Finland). Research permission was obtained from the local Blood Service Board (Finnish Red Cross Blood Service, Finland).

Consent for publication

Not applicable.

Availability of data and materials

All sequencing data were deposited in NCBI's Gene Expression Omnibus (54), and they are accessible through the following GEO Series accession numbers: bulk RNAseq -data GSE108378 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108378), and single-cell RNAseq data GSE108241 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108241).

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Finnish Funding Agency for Innovation (TEKES) as part of the SalWe research program Personalized Diagnostics and Care (GET IT DONE, grant No. 3986/31/2013).

Authorship Contributions

EK and PM designed the study. EK, JL, and PM planned and carried out the experiments. EK and UI analyzed the imaging data, JL, AL, and LM analyzed the sequencing data. EK, JL and PM interpreted the results. EK, JL and PM wrote the manuscript.

Acknowledgments

We would like to thank Birgitta Rantala, Lotta Sankkila, Lotta Andersson, Sami Valkonen, Anne Vaittinen and Reija Soukka for their excellent technical assistance. Anne Valkeajärvi and Juha Eronen are acknowledged for their expertise in RBC unit production. We thank the FIMM Sequencing unit supported by Biocenter Finland for their contribution in the sequencing of the RBC samples.

References

1. Hattangadi SM, Wong P, Zhang L, Flygare J, Lodish HF. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. Blood. 2011;118(24):6258–68.

2. Johnstone R, Adam M, Hammond J, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). J Biol Chem. 1987;262(19):9412–20.

3. Johnstone RM, Bianchini A, Teng K. Reticulocyte maturation and exosome release: transferrin receptor containing exosomes shows multiple plasma membrane functions. Blood. 1989;74(5):1844–51.

4. Gasko O, Danon D. Endocytosis and Exocytosis in Membrane Remodelling during Reticulocyte Maturation. Br J Haematol. 1974;28(4):463–70.

5. Heynen MJ, Verwilghen RL. A quantitative ultrastructural study of normal rat erythroblasts and reticulocytes. Cell Tissue Res. 1982;224(2):397–408.

6. Ney PA. Normal and disordered reticulocyte maturation. Curr Opin Hematol. 2011;18(3):152–7.

7. Bastos RN, Volloch Z, Aviv H. Messenger RNA population analysis during erythroid differentiation: a kinetical approach. J Mol Biol. 1977;110(2):191–203.

8. Lodish HF, Small B. Different Lifetimes of Reticulocyte Messenger RNA. Cell. 1976;7(Jan):59–65.

9. Goh SH, Lee YT, Bouffard GG, Miller JL. Hembase: browser and genome portal for hematology and erythroid biology. Nucleic Acids Res. 2004;32:D572-4.

10. Bonafoux B, Lejeune M, Piquemal D, Quere R, Baudet A, Assaf L, et al. Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage. Haematologica. 2004;89(12):1434–8.

11. Goh SH, Josleyn M, Lee YT, Danner RL, Gherman RB, Cam MC, et al. The human reticulocyte transcriptome. Physiol Genomics. 2007;30(2):172–8.

12. Doss JF, Corcoran DL, Jima DD, Telen MJ, Dave SS, Chi JT. A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. BMC Genomics. 2015;16:952.

13. Paralkar VR, Weiss MJ. Blood Spotlight Long noncoding RNAs in biology and hematopoiesis. Blood. 2014;121(24):4842–7.

14. Aviv H, Voloch Z, Bastos R, Levy S. Biosynthesis and stability of globin mRNA in cultured erythroleukemic friend cells. Cell. 1976;8(4):495–503.

15. Heynen M. Ultrastructural Changes during Erythroid Development. In: Harris JR, editor. Blood Cell Biochemistry. Springer Science+Business Media, LLC; 1990. p. 1–26.

16. Waggoner S, Liebhaber S. Regulation of alpha-Globin mRNA Stability. Exp Biol Med. 2003;228(4):387–95.

17. Schoenberg DR. Mechanisms of endonuclease-mediated mRNA decay. Wiley Interdiscip Rev RNA. 2011;2(4):582–600.

18. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654–9.

19. Yáñez-Mó M, Siljander PR-M, Andreu Z, Bedina Zavec A, Borràs FE, Buzas EI, et al. Biological properties of extracellular vesicles and their physiological functions. J Extracell Vesicles. 2015;4(1):27066.

20. Berckmans RJ, Nieuwland R, Boing AN, Romijn FP, Hack CE, Sturk A. Cell-derived microparticles circulate in healthy humans and support low grade thrombin generation. Thromb Haemost. 2001;85(4):639–46.

21. Antonelou MH, Seghatchian J. Update on extracellular vesicles inside red blood cell storage units: Adjust the sails closer to the new wind. Transfus Apher Sci. 2016;55(1):92–104.

22. Hill AF, Pegtel DM, Lambertz U, Leonardi T, O'Driscoll L, Pluchino S, et al. ISEV position paper: extracellular vesicle RNA analysis and bioinformatics. J Extracell Vesicles. 2013;2:eCollection.

23. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:1–12.

24. Ma X-Y, Wang J-H, Wang J-L, Ma CX, Wang X-C, Liu F-S. Malat1 as an evolutionarily conserved lncRNA, plays a positive role in regulating proliferation and maintaining undifferentiated status of early-stage hematopoietic cells. BMC Genomics. 2015;16(1):676.

25. Alvarez-Dminguez JR, Hu W, Yuan B, Shi J, Park SS, Gromatzky AA, et al. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. Blood. 2014;123(4):570–81.

26. Pircher A, Gebetsberger J, Polacek N. Ribosome-associated ncRNAs: An emerging class of translation regulators. RNA Biol. 2014 Nov 2;11(11):1335–9.

27. Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. Proc Natl Acad Sci. 2017;114(46):E10018–27.

28. Carlevaro-Fita J, Rahim A, Vardy LA, Johnson R, Guigó R, Vardy LA, et al. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. RNA. 2016;22(6):1–16.

29. Li Z, Li J, Tang N. Long noncoding RNA Malat1 is a potent autophagy inducer protecting brain microvascular endothelial cells against oxygen-glucose deprivation/reoxygenation-induced injury by sponging miR-26b and upregulating ULK2 expression. Neuroscience. 2017;354:1–10.

30. Chen SY, Wang Y, Telen MJ, Chi JT. The genomic analysis of erythrocyte microRNA expression in sickle cell diseases. PLoS One. 2008;3(6):e2360.

31. Azzouzi I, Moest H, Winkler J, Fauchere JC, Gerber AP, Wollscheid B, et al. MicroRNA-96 directly inhibits gamma-globin expression in human erythropoiesis. PLoS One. 2011;6(7):e22838.

32. Azzouzi I, Moest H, Wollscheid B, Schmugge M, Eekels JJ, Speer O. Deep sequencing and proteomic analysis of the microRNA-induced silencing complex in human red blood cells. Exp Hematol. 2015;43(5):382–92.

33. Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell. 2008;135(5):919–32.

34. Serke S, Huhn D. Identification of CD71 (transferrin receptor) expressing erythrocytes by multiparameter-flow-cytometry (MP-FCM): correlation to the quantitation of reticulocytes as determined by conventional microscopy and by MP-FCM using a RNA-staining dye. Br J Haematol. 1992;81(3):432–9.

35. Malleret B, Xu F, Mohandas N, Suwanarusk R, Chu C, Leite JA, et al. Significant biochemical, biophysical and metabolic diversity in circulating human cord blood reticulocytes. PLoS One. 2013;8(10):e76062.

36. Knight ZA, Schmidt SF, Birsoy K, Tan K, Friedman JM. A critical role for mTORC1 in erythropoiesis and anemia. Elife. 2014;Sep 8;3:e0.

37. Liu J, Guo X, Mohandas N, Chasis JA, An X. Membrane remodeling during reticulocyte maturation. Blood. 2010;115(10):2021–7.

38. Thom C, Traxler E, Khandros E, Nickas JM. Trim58 degrades dynein and and regulates terminal eryhtropoiesis. Dev Cell. 2014;30(6):688–700.

39. Ney PA. Mitochondrial autophagy: Origins, significance, and role of BNIP3 and NIX. Biochim Biophys Acta - Mol Cell Res. 2015;1853(10):2775–83.

40. Chu T, Sinha A, Malleret B, Suwanarusk R, Park JE, Naidu R, et al. Quantitative mass spectrometry of human reticulocytes reveal proteome-wide modifications during maturation. Br J Haematol. 2018;180(1):118–33.

41. Kundu M, Lindsten T, Yang CY, Wu J, Zhao F, Zhang J, et al. Ulk1 plays a critical role in the autophagic clearance of mitochondria and ribosomes during reticulocyte maturation. Blood. 2008;112(4):1493–502.

42. Sandoval H, Thiagarajan P, Dasgupta SK, Schumacher A, Prchal T, Chen M, et al. Essential role for Nix in autophagic maturation of erythroid cells. Nature. 2009;454(7201):232–5.

43. Zhang J, Loyd MR, Randall MS, Waddell MB, Kriwacki RW, Ney PA. A short linear motif in BNIP3L (NIX) mediates mitochondrial clearance in reticulocytes. Autophagy. 2012;8(9):1325–32.

44. Griffiths RE, Kupzig S, Cogan N, Mankelow TJ, Betin VMS, Trakarnsanga K, et al. Maturing reticulocytes internalize plasma membrane in glycophorin A-containing vesicles that fuse with autophagosomes before exocytosis. Blood. 2012;119(26):6296–306.

45. Rowley JW, Oler AJ, Tolley ND, Hunter BN, Low EN, Nix DA, et al. Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. Blood. 2011;118(14):e101-11.

46. Parolini I, Federici C, Raggi C, Lugini L, Palleschi S, Milito A De, et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. J Biol Chem. 2009;284(49):34211–22.

47. Laurén E, Tigistu-Sahle F, Valkonen S, Westberg M, Valkeajärvi A, Eronen J, et al. Phospholipid composition of packed red blood cells and that of extracellular vesicles show a high resemblance and stability during storage. Biochim Biophys Acta. 2018;1863(1):1–8.

48. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525–7.

49. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14(7):687–90.

50. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007. http://bioinfogp.cnb.csic.es/tools/venny/index.html.

51. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2017;161(5):1202–14.

52. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495–502.

53. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2012;14(2):178–92.

54. Edgar R, Domrachev M, Lash A. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.

Table

Table 1. The most enriched pathways and functions in red blood cell (RBC) and RBC-derived extracellular vesicle (EV) transcriptomes. The top five categories are shown in each group, as well as p-values or p-value ranges.

RBC transcriptome

Canonical Pathways	p-value
EIF2 Signaling	2,53E-54
Regulation of eIF4 and p70S6K Signaling	1,65E-24
Mitochondrial Dysfunction	1,05E-18
Protein Ubiquitination pathway	2,77E-17
Oxidative Phosphorylation	1,78E-16

Molecular and Cellular Functions	p-value range
Gene expression	5,17E-31 - 2,57E-07
Protein synthesis	4,81E-30 - 7,75E-05
Cell Death and Survival	1,47E-25 - 1,68E-04
RNA Post-Transcriptional Modification	7,82E-25 - 2,78E-05
Cell Cycle	1,86E-16 - 1,43E-04

Physiological System Development and Function	p-value range
Organismal Survival	4,93E-14 - 4,39E-06
Connective Tissue Development and Function	4,00E-09 - 5,15E-05
Hematological System Development and Function	4,25E-09 - 4,53E-05
Hematopoiesis	4,25E-09 - 4,53E-05
Tissue Morphology	4,25E-09 - 4,53E-05

EV transcriptome

Canonical Pathways	p-value
EIF2 Signaling	1.14E-64
Oxidative Phosphorylation	3,21E-25
Mitochondrial Dysfunction	1,60E-24
Regulation of eIF4 and p70S6K Signaling	2,90E-18
mTOR Signaling	1,68E-13

Molecular and Cellular Functions	p-value range
Cell Death and Survival	2,92E-31 - 4,49E-03
Protein synthesis	8,21E-24 - 4,27E-03
Gene expression	3,53E-19 - 2,95E-03
RNA Post-Transcriptional Modification	4,11E-13 - 7,87E-04
Free Radical Scavenging	3,24E-09 - 2,28E-03

Physiological System Development and Function	p-value range
Hematological System Development and Function	6,36E-07 - 3,56E-03
Connective Tissue Development and Function	1,10E-06 - 4,55E-03
Hematopoiesis	4,02E-06 - 4,07E-03
Tissue Morphology	4,02E-06 - 1,47E-03
Organismal Development	7,03E-06 - 2,28E-03

Statistically significant overrepresentation of genes was examined on canonical pathways, molecular and cellular functions and physiological system development and function categories with QIAGEN's Ingenuity Pathway Analysis (Summer, 2018). The significance values are calculated by Fisher's exact test right-tailed.

Supplementary File Legends

File Name: Additional File 1.

File format: .pdf

Title of the data: Supplementary Material

Description: The file includes Supplementary Methods for EV characterization, Supplementary Figure Legends and altogether five Supplementary Figures named S1-S5.

File Name: Additional File 2.

File format: .xlsx

Title of the data: Supplementary Tables

Description: The file includes altogether 11 Supplementary Tables

Supplementary Table 1. Alignment statistics for RNAseq.

Supplementary Table 2. Expressed genes in RBCs.

Supplementary Table 3. Expressed genes in EVs.

Supplementary Table 4. Estimates for transcript abundances in RBCs.

Supplementary Table 4. Estimates for transcript abundances in EVs.

Supplementary Table 6. Pathway enrichment of expressed genes in RBCs.

Supplementary Table 7. Pathway enrichment of expressed genes in EVs.

Supplementary Table 8. Top overrepresented molecular and cellular functions and physiological system development and function categories in RBC transcriptome.

Supplementary Table 9. Top overrepresented molecular and cellular functions and physiological system development and function categories in EV transcriptome.

Supplementary Table 10. The genes with 5 or more total UMIs across the cells in the combined gene-cell barcode matrix.

Supplementary Table 11. Positive marker genes for cell clusters.

Figures
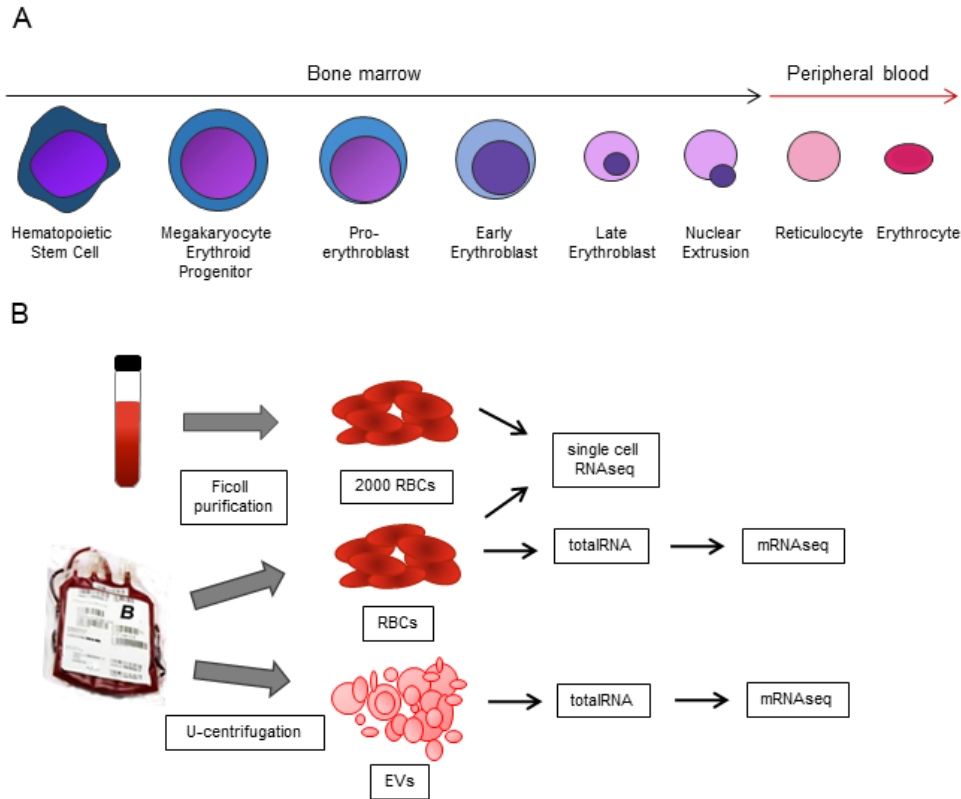
Figure 1

(A) Simplified scheme of erythropoiesis showing hematopoietic stem cell differentiation into megakaryocyte erythroid progenitors and further into pro-erythroblasts which undergo consecutive maturation steps, culminating in nuclear extrusion to yield reticulocytes and subsequently, biconcave erythrocytes. (B) Experimental setup for bulk and single-cell sequencing. For bulk sequencing, red blood cells (RBCs) from leukocyte-filtered blood units (n = 4) were purified with density gradient centrifugation

(Ficoll), and corresponding cell-free supernatant samples were ultracentrifuged to obtain EVs. For single-cell sequencing, a fresh blood sample (n = 1) and a sample from a blood unit (n = 1) were purified with density gradient centrifugation, as well as a fresh blood sample with white blood cell filtering. The total RNA was isolated and used for RNA sequencing. RBC, red blood cell; EV, extracellular vesicle; RNAseq, RNA sequencing.

A

SYMBOL	ENSEMBL ID
HBB	ENSG00000244734
HBA1/HBA2	ENSG00000188536
FTL	ENSG00000087086
MT-RNR2	ENSG00000210082
ALAS2	ENSG00000158578
HBD	ENSG00000223609
FBXO7	ENSG00000100225
BNIP3L	ENSG00000104765
UBB	ENSG00000170315
RPL21	ENSG00000122026
OAZ1	ENSG00000104904
UBA52	ENSG00000221983
SERF2	ENSG00000140264
HBG2	ENSG00000196565
PRDX6	ENSG00000117592
EPB41	ENSG00000159023
HEMGN	ENSG00000136929
SLC25A39	ENSG00000013306
RPS12	ENSG00000112306
GLUL	ENSG00000135821
C9orf78	ENSG00000136819
NFE2	ENSG00000123405
LGALS3	ENSG00000131981
PSMF1	ENSG00000125818
BTF3	ENSG00000145741

B

SYMBOL	ENSEMBL ID
HBB	ENSG00000244734
MT-RNR2	ENSG00000210082
BNIP3L	ENSG00000104765
HBA1/HBA2	ENSG00000188536
FTL	ENSG00000087086
FBXO7	ENSG00000100225
MT-RNR1	ENSG00000211459
HEMGN	ENSG00000136929
RPL21	ENSG00000122026
MT-CO1	ENSG00000198804
UBB	ENSG00000170315
ALAS2	ENSG00000158578
HBD	ENSG00000223609
RIOK3	ENSG00000101782
SEC62	ENSG00000008952
MBNL3	ENSG00000076770
EPB41	ENSG00000159023
YIPF6	ENSG00000181704
B2M	ENSG00000166710
UBA52	ENSG00000221983
RPS12	ENSG00000112306
TMSB10/TMSB4X	ENSG00000205542
C9orf78	ENSG00000136819
CA1	ENSG00000133742
SERF2	ENSG00000140264

Figure 2

Red blood cell (A) and extracellular vesicle (B) transcriptomes. Venny's diagrams show the number of expressed genes for individual samples as well as those shared in all samples. The gene lists (minimum 5 counts/gene) from RBC and EV samples were compared with the Venny 2.1.0 tool using Ensembl gene IDs. On the right, the 25 most highly expressed genes are shown in the corresponding transcriptomes. The full lists with gene names and expression values are shown in Supplementary Table S2 and S3. Data was annotated using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN, Dec 16th, 2016) or Ensembl genome browser (http://www.ensembl.org, release 88 - Mar 2017).
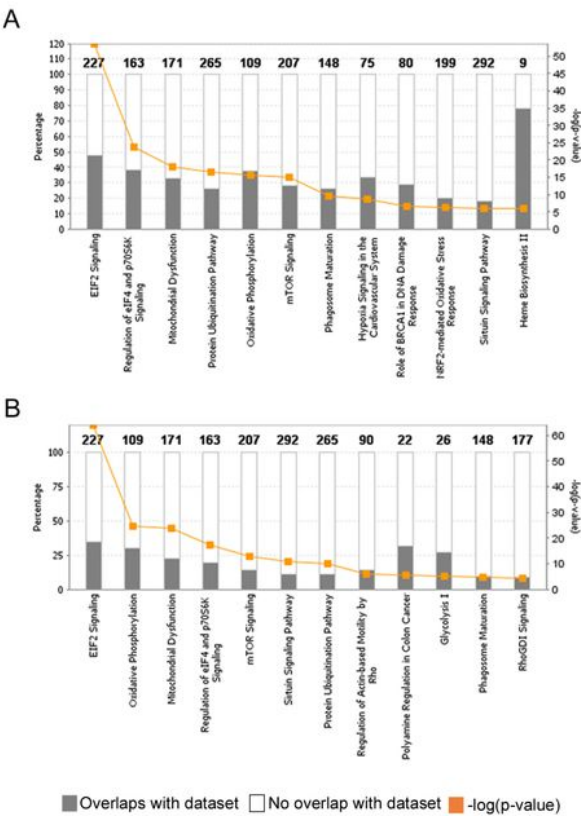
Figure 3

The 12 most significant canonical pathways in RBC (A) and EV (B) transcriptomes, as analyzed with QIAGEN's Ingenuity Pathway Analysis (IPA). Overlaps are shown, as well as the -log(p-values) obtained by Fisher's exact test right-tailed.
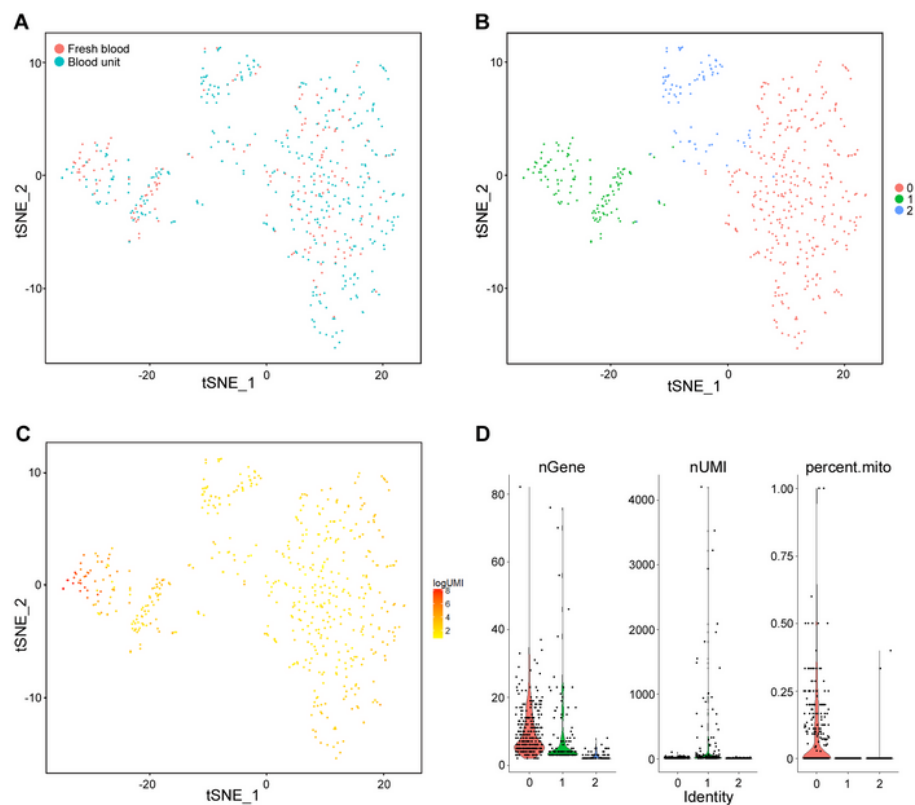
Figure 4

Single-cell analysis of red blood cells (RBCs) reveals three distinct cell populations. (A) 2D t-distributed stochastic neighbor embedding (t-SNE) projection of 554 RBCs, with each cell colored based on their sample ID (blue; cells from blood unit, red; cells from the fresh sample. (B) The tSNE projection of 554 RBCs, where each cell is grouped into one of the identified 3 clusters (0, 1, 2). (C) tSNE projection of 554 RBCs, with each cell colored based on their total unique molecular identifier (UMI) counts in logarithmic scale. (D) Violin plots showing the distribution of the total number of genes (nGene), total UMIs (nUMI) and the percentage of mitochondrial gene expression (percent.mito) per cell for the identified three RBC clusters.

Figure 5

Three red blood cell (RBC) populations are characterized by specific marker gene expression profiles. (A) A heatmap showing the top 30 marker genes for each RBC cluster. The expression of each gene is normalized against the total expression within the cell. (B) Violin plots showing the expression of MALAT1, HBB, HBA1 and HBA2 genes as unique molecular identifier (UMI) counts per cell. (C) T-distributed stochastic neighbor embedding (tSNE) projection of 554 RBCs, with each cell colored based

on its expression of MALAT1, HBB, HBA1 and HBA2 genes. A blue color indicates a positive expression, and gray indicates a negative expression.
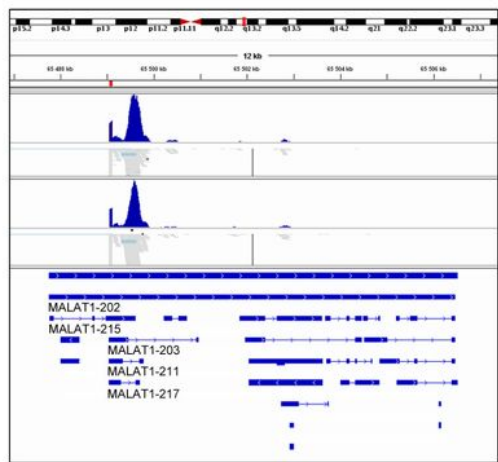
Figure 6

Alignment of sequencing reads mapping to MALAT1 gene from single-cell RNA sequencing analysis. The alignment was performed with an integrative genomic viewer (IGV) using Ensembl gene annotations to visualize short reads matching to the MALAT1 gene.
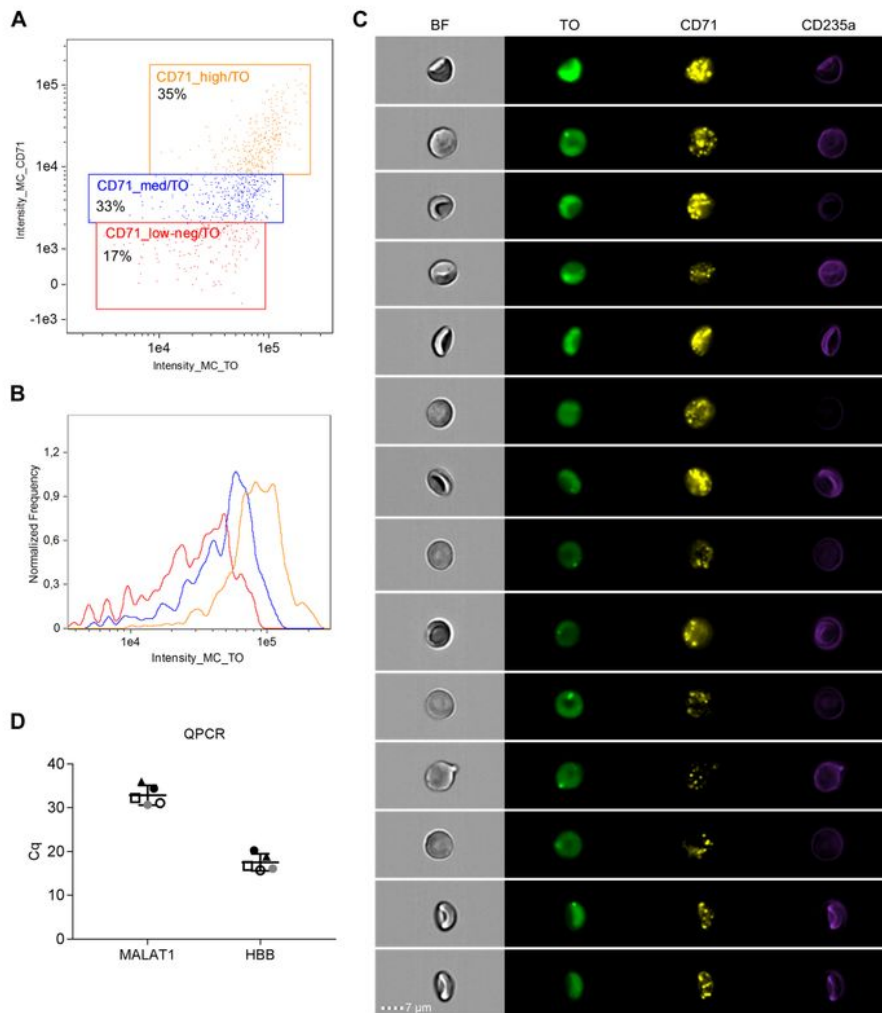
Figure 7

Characterization of enriched reticulocytes using imaging flow cytometry and quantitative PCR. (A) Imaging flow cytometry was performed after staining of CD71-enriched cells with CD71-PE and CD235a-PB antibodies and thiazole orange (TO) for RNA content. Double positive cells for CD71 and TO are shown as well as gates for CD71_high, CD71_medium, and CD71_low-neg populations from one representative sample. (B) Intensity range of TO RNA stain in different CD71-populations. (C)

Representative image gallery showing different morphologies and staining patters for CD71-enriched cells. Each cell is represented by a row of four images acquired simultaneously in flow, from left to right: brightfield (BF), the RNA binding dye TO (green), CD71-PE (yellow), CD235a-Pacific Blue (violet). Scale bar 7 µm. (D) The expression of the MALAT1 and hemoglobin beta (HBB) genes in the CD71-enriched samples (n=5) is shown as quantification cycle (Cq) values. Lines depict mean±SD. Corresponding samples are illustrated with matching symbols.