



OPEN

# Transcriptomic landscape of blood platelets in healthy donors

Anna Supernat<sup>1</sup>✉, Marta Popęda<sup>1</sup>, Krzysztof Pastuszak<sup>1,2</sup>, Myron G. Best<sup>3,4</sup>, Peter Grešner<sup>1</sup>, Sjors In 't Veld<sup>3,4</sup>, Bartłomiej Siek<sup>5</sup>, Natalia Bednarz-Knoll<sup>1</sup>, Matthew T. Rondina<sup>6,7,8,9</sup>, Tomasz Stokowy<sup>10</sup>, Thomas Wurdinger<sup>3,4</sup>, Jacek Jassem<sup>11</sup> & Anna J. Żaczek<sup>1</sup>

Blood platelet RNA-sequencing is increasingly used among the scientific community. Aberrant platelet transcriptome is common in cancer or cardiovascular disease, but reference data on platelet RNA content in healthy individuals are scarce and merit complex investigation. We sought to explore the dynamics of platelet transcriptome. Datasets from 204 healthy donors were used for the analysis of splice variants, particularly with regard to age, sex, blood storage time, unit of collection or library size. Genes *B2M*, *PPBP*, *TMSB4X*, *ACTB*, *FTL*, *CLU*, *PF4*, *F13A1*, *GNAS*, *SPARC*, *PTMA*, *TAGLN2*, *OAZ1* and *OST4* demonstrated the highest expression in the analysed cohort, remaining substantial transcription consistency. *CSF3R* gene was found upregulated in males (fold change 2.10, FDR  $q < 0.05$ ). Cohort dichotomisation according to the median age, showed upregulated *KSR1* in the older donors (fold change 2.11, FDR  $q < 0.05$ ). Unsupervised hierarchical clustering revealed two clusters which were irrespective of age, sex, storage time, collecting unit or library size. However, when donors are analysed globally (as vectors), sex, storage time, library size, the unit of blood collection as well as age impose a certain degree of between- and/or within-group variability. Healthy donor platelet transcriptome retains general consistency, with very few splice variants deviating from the landscape. Although multidimensional analysis reveals statistically significant variability between and within the analysed groups, biologically, these changes are minor and irrelevant while considering disease classification. Our work provides a reference for studies working both on healthy platelets and pathological conditions affecting platelet transcriptome.

Circulating blood platelets are anucleate cytoplasm fragments, derived from megakaryocytes. Once released, they circulate in the bloodstream for up to 10 days, playing a crucial role in the maintenance of homeostasis and regulation of inflammation<sup>1,2</sup>. Platelets are composed of a cell membrane, organelles, elements of cytoskeleton and granules. Equipped with mitochondria but devoid of the nucleus, they lack genomic DNA, remaining incapable of *de novo* transcription<sup>3-5</sup>. Thus, they rely on a rich repertoire of RNAs, complemented by the presence of translational machinery, which allows translation of nascent transcripts into proteins under environmental stress<sup>6</sup>. Platelets host a broad spectrum of transcripts, including a variety of messenger RNAs, diverse classes of non-coding RNAs, microRNAs and circular RNAs<sup>3,5</sup>.

Historically, first studies on platelet RNA content involved Northern blot hybridization method, later replaced with a more robust RT-qPCR (Reverse Transcription quantitative Polymerase Chain Reaction). Along with the implementation of microarrays, it became feasible to further characterize the platelet transcriptome<sup>7,8</sup>. The introduction of RNA sequencing (RNA-seq) provided even more detailed information on the number and biotypes of RNAs in platelets. This method, combined with platelet sorting, allowed investigators to characterize and compare

<sup>1</sup>Laboratory of Translational Oncology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Dębinki 1, 80-211 Gdańsk, Poland. <sup>2</sup>Department of Algorithms and Systems Modelling, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland. <sup>3</sup>Department of Neurosurgery, Brain Tumor Center Amsterdam, Cancer Center Amsterdam, Amsterdam UMC, VU University Medical Center, Amsterdam, The Netherlands. <sup>4</sup>Brain Tumor Center Amsterdam, Amsterdam UMC, VU University Medical Center, Amsterdam, The Netherlands. <sup>5</sup>Department of History and Philosophy of Medical Sciences, Medical University of Gdańsk, Gdańsk, Poland. <sup>6</sup>University of Utah Molecular Medicine Program, Salt Lake City, UT, USA. <sup>7</sup>Department of Internal Medicine, Division of General Internal Medicine, University of Utah, Salt Lake City, UT, USA. <sup>8</sup>George E. Wahlen Veterans Affairs Medical Center Department of Internal Medicine and the Geriatric Research Education and Clinical Center (GRECC), Salt Lake City, UT, USA. <sup>9</sup>Department of Pathology, University of Utah, Salt Lake City, UT, USA. <sup>10</sup>Department of Clinical Science, University of Bergen, Bergen, Norway. <sup>11</sup>Department of Oncology and Radiotherapy, Medical University of Gdańsk, Gdańsk, Poland. ✉email: anna.supernat@gumed.edu.pl

young, newly released platelets to platelets that had been circulating for a longer period of time<sup>9</sup>. The ultimate golden trove for platelet transcriptome analysis remains single platelet sequencing. This approach, however, is challenging due to the limited amount of input material, with one platelet estimated to have ~2.2 femtograms of total RNA mass<sup>10</sup>. Currently, only single megakaryocyte RNA sequencing is feasible<sup>11,12</sup>.

While maintaining homeostasis and mediating inflammation, platelets are implicated in other pathophysiological processes. Their function, and thus content, might be affected by the ongoing disease, and the list of ailments with altered platelet transcriptome keeps expanding. A large number of studies concerns heart conditions and cancer. Aberrant RNA profiles have been discovered in cardiovascular disease<sup>13</sup>, atrial fibrillation<sup>14</sup>, acute myocardial infarction<sup>15</sup>, as well as in myeloproliferative neoplasms<sup>16</sup>, and solid tumours<sup>17,18</sup>. In the latter case, platelets have even earned a separate name: TEPs (Tumor Educated Platelets). An altered platelet transcriptomic profile has also been observed in sickle cell disease<sup>19</sup> and sepsis<sup>20</sup>.

In healthy humans, the number of blood cells remains tightly controlled within narrow physiologic ranges, with cytokine-mediated signalling and transcriptional circuits regulating hematopoietic dynamics<sup>21</sup>. The platelet transcriptome is considered stable in healthy individuals over time<sup>22</sup>. Nevertheless, RNA expression in hematopoietic stem cells is subject to changes as the cells age<sup>23</sup> or become exposed to sex hormones<sup>24</sup>. Thus, we wondered about the extent of dynamics in the platelet transcriptional profile. Available reports concerning this subject included a limited number of samples or used healthy donor samples as a frame of reference, not as the direct subject of study<sup>25–28</sup>. As platelet RNA-seq is increasingly used by groups around the world, we sought to combine and analyze platelet RNA-sequencing data sets from 204 healthy donors. We hypothesized that the highest ever reported number of samples would enable us to more robustly identify and control for variability in the platelet transcriptome which might not otherwise be apparent in smaller datasets. We analyzed platelet transcriptome with respect to the age, gender or technical variables, such as sequencing library size, storage time, and blood collection conditions (collecting unit). In addition to investigating separate RNA splice variants, we also included multidimensional sample vector analysis. Our work might be considered a reference for studies on both healthy platelets and different pathological conditions.

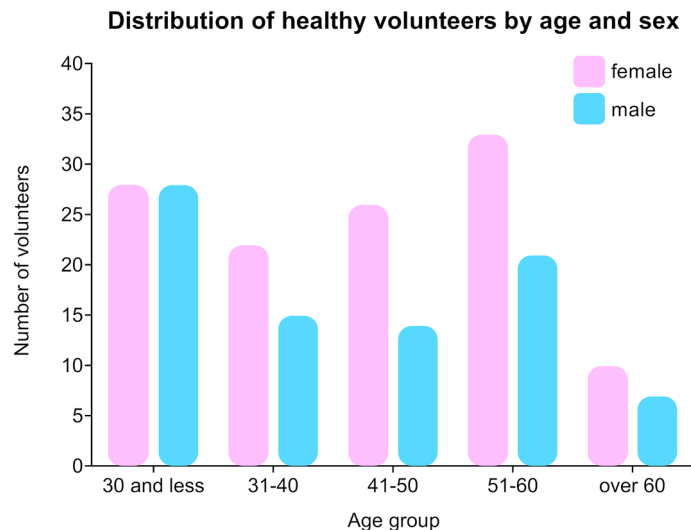
## Materials and methods

**Study cohort.** Hoping to refine the current knowledge on the platelet transcriptome landscape, we included samples collected from 204 asymptomatic controls. These were all the samples available at the moment of analysis. Assuming the variance of splice variant expression levels in platelets of approximately 15% (interquartile range: IQR of 11–19%), statistical power of 80%, the level of statistical significance of 0.05 and the use of double-sided statistical tests, such sample size was far enough to indicate between-group statistically significant differences as little as only 5% (IQR 4–7%). It is important to emphasize that when comparing biological differences among the studied subgroups, we were focusing on the expression changes of at least 50%, while considering at least twofold (100%) differences as possibly biologically meaningful.

All the donors provided written informed consent and reported to be disease-free. The individuals were not subjected to extensive diagnostic tests to rule-out undiagnosed conditions. Whole blood was collected into 6-mL EDTA Vacutainer tubes and then processed according to thromboSeq protocol published before<sup>29</sup>, with the majority of the samples processed between 2 and 12 h after collection. The study was conducted in accordance with the principles of the Declaration of Helsinki. Approval for this study was issued by the Medical Ethical Committee, VU University Medical Center, Amsterdam, The Netherlands. The detailed list of cases is presented as Supplemental Table S1. RNA-seq data is available at Gene Expression Omnibus (GEO) under accession number: GSE89843. Prior to bioinformatics processing, all samples passed bioinformatics quality checks included in thromboSeq protocol<sup>29</sup>. In brief, the intron-spanning spliced RNA reads were selected to eliminate possible DNA contamination. Next, gene body coverage analysis was performed. We additionally applied a cut-off of 100 k reads per sample to improve the reliability of results. Our cohort consisted of 195 samples collected at the Amsterdam UMC VU University Medical Center and 9 samples collected in Amsterdam UMC Academic Medical Center Amsterdam. One hundred nineteen patients were female, and 85 were male. Details on age and sex distribution in the cohort are presented in Fig. 1. The reason why more females than males were included in the dataset is due to the type of study this reference cohort was collected for. All donors were adult, representing all age groups (range 18–86). The mean age was 42.11 years [standard deviation (SD) 13.890], and the median age was 42.5 years (IQR 29–53). Samples were sequenced in several, randomly allocated batches between November 2015 and July 2016.

**Processing of RNA-seq data.** Raw RNA-seq data encoded in FASTQ-files were subjected to a standardized RNA-seq alignment pipeline, as described previously<sup>17,29</sup>. Briefly, reads were subjected to trimming of sequence adapters by Trimmomatic (v. 0.22), mapped to the human reference genome (hg19) using STAR (v. 2.3.0), and summarized using HTSeq (v. 0.6.1), guided by the Ensembl gene annotation version 75. Samples with less than 100 k total reads were excluded from further analysis. Ensembl IDs were converted into gene names using Gencode v19 GRCh37 annotation. Genes with known status were accepted for further analysis. In cases with two IDs mapped to the same gene name, level 1 records were used (Comprehensive gene annotation files, Gencode).

**Statistical analysis.** Subsequent analysis of generated read counts was performed in R (version 3.6.1) and R-studio (version 1.2.5019). For normalization, the DESeq2 R package<sup>30</sup> with Variance Stabilizing Transformation<sup>31</sup> was used. Transcripts with mean log<sub>2</sub> count below four in all groups were excluded. Basic statistical analysis involved t-tests with Benjamini & Hochberg false discovery rate (FDR) correction<sup>32</sup>. Mean-based fold changes (FC) between age-sex-related groups and sample clusters were evaluated. Correlation between spliced junction reads and age was estimated with the Pearson correlation coefficient. Unsupervised cluster-



**Figure 1.** Age and sex distribution of healthy donors.

ing of the samples was performed using Ward D2 agglomerative clustering algorithm<sup>33</sup> as implemented in R package stats (source: <https://www.rdocumentation.org/packages/stats/>). Heatmaps were generated for the most differing transcripts using heatmap3 R package (source: <https://cran.r-project.org/web/packages/heatmap3/>). Selected transcripts were functionally annotated with Reactome pathways using over-representation analysis tool by ConsensusPathDB<sup>34</sup>. Interactions between protein products of selected transcripts were visualized using STRING v11<sup>35</sup>. Canonical platelet transcripts were defined according to the platelet markers listed in PanglaoDB<sup>36</sup>. To further assess variability, splice variants were analysed globally as multidimensional vectors. Variability in platelet transcriptome was assessed by testing the between-group differences in respective mean RNA transcript vectors, between-group differences in within-group multidimensional RNA transcript vector distribution, and between-group differences in within-group multidimensional similarity among RNA transcript vectors. The between-group differences in mean splice variant vectors and their within-group multidimensional distribution were analysed using the gene set analysis approach by means of the two-sample nonparametric multivariate test of means based on the minimum spanning tree and Kolmogorov–Smirnov statistic and the multivariate generalization of the Wald–Wolfowitz runs test, respectively, as implemented in the GSAR R package obtained from Bioconductor. Within-group multidimensional similarity among splice variant vectors was calculated as median reciprocal to all pair-wise Euclidean distances among all possible pairs of RNA splice variant vectors from a given group. Statistical significance of between-group differences in such multidimensional within-group similarity among RNA transcript vectors was assessed by testing the difference between median similarities in the compared groups against its permutational distribution obtained from 10,000 permutations. Statistical significance was inferred for  $p < 0.05$ , with Bonferroni correction for multiple testing where necessary.

## Results

**General characterization of platelet transcriptome in healthy individuals.** Expression of spliced junction reads in human platelets ranged up to six orders of magnitude. The average number of reads per sample was 811,857.6 in the evaluated cohort of 204 healthy donors. Initially, 57,736 different splice variants were mapped. As the quality of the mapped transcripts varied, and a number of them were expressed only in a small subset of samples, reads were subjected to quality control and filtering of low expression RNAs, using the exclusion criteria of normalized log<sub>2</sub> counts below four in all groups. Eventually, 3954 transcripts were qualified as appropriate for further analysis. RNA biotypes before and after filtering are presented as Supplementary Fig. S1. Importantly, SMARTer Kit used in the sequencing protocol enriches for poly-A tailed RNAs, which affects the types of RNAs identified.

Most of the analysed variants in the cohort included mRNAs (98.6%), with a small number of pseudogene transcripts (1.1%). Genes *B2M*, *PPBP*, *TMSB4X*, *ACTB*, *FTL*, *CLU*, *PF4*, *F13A1*, *GNAS*, *SPARC*, *PTMA*, *TAGLN2*, *OAZ1* and *OST4* demonstrated the highest expression in the dataset, retaining substantial expression stability (mean log<sub>2</sub> normalized counts > 12, SD range 0.51–1.04, Supplementary Table S2). STRING analysis revealed a strong network interaction occurring between the protein products of these 14 transcripts, as shown in Supplementary Fig. S2. These mRNAs were heavily involved in vesicle-mediated transport and secretion. Further functional annotation analysis of enriched Reactome pathways (CPDB, ConsensusPathDB database) demonstrated that the top 1000 transcripts were typically involved in protein synthesis and platelet activation signalling, as presented in Fig. 2. For clarity of depiction, we are showing only the first top 20 expressed pathways. We also examined the variance in the expression of spliced junction reads in our cohort. Not only were the platelets of healthy donors consistent in terms of canonical platelet splice variants defined according to PanglaoDB (SD

## Functional annotation analysis of top1000 transcripts



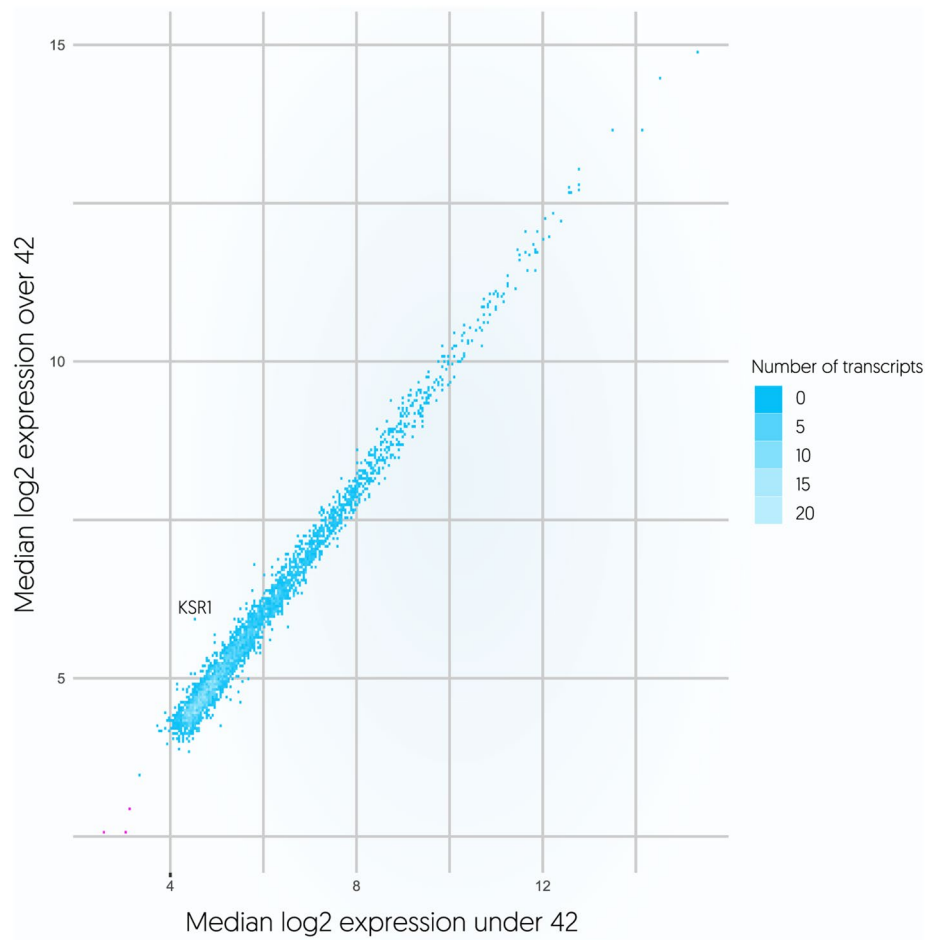
**Figure 2.** Functional annotation analysis of top expressed spliced genes among the analysed sex and age groups. For each group, 1000 spliced genes with the highest expression level were subjected to over-representation analysis and top20 Reactome pathways were identified based on  $-\log_{10}(\text{q-value})$ .

range 0.44–1.94), but also remained considerably stable across the remaining non-canonical transcripts (SD range 0.29–2.57) (Supplementary Table S2).

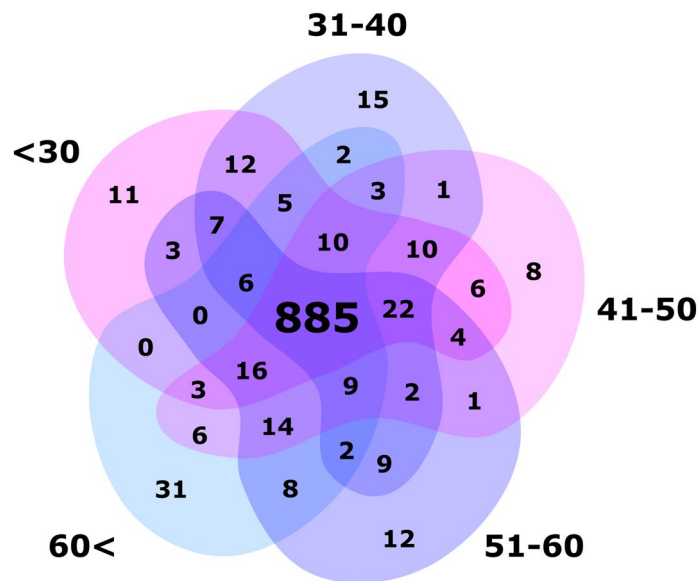
**Platelet splice variants according to sex and age.** When comparing transcriptomic profiles between males and females, mean spliced reads expression analysis per chromosome confirmed that the major discrepancy was observed for chromosome Y (Supplementary Fig. S3). Except for Y-specific differences, the female and male platelet transcriptome demonstrated substantial similarity, reflected by 94.4% overlap of top 1000 transcripts in both groups. Only one gene—*CSF3R*, coding for Colony Stimulating Factor 3 Receptor, was found significantly upregulated in males when compared to females ( $\text{FC} = 2.10$ , FDR-corrected  $q\text{-value} < 0.05$ ). We also observed a moderate upregulation ( $\text{FC} > 1.5$ , FDR-corrected  $q\text{-value} < 0.05$ ) of another 82/3954 (2.1%) transcripts that were predicted to be involved in neutrophil-mediated immune response and extracellular matrix organization (Supplementary Table S3).

We next investigated the effect of ageing on the platelet transcriptome. Direct correlation analysis demonstrated a modest association ( $r > 0.3$  and  $r < -0.3$ ) between the expression of 32/3954 (0.8%) transcripts and age. The gene list included: *KSR1*, *STXBP2*, *CAPZB*, *KIAA2013*, *FCGR2A*, *PPIF*, *TALDO1*, *MGAT4B*, *RAB6B*, *SF3A2*, *GPSM3*, *CAPNS1*, *ZYX*, *TMED3*, *TMSB4X*, *RAB11A*, *BROX*, *UBE2K*, *CALM1*, *TADA2A*, *THOC2*, *TRAPPC2*, *OPA1*, *HIST1H4H*, *EIF5*, *CALM2*, *AP1S2*, *HIST1H2BC*, *STK4*, *MAP1LC3B*, *CARD8* and *KRAS*. The highest coefficient in the whole dataset was observed for the gene *KSR1* ( $r = 0.409$ , FDR corrected  $q\text{ value} < 0.001$ ) encoding Kinase suppressor of Ras 1. Cohort dichotomization, according to the median age (42.5 years) has confirmed that *KSR1* was the only significantly differing gene between the older and younger participants ( $\text{FC} = 2.11$ , Fig. 3). When considering 50% differences in spliced read counts, 4 transcripts—*HDAC8*, *TGFBI*, *FOS* and *SAMD14*, were upregulated ( $\text{FC} > 1.5$ ) and 4 transcripts—*IL7R*, *LEF1*, *HIST1H4H* and *ANXA3*, were downregulated ( $\text{FC} < 0.67$ ) in older versus younger donors (all FDR corrected  $q\text{ value} < 0.05$ ). Nonetheless, analysis of mean spliced reads per chromosome revealed substantial consistency among younger and older donors. To further explore the dynamics of transcriptome variability in donors we applied various age cut offs (Supplementary Figs. S4–7). This additional discrimination confirmed considerable homogeneity across splice variants analysed separately, with high top 1000 transcript overlap (88.5%) between 5 age subgroups (Fig. 4).

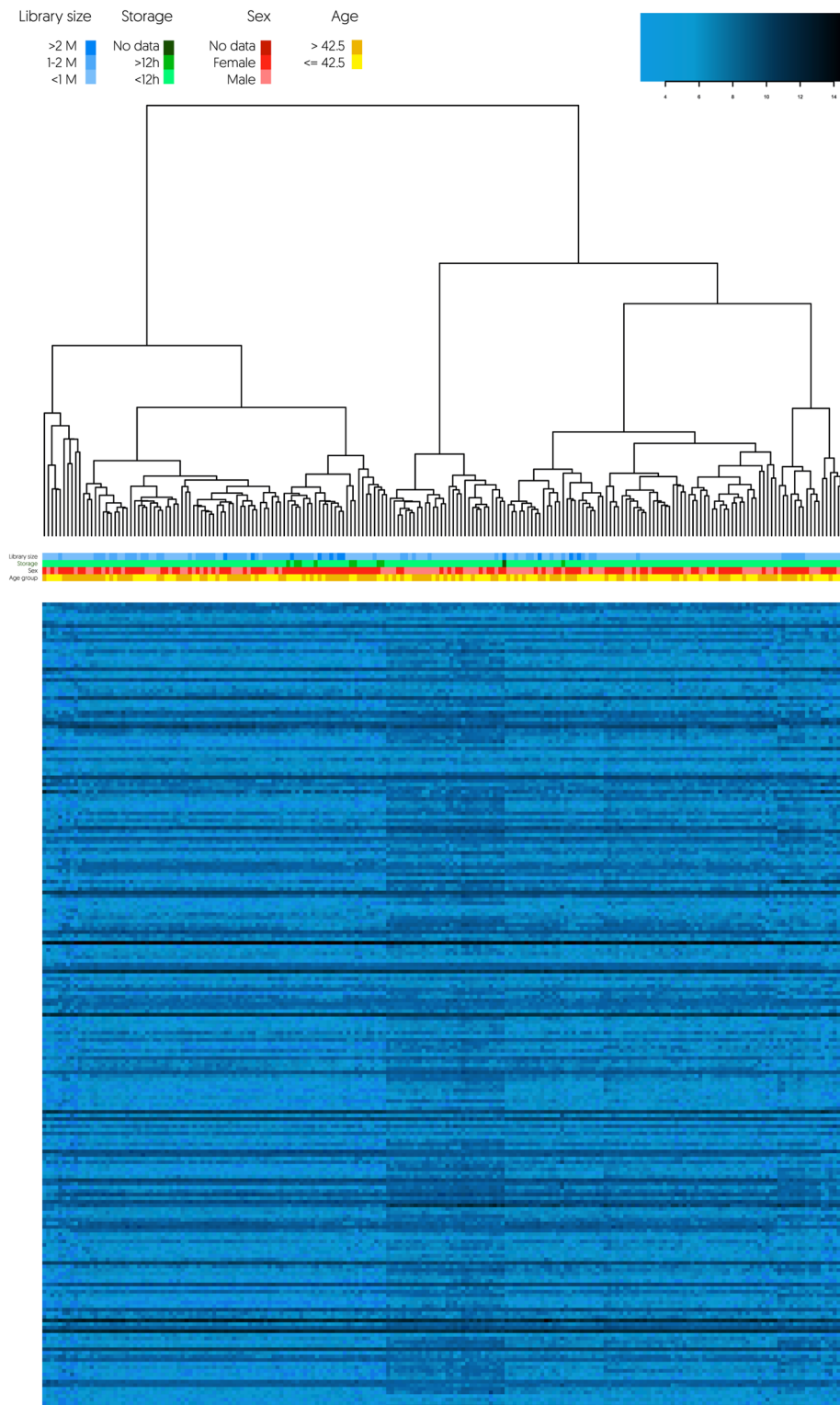
**Global variability of platelet transcriptome.** To further explore platelet transcriptome, we decided to verify whether there was any other variability within the analysed dataset. Unsupervised hierarchical clustering revealed two clusters of samples, depicted in Fig. 5. In this process of unsupervised clustering, samples are assigned to groups with respect to their expression profile, not parameters such as sex or age. Eventually, samples similar to one another in terms of splice variant expression are clustered together, while samples with differing expression profiles will be clustered apart. This means that clusters are built based on the similarity of subsets of samples. Importantly, all the analysed factors (sex, age, unit of blood collection or library size) were randomly distributed across both clusters. Hence, there was no bias towards any of the clusters with respect to sex, age, unit of blood collection or library size. This suggests that these variables did not affect platelet transcriptome



**Figure 3.** Plot representing the relation between log2 median spliced platelet RNA profile of donors younger (X axis) and older than 42 years (Y axis).



**Figure 4.** Top 1000 splice variants notably overlap between the studied five age groups.



**Figure 5.** Heatmap depicts the expression of 224 spliced genes with the greatest, at least fourfold mean changes between the two clusters revealed by unsupervised hierarchical clustering. Colour legend indicates the sex, age group, blood collection unit, storage time and library size of each sample. The clusters presented substantial similarity in terms of the expression of spliced junction reads. 62% of spliced platelet RNAs demonstrated less than a twofold expression difference. Only 16% of spliced genes in the dataset (649/3954) demonstrated at least a threefold expression difference, while at least fourfold difference was observed in only 6% of the spliced reads (224/3954). Due to prior clustering, all these observations were statistically significant (FDR corrected  $q$  value  $< 0.05$ ).

Analysed parameter potentially affecting platelet transcriptome	Mean splice reads vectors ( $p_{KS}$ )	Within-group splice read distribution ( $p_{WW}$ )	Within-group similarity among splice read vectors ( $p$ )
<b>Gender</b>			
Female vs. male	<b>0.0186<sup>a</sup></b>	<b>0.0010<sup>a</sup></b>	0.8502
<b>Age (Group division I)</b>			
< 30 vs. 31–40	0.7012	0.3497	0.0196
< 30 vs. 41–50	0.3761	0.1009	0.7144
< 30 vs. 51–60	0.3843	0.2348	0.0552
< 30 vs. > 60	0.8304	0.0140	0.9676
31–40 vs. 41–50	0.6157	0.0809	0.3815
31–40 vs. 51–60	0.9509	0.1838	0.0259
31–40 vs. > 60	0.6137	0.1968	0.7439
41–50 vs. 51–60	0.0653	0.5475	0.3360
41–50 vs. > 60	0.4794	0.5894	0.0234
51–60 vs. > 60	0.7843	0.9341	0.7600
<b>Age (Group division II)</b>			
< 40 vs. 41–60	0.6985	<b>0.0130<sup>a</sup></b>	0.2115
< 40 vs. > 60	0.8724	0.1189	0.2544
41–60 vs. > 60	0.3450	0.9151	0.0240
<b>Age (median)</b>			
< 42 vs. > 42	0.5674	<b>0.0020<sup>a</sup></b>	0.1150
<b>Storage time</b>			
< 12 h vs. > 12 h	<b>0.0010<sup>a</sup></b>	0.0629	0.0631
<b>Cluster</b>			
Cluster 1 vs. Cluster 2	<b>0.0010<sup>a</sup></b>	<b>0.0010<sup>a</sup></b>	<b>0.0004<sup>a</sup></b>
<b>Library size</b>			
1 vs. 2	<b>0.0005<sup>b</sup></b>	0.2398	<b>&lt;&lt;0.0001<sup>b</sup></b>
1 vs. 3	<b>0.0029<sup>b</sup></b>	0.5225	<b>0.0013<sup>b</sup></b>
2 vs. 3	<b>0.0159<sup>b</sup></b>	0.6943	0.5301

**Table 1.** Summary of analysis of variability associated with selected factors. Values below the assumed thresholds are marked in bold. <sup>a</sup> $p_{\text{threshold}} = 0.0500$ . <sup>b</sup> $p_{\text{threshold}} = 0.0167$ .

considerably. Of note, all the samples passed our QC filtering criteria (the poorest quality analysed was 130,709 reads). Intriguingly, all of the differentially expressed spliced junction reads with at least fourfold difference were upregulated in cluster 2. According to functional analysis, those 224 genes were involved in processes related to neutrophil-mediated immune reaction and interactions at the vascular wall, as depicted in Supplementary Fig. S8. To this end, we named the first cluster “Immune cold”, while the second “Immune hot”. There were 6 splice variants (*FCN1*, *ITGB2*, *LYZ*, *PTPRC*, *CTSS* and *CPVL*) with over tenfold difference between the clusters, all associated with neutrophil degranulation process.

Noteworthy, the clusters demonstrated a substantial similarity in terms of the most abundant transcripts. At this stage, we inspected the platelet RNA profiles manually, by looking at the top 100 spliced genes in each cluster. Altogether, we identified 134 unique splice variants, with an overlap of 78 transcripts and 28 transcripts specific for each cluster. The top 78 spliced genes common for both clusters were mainly platelet-specific, involved in haemostasis, cell–extracellular matrix interactions and smooth muscle contractions processes (CPDB). When focusing on differences, top spliced reads unique for cluster “Immune cold” were implicated in the synthesis and metabolism of fatty acids, whereas those unique for cluster “Immune hot” were associated with translation processes.

Next, we assessed the variability of platelet transcriptome considering each donor’s transcriptome a multidimensional vector and analysed between and pithing group variability (Table 1). Statistically significant differences in multidimensional mean RNA transcript vector were found between groups obtained based on gender, storage time, unsupervised Ward-based hierarchical clustering, and library size. The within-group multidimensional distribution of RNA transcript vectors was found to differ between groups obtained based on gender, age, clustering, median, and processing time. The within-group similarity among RNA transcript vectors were found to differ between groups obtained based on clustering, library size, and processing time. Detailed numerical data are available as Supplementary Table S4.

## Discussion

As great efforts are invested in translating the sequencing of liquid biopsies into clinical practice, it is becoming imperative to provide detailed information on RNA blood content. The available literature on healthy platelet transcriptome mostly relies on a limited sample size<sup>22,25,27,28</sup>. Small sample size is bound to be prone to high FDR

q values, and thus conclusions need to be derived with caution. We did not compare our results directly to other publicly available datasets<sup>22,37</sup> as these data differed from our study with respect to the: participant enrolment procedure, blood processing protocol, RNA extraction protocol, RNA reverse transcription, library preparation, sequencing settings, bioinformatic pipeline and eventually, splice variant filtering criteria. In the presented study, which included over 200 samples, we have shown consistency of separately analysed splice reads which rarely surpassed criteria of statistical significance. Homogeneity regarding fold changes prevailed in direct comparison (female versus male, young versus old), with a few consistent tendencies (increase or decrease) across the analysed sex and age groups: *CSF3R*, *KSR1*, *IL7R*, *TGFB1*, *IGTA2B*. Despite these tendencies, a considerable expression overlap would be observed between the studied groups. Interestingly, Segal and Moliterno reported platelet counts to vary according to age, sex, and also ethnicity<sup>38</sup>. Their study performed on 12,142 donors has shown a very small but steady decline in platelet counts with age, the higher platelet count in women, and the lowest platelet count in whites when compared to other ethnicities. We found rather minute differences among age and sex groups, with considerable overlap in splice variant expression between the studied sub-cohorts.

Age-related changes in platelet transcriptome were published before<sup>39,40</sup>. Campbell et al. identified 514 differentially expressed transcripts, with most (455) being increased in older adults aged 65 years and older. However, the exact same transcripts, when analysed in a large cohort of our study, would range only between 0.61 to 1.46 in terms of fold change (when over 60 and 40 and less groups were compared). Importantly, as the study of Campbell et al. validated the expression and function of Granzyme A by multiple transcriptomic, proteomic, and functional biochemical assays, the discrepant findings are likely the consequence of different sequencing techniques, analytics, filtering methods, and healthy donor populations.

Another interesting age-related study included six adults and six neonates, with 201 genes found differentially expressed<sup>41</sup>. According to this report, neonatal platelets would contain high RNA levels of transcripts associated with protein synthesis and processing, simultaneously carrying lower levels of genes involved in calcium transport, calcium metabolism and cell signalling. Our study did not contain subjects younger than 18, but then again, the low sample size could be the source of bias in the aforementioned report.

The study of Edelstein et al. reported racial differences in human platelet transcriptome, with a very thoroughly characterised healthy volunteer group of 154 donors<sup>26</sup>. Unfortunately, our data lacked information on ethnicity. However, the included samples were collected in Amsterdam. Hence, roughly half of the donors were likely to be Dutch, with most of the remaining participants of European, Moroccan, Surinamese and Turkish descent. Additionally, our study was based on RNA-seq, not microarrays. As much as it renders any comparisons challenging, it is worth noting that, contrary to our observations, Edelstein et al. reported multiple mRNAs and miRNAs to be differentially expressed between the compared groups. These discrepancies are likely to stem from our assumed strict filtering criteria, adjusted to thromboSeq classification<sup>29</sup>. Any compromise in filtering would yield lower quality and thus the credibility of the performed analysis. Of note, the presented QC eliminated most of the protein-coding transcripts, which were reported to have low expression, and all miRNAs sequenced.

Our unsupervised hierarchical clustering revealed two, relatively evenly distributed clusters of samples. Importantly, the clusters were irrelevant of age, sex, or the unit of blood collection. We hypothesise that patients belonging to the “Immune hot” cluster may have been undergoing covert infection. Apart from all this stated above, the landscape of platelet transcriptome presents slightly differently when donors’ transcriptomes are analysed globally, as multidimensional vectors. To the best of our knowledge, there are no previous studies reporting expression data as vectors. In such analyses, factors such as gender, storage time, library size, the unit of blood collection as well as age and processing time were identified as factors capable of introducing between-group variability in the mean distribution of RNA splice variant vectors and/or the within-group similarity among them. Importantly, while statistically significant differences need to be acknowledged, biological significance of these changes seems of far lesser importance. Hence, the transcriptome of healthy subjects is exquisitely durable while being dynamically altered in the state of disease, as also highlighted by Davizon-Castillo et al.<sup>11</sup>. Consequently, artificial intelligence-based classifiers such as thromboSeq or imPlatelet, while performing discrimination between healthy and diseased individuals appear irrelevant of variables such as age or sex<sup>18,42</sup>.

One of the limitations of the study was the lacking information on platelet count of each donor. It might be speculated that re-scaling of the obtained read counts accounting for this variable, or factors such as menopause status or hormone replacement therapy, would refine the performed analysis, demonstrating even further consistency. Unfortunately, sample enrolment could not rule out cases with undiagnosed cancer, an ongoing inflammatory or infectious disease. Due to the anonymization of the sample collection procedure, we could not trace back the healthy donors who donated blood for clinical follow-up. Another limitation was the application of RNA-seq approach only, without further validation by other methods, such as RT-qPCR or proteomics analysis. It is also crucial to point out the weaknesses of the applied method: used reagents enrich for polyA splice variants, whereas bioinformatic pipeline causes the loss of one-exon transcripts and miRNAs which might also be of relevance in the disease process. The last important limitation was an occasional long period between blood collection and processing. However, we did not observe any crucial variability among the samples which were processed later than 12 h after blood collection. This information is important as high stability makes platelets collected in EDTA-coated tubes a very convenient liquid biopsy material. Finally, an important feature for platelet transcriptome analysis is the risk of leukocyte contamination. It has been previously shown that contamination of such nucleated cells using platelet isolation protocol is low<sup>17,18,29</sup>—also without significant platelet activation—though we cannot rule out co-isolation of other cell fragments of similar size as platelets.

In summary, platelets’ transcriptome carries diagnostic information relevant in conditions such as cancer or cardiovascular disease. As reference data on the platelet transcriptome in healthy individuals are still scarce and merit complex investigation, we performed an in-depth analysis of platelet transcriptome landscape. Our study, based on a large dataset, proves that healthy human platelets have a unique and reproducible transcript profile. The knowledge on the transcriptome landscape can be used for example in RT-qPCR setting, where top



spliced genes, such as *B2B*, *PPBP* and *ACTB*, could potentially constitute reference genes when analysing gene expression. Platelet profile demonstrates consistency, which remains an important implication when recruiting a control cohort for any experiments. We show that separately analysed platelet splice variants retain general consistency, with a few interesting deviating genes from these general trends. When assessed globally (as vectors representing corresponding samples), factors as sex, age, storage time or library size impose a certain degree of between- and/or within-group variability. However, for disease classification purposes, especially cancer, these differences remain negligible<sup>18,42</sup>. The presented results suggest that, when planning an oncology-related study, control cases do not necessarily need to be sex or age-matched. Ideally, they should be, but even without the matching, the classification most probably will not be affected to a classification output extent. Similarly, blood samples should be processed as soon as possible after collection, in room temperature (to avoid platelet activation), but even when processed after over 12 h, the results are still meaningful. This is especially important when considering logistic aspects of diagnostics. In many cases blood material is not directly processed at the hospital, but rather transported prior to analysis. To conclude, factors such as age or storage time introduce small variability, unlikely to have strong biological effects.

Received: 4 January 2021; Accepted: 22 June 2021

Published online: 03 August 2021

## References

- Harker, L. A. *et al.* Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. *Blood* **95**, 2514–2522 (2000).
- Patel, S. R., Hartwig, J. H. & Italiano, J. E. The biogenesis of platelets from megakaryocyte proplatelets. *J. Clin. Investig.* **115**, 3348–3354 (2005).
- Hartwig, J. H. The platelet: Form and function. *Semin. Hematol.* **43**, S94–100 (2006).
- Melchinger, H., Jain, K., Tyagi, T. & Hwa, J. Role of platelet mitochondria: Life in a nucleus-free zone. *Front. Cardiovasc. Med.* **6**, 153 (2019).
- Provost, P. The clinical significance of platelet microparticle-associated microRNAs. *Clin. Chem. Lab. Med.* **55**, 657–666 (2017).
- Kapur, R., Zufferey, A., Boilard, E. & Sempole, J. W. Nouvelle cuisine: Platelets served with inflammation. *J. Immunol.* **194**, 5579–5587 (2015).
- Bahou, W. F. & Gnatenko, D. V. Platelet transcriptome: The application of microarray analysis to platelets. *Semin. Thromb. Hemost.* **30**, 473–484 (2004).
- Macaulay, I. C., Carr, P., Farrugia, R. & Watkins, N. A. Analysing the platelet transcriptome. *Vox Sang.* **87**(Suppl 2), 42–46 (2004).
- Clancy, L., Beaulieu, L. M., Tanriverdi, K. & Freedman, J. E. The role of RNA uptake in platelet heterogeneity. *Thromb. Haemost.* **117**, 948–961 (2017).
- Teruel-Montoya, R. *et al.* MicroRNA expression differences in human hematopoietic cell lineages enable regulated transgene expression. *PLoS One* **9**, e102259 (2014).
- Davizon-Castillo, P., Rowley, J. W. & Rondina, M. T. Megakaryocyte and platelet transcriptomics for discoveries in human health and disease. *Arterioscler. Thromb. Vasc. Biol.* **40**, 1432–1440 (2020).
- Psaila, B. *et al.* Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol.* **17**, 83 (2016).
- Hartman, R. J. G. *et al.* Platelet RNA modules point to coronary calcification in asymptomatic women with former preeclampsia. *Atherosclerosis* **291**, 114–121 (2019).
- Wysokinski, W. E. *et al.* Impact of atrial fibrillation on platelet gene expression. *Eur. J. Haematol.* **98**, 615–621 (2017).
- Eicher, J. D. *et al.* Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets* **27**, 230–239 (2016).
- Zini, R. *et al.* CALR mutational status identifies different disease subtypes of essential thrombocythemia showing distinct expression profiles. *Blood Cancer J.* **7**, 638 (2017).
- Best, M. G. *et al.* RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).
- Best, M. G. *et al.* Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* **32**, 238–252.e9 (2017).
- Jain, S. *et al.* Expression of regulatory platelet microRNAs in patients with sickle cell disease. *PLoS One* **8**, e60932 (2013).
- Middleton, E. A. *et al.* Sepsis alters the transcriptional and translational landscape of human and murine platelets. *Blood* **134**, 911–923 (2019).
- Wu, S. *et al.* BLVRB redox mutation defines heme degradation in a metabolic pathway of enhanced thrombopoiesis in humans. *Blood* **128**, 699–709 (2016).
- Rondina, M. T. *et al.* Longitudinal RNA-Seq analysis of the repeatability of gene expression and splicing in human platelets identifies a platelet SELP splice QTL. *Circ. Res.* **126**, 501–516 (2020).
- Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).
- Calvanese, V., Lee, L. K. & Mikkola, H. K. A. Sex hormone drives blood stem cell reproduction. *EMBO J.* **33**, 534–535 (2014).
- Bray, P. F. *et al.* The complex transcriptional landscape of the anucleate human platelet. *BMC Genom.* **14**, 1 (2013).
- Edelstein, L. C. *et al.* Racial difference in human platelet PAR4 reactivity reflects expression of PCTP and miR-376c. *Nat. Med.* **19**, 1609–1616 (2013).
- Frey, C. *et al.* Longitudinal assessment of the platelet transcriptome in advanced heart failure patients following mechanical unloading. *Platelets* <https://doi.org/10.1080/09537104.2020.1714573> (2020).
- Heffron, S. P., Marier, C., Parikh, M., Fisher, E. A. & Berger, J. S. Severe obesity and bariatric surgery alter the platelet mRNA profile. *Platelets* **30**, 967–974 (2019).
- Best, M. G., Veld, S. G. J. G. I., Sol, N. & Würdinger, T. RNA sequencing and swarm intelligence-enhanced classification algorithm development for blood-based disease diagnostics using spliced blood platelet RNA. *Nat. Protoc.* **14**, 1206–1234 (2019).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**(Suppl 1), S96–104 (2002).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).

33. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?. *J. Classif.* **31**, 274–295 (2014).
34. Kamburov, A. *et al.* ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Res.* **39**, D712–717 (2011).
35. Szklarczyk, D. *et al.* STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
36. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, 1–9 (2019).
37. Marcantoni, E. *et al.* Platelet transcriptome profiling in HIV and ATP-binding cassette subfamily C member 4 (ABCC4) as a mediator of platelet activity. *JACC Basic Transl. Sci.* **3**, 9–22 (2018).
38. Segal, J. B. & Moliterno, A. R. Platelet counts differ by sex, ethnicity, and age in the United States. *Ann. Epidemiol.* **16**, 123–130 (2006).
39. Campbell, R. A. *et al.* Granzyme A in human platelets regulates the synthesis of proinflammatory cytokines by monocytes in aging. *J. Immunol.* **200**, 295–304 (2018).
40. Simon, L. M. *et al.* Human platelet microRNA–mRNA networks associated with age and gender revealed by integrated plateletomics. *Blood* **123**, e37–45 (2014).
41. Caparrós-Pérez, E. *et al.* Comprehensive comparison of neonate and adult human platelet transcriptomes. *PLoS One* **12**, e0183042 (2017).
42. Pastuszak, K. *et al.* imPlatelet classifier: Image-converted RNA biomarker profiles enable blood-based cancer diagnostics. *Mol. Oncol.* <https://doi.org/10.1002/1878-0261.13014> (2021).

## Acknowledgements

This research was supported by the SONATA grant of The National Science Centre (2018/31/D/NZ5/01263) and Medical University of Gdańsk statutory work (ST-23, 02-0023/07). We wish to thank Bartosz Supernat for artwork design and preparation. We would like to acknowledge Adam Wyszomirski for providing biostatistics consultation within the services of the Computational Core located at Medical University of Gdańsk, Poland. The Core Facility is working as part of “Excellence Initiative—Research University” Grant No. MNiSW 07/IDUB/2019/94.

## Author contributions

T.W. and M.G.B. are inventors on relevant patents, and T.W. is shareholder of GRAIL Inc. M.T.R. is an inventor on a relevant patent. A.S., B.S., N.B.K., A.J.Ż. conceptualized and designed the study. K.P., T.S. carried out bioinformatics data processing. A.S., M.P., K.P., T.S., P.G. carried out statistical analysis. A.S., M.P., K.P., P.G., M.G.B., S.I.V., M.T.R., T.W. interpreted the results. M.G.B., S.I.V., M.T.R., T.W. provided feedback on additional analyses. M.P., K.P., T.S. visualized the data. A.S., M.P., K.P., P.G., M.G.B., M.T.R. wrote the manuscript. A.S., M.P., K.P., P.G. prepared tables and supplementary information. A.S., A.J.Ż., J.J. supervised the study and provided funding. All authors have critically reviewed the manuscript.

## Competing interests

TW, MGB, MTR are inventors on relevant patents. TW is shareholder of GRAIL Inc. AS, MP, KP, PG, SIV, BS, NBK, TS, JJ, AJS have no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-94003-z>.

**Correspondence** and requests for materials should be addressed to A.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021