Multi-parameter phenotyping of platelets and characterisation of the effects of agonists using machine learning

Ami Vadgama[1], James Boot PhD[2], Nicola Dark[1], Harriet E. Allan PhD[1], Charles A. Mein PhD[2], Paul C. Armstrong PhD*[1], Timothy D. Warner PhD*[1]

[1]Centre for Immunobiology, [2]Genome Centre; Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

*PCA and TDW contributed equally to this work

Corresponding author: Dr Paul C. Armstrong, p.c.armstrong@qmul.ac.uk, Centre for Immunobiology, Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University London, 4 Newark Street, London, E1 2AT, United Kingdom

Short title: Platelet immunophenotyping using machine learning

Main body of text word count: 3890

Essentials

- Platelet function is directed by the expression of specialised surface markers

- Circulating platelet sub-populations are incompletely characterised

- Multi-parameter spectral flow cytometry allows robust and comprehensive phenotyping of platelets

- Coupling with machine learning offers a powerful method to determine platelet sub-populations

Abstract

Background: Platelet function is driven by the expression of specialised surface markers. The concept of distinct circulating sub-populations of platelets has emerged in recent years, but their exact nature remains debatable.

Objective: To design a spectral flow cytometry-based phenotyping workflow to provide a more comprehensive characterisation, at a global and individual level, of surface markers in resting and activated healthy platelets. Secondly, to apply this workflow to investigate how responses differ according to platelet age.

Methods: A 14-marker flow cytometry panel was developed and applied to vehicle- or agonist-stimulated platelet-rich plasma and whole blood samples obtained from healthy volunteers, or to platelets sorted according to SYTO-13 staining intensity as an indicator of platelet age. Data were analysed using both user-led and independent approaches incorporating novel machine learning-based algorithms.

Results: The assay detected differences in marker expression in healthy platelets, at rest and on agonist activation, in both platelet rich plasma and whole blood samples, that are consistent with the literature. Machine learning identified stimulated populations of platelets with high accuracy (>80%). Similarly, machine learning differentiation between young and old platelet populations achieved 76% accuracy, primarily weighted by FSC-A, CD41, SSC-A, GPVI, CD61, and CD42b expression patterns.

Conclusions: Our approach provides a powerful phenotypic assay coupled with robust bioinformatic and machine learning workflows for deep analysis of platelet sub-populations. Cleave-able receptors, GPVI and CD42b, contribute to defining shared and

unique sub-populations. This adoptable, low-volume approach will be valuable in deep characterisation of platelets in disease. *(244 words)*

Key words: computational biology; flow cytometry; haemostasis; thrombosis, machine learning

Introduction

Haemostasis is a carefully orchestrated process in which platelets are primary players. These metabolically-active cell fragments circulate for approximately 7-10 days in healthy individuals before being degraded by the spleen or liver[1–3]. Platelet function is mediated through the expression of specialised surface markers, with resting and activated platelets showing different expression profiles consistent with response heterogeneity in thrombus formation[4,5]. These variations in surface marker expression are thought to be conferred during production from megakaryocytes, activation history, and ageing in the circulation[6,7], and may describe dynamically discrete and specialised platelet sub-populations[8,9].

As platelets age in the circulation, they lose messenger ribonucleic acids (mRNA) remaining from their progenitor megakaryocytes[10–12]. Platelets have minimal capacity to make new mRNA, and thus these residual mRNAs can be used as a surrogate measure for age[13–15]. Newly-formed or 'young' platelets (also termed reticulated platelets or the immature platelet fraction) have the highest levels of mRNA, while 'old' platelets have the lowest[14,16,17]. Young platelets are hyper-reactive and have an elevated thrombotic potential[18–20]. This is apparent in several pathological states, such as diabetes mellitus[21] and major trauma[18], in which there are relative increases in young platelets resulting from altered platelet turnover and lifespan associated with increased incidences of thromboembolic events[22–25] and decreased efficacy of anti-platelet therapies[26–29].

Flow cytometry with the inclusion of fluorescently-tagged antibodies is frequently used to investigate platelet protein expression and function as it requires only small volumes of blood and a relatively low number of platelets, making it ideal for analysis of clinical samples. However, the number of parameters that can be measured concurrently using this method is limited by overlap of fluorescent emission spectra, resulting in antibody panels that typically determine, at most, three to four markers in any one sample [30,31]. Spectral flow cytometry is a next-generation technique that allows the simultaneous measurement and discrimination of multiple fluorophores by evaluation of full emission spectral signatures. Accounting for steric hindrance, this platform can therefore be used for analysis of 10-15 markers on platelets. Previous work from Blair *et al*.[8] has used mass cytometry to develop a panel to study platelet function, but spectral flow cytometry offers several considerable advantages including cost of reagents, availability of equipment, and simplicity of technique.

With this in mind, we sought to develop a 14-marker spectral flow cytometry panel based upon the panel of Blair *et al*. (noting in this report that Blair *et al*. have subsequently published reports utilizing spectral flow cytometry[30,32]) to provide a more comprehensive characterisation of platelets, at a global and individual-platelet level. In our studies we subjected our data to a powerful computational analytical approach employing machine learning to explore the potential existence of platelet sub-populations in the human circulation with a particular focus on young and old platelets.

Materials and Methods

Ethical statement: human studies

All studies were conducted according to the principles of the Declaration of Helsinki and were approved by St. Thomas' Hospital Research Ethics Committee (Ref.: 07/Q0702/24). Healthy volunteers were aged 23-40, screened prior to entering the study (non-smokers; had not taken non-steroidal anti-inflammatory drugs $\leq$10 days prior to donating blood; no health problems contraindicating study involvement) and gave written informed consent.

Collection of blood and preparation of platelet-rich plasma

Blood was drawn from volunteers by venepuncture into trisodium citrate vacutainers (3.2%; BD Biosciences), and platelet-rich plasma (PRP) was isolated as previously published[24].

Flow cytometric measurement of activation markers

PRP, or whole blood (WB), was diluted 1:40 with 2mM Ca^{2+}-buffered, filtered PBS, and added to a 96-well plate with wells containing vehicle (phosphate-buffered saline, PBS) or agonist: 0.3-30µM adenosine diphosphate (ADP; Labmedics); 0.3-30µM thrombin-receptor activating peptide 6 (TRAP-6; Cambridge Biosciences); 0.3-30µM U46619 (Enzo Life Sciences); 3-100µM protease-activated receptor 4 agonist (PAR-4; Cambridge Biosciences); 0.03-3µM collagen-related peptide (CRP-XL; University of Cambridge). In initial experiments, an antibody mix comprising anti-CD42b-BV421 (1:70; clone HIP1;

BioLegend), PAC-1-FITC (1:10; BD Biosciences), and anti-CD62P-APC (1:100; clone AK4; BioLegend) was added to each well. Alternatively, an antibody master mix (Supplementary table 1 and Supplementary figure 1) and staining buffer (Brilliant Stain Buffer, BD Biosciences) were used. The plate was then mixed (200rpm; 37°C) for 20 minutes in the dark (BioShake iQ, Quantifoil Instruments GmbH). Samples were fixed with 1% formalin and run on the Cytek Aurora 5-laser flow cytometer (Cytek Biosciences). Platelets were gated on SSC-A/CD42b-A, and 10,000 CD42b+ events were collected (Supplementary figure 2).

Flow cytometric sorting of young and old platelets

Adapting our previously published approach[33], PRP was stained with SYTO-13 (750nM; Thermo-Fisher Scientific) prior to sorting. 10 million platelets per condition were sorted using a BD FACS Aria IIIu Fusion Cell Sorter (70µM nozzle, 70Ps, ≤10,000 events/second; BD Biosciences) with the top 20% SYTO-13 fluorescence being taken as 'young' and the bottom 30% SYTO-13 as 'old'. Platelets were pelleted in the presence of prostacyclin (Epoprostanol; 2µmol/L; Tocris Biosciences) at x1000g for 10 minutes and re-suspended in calcium chloride ($CaCl_2$; 2mmol/L; Sigma-Aldrich) -buffered MTH buffer. Panel markers were then measured as described above using the Cytek Aurora.

Statistical, bioinformatics, and machine learning analysis

d

Using the Caret v6.0-93 (Classification And REgression Training) R package[37], we developed a machine learning model to predict whether platelets were treated with

vehicle or agonist. Platelet data were loaded and processed as already detailed, with 10,000 platelets per healthy donor per treatment being loaded, unless otherwise stated. Balanced in number for each condition, from 16 donors, a Random Forest model was given a training dataset made up of 80% of the total data to learn from. The remaining 20% of the total data was used for preliminary validation. The remaining 5 donors were used as an "unseen" validation data set to test the model. A feature importance comparison was then run to determine which markers were most important in the classification. The model was trained using a 10-fold cross-validation using 3 repeats. This means that the training data is first randomly shuffled and split into 10 "folds", then, in turn, each fold is excluded from the training data whilst the remaining 9 folds are used to train a model; the excluded fold is used to assess the accuracy of the trained model. This process was then repeated a total of 3 times; each time the data was shuffled randomly, producing new folds. The metric used to assess the quality of the model was the Receiver Operating Characteristic (ROC); overall accuracy was also reported.

Using the Caret v6.0-93 (Classification And REgression Training) R package[37] we developed Random Forest machine learning models to predict whether platelets were young or old in the presence of vehicle or individual agonists. Platelet data were loaded and processed as already detailed, with 9,000 platelets per young and per old sample from each donor being loaded, outlier events/platelets were removed, and the measurements for PAC-1 excluded. In total there were 8 donors, each with a young and old sample. 6 of the 8 donors were used to create a training and preliminary validation dataset; 80% of the data from the 6 donors was used for training whilst 20% was used for preliminary validation to assess the accuracy of the model. Data from the remaining 2

donors was kept aside to use as an "unseen" validation dataset. As previously, a feature importance comparison was then run to determine which markers were most important in the classification. The model was trained using a 10-fold cross-validation using 3 repeats, and ROC used to assess the quality of the model, as already detailed.

Results

Assay capable of detecting significant changes in marker expression on activation in healthy platelets

To select optimum concentrations of agonists for use in subsequent experiments, flow cytometry was used to assess changes in PAC-1 binding and CD62P expression in response to increasing concentrations of TRAP-6, PAR-4, CRP-XL, ADP, and U46619 (Supplementary figure 3). The following concentrations were selected for their ability to induce a robust response: 10µM TRAP-6, 100µM PAR-4, 3µM CRP-XL, 30µM ADP, and 10µM U46619.

The chosen concentrations of agonists were then tested using the extended phenotyping panel of markers as well as forward scatter (FSC-A) and side scatter (SSC-A). All agonists tested caused significant increases in the expression of CD62P, PAC-1, CD63, CD107a, CD61, CD29, and CD9, and decreases in CD42b; none of the agonists caused any changes in CLEC-2 (Figure 1). TRAP-6 was the only agonist to decrease CD31 (4336±233 vs. 3973±223, p=0.03), and CRP-XL was the only agonist to decrease GPVI (5735±512 vs. 4030±600, p=0.006). CD42a expression was decreased by all agonists

except CRP-XL. Across all conditions, the physical characteristics FSC-A and SSC-A demonstrated the least inter-individual variation (Supplementary figure 4).

High-dimensionality analysis allows visualisation and in-depth interrogation of marker changes in response to activation

Unsupervised dimensionality reduction and visualisation of the entire data sets using tSNE revealed the same shifts in receptor patterns. Namely, agonist stimulation caused visible increases in the expression of PAC-1, CD62P, CD63, CD107a, CD61, and CD9, decreases in CD42b, CD42a, GPVI, CLEC-2, and CD31; CD29 remained unchanged (Figure 2). The same visualisation approach also confirmed that the detected marker expression patterns were shared across donors and not driven by donor or batch effect (Supplementary figure 5). Auto-clustering analysis produced 5 clusters in vehicle-treated platelets and between 9 and 11 clusters in agonist-treated platelets. Each formed cluster was not dominated by individual donors but rather reflected the gradation in difference of expression (Supplementary figure 6). However, hierarchical dendrograms within the clustering indicated that interlinked relationships between the expression of each marker differed by agonist stimulation (Supplementary figure 6).

Machine learning reveals the most important markers in distinguishing effects of agonists

Machine learning (ML) was used to further analyse the data at the single platelet level in an unbiased fashion. Following training, the accuracy of differentiation of vehicle-treated

from agonist-treated platelets was determined. Accuracy rates for unseen datasets were highest for TRAP-6, PAR-4, and CRP of 0.92, 0.91, and 0.88, respectively. Comparatively, rates for ADP and U46619 were 0.80 and 0.77, respectively (Table 1), with the greatest fall between the training and unseen sets also occurring with these agonists.

Rankings of distinguishing markers within each prediction model was examined with those with a weighting of greater than 20 considered important (Figure 3). For identification of TRAP-6-, PAR-4-, and CRP-XL-stimulated platelets, CD62P was considered the most important, followed by PAC-1, CD42b, and CD107a (Figure 3A). Identification of ADP-treated platelets was based predominantly on the expression patterns of PAC-1, CD62P, and CD42b (Figure 3B-D). This was similar to platelets activated with U46619, for which CD62P, CD42b, and PAC-1 were highest (Figure 3E).

CD41/CD61, FSC-A, SSC-A, GPVI, CLEC-2, and CD61 are the primary markers used by machine learning to differentiate between young and old platelets

Next, we undertook phenotypic analysis of young and old platelets, as determined by SYTO-13 RNA staining, following vehicle or agonist stimulation. We first analysed the data on a traditional by-population basis by looking at raw MFI values and changes in MFI (Supplementary figure 7), before applying machine learning. Predictive capability within each training set ranged from 0.86 to 0.90 and maintained high level of accuracy in the unseen datasets (range 0.74 to 0.78; Table 2).

Markers rated greater than an importance of 20 within vehicle-treated platelets identified FSC-A, CD41, SSC-A, GPVI, and CD61 as most important (Figure 4). FSC-A, SSC-A,

and CD41 were the top three discriminators for all agonists tested, with CD61, CD42b, and GPVI being also seen at above 20 in all. CLEC-2 was seen in TRAP-6, and CD61 in CRP-XL.

Subsequently we applied the markers rated greater than an importance of 20 from our separated young and old platelet populations to the mixed healthy platelet data presented previously. This analysis indicated an increased probability of young platelets being present in clusters with higher markers of activation, both in control conditions (i.e. vehicle-treated) and following exposure to platelet agonists (Figure 5).

Assay detects comparable alterations in marker expression following stimulation and labelling of whole blood

We next wished to test the suitability of this assay for use with whole blood, rather than PRP. Agonist stimulation elicited significant increases in the expression of CD62P, PAC-1, CD63, CD107a, CD61, CD29, and CD9, and decreases in CD31 andCD42b; none of the agonists caused any changes in CLEC-2 (Supplementary figure 8). This pattern of altered marker expression was consistent with those observed in PRP samples, but with a significantly greater loss of CD31 (Supplementary figure 9).

High-dimensionality analysis of stimulated whole blood samples reveals three common sub-populations of activated platelets.

tSNE based dimensionality reduction and visualisation confirmed that the detected alterations in marker expression (i.e.: omitting vehicle treated samples) in whole blood was again not driven by donor or batch effect (Figure 6A). Auto-clustering analysis of the agonist-stimulated only data set, produced 5 clusters (Figure 6B), with 3 clusters (1,2 and 3) accounting for more than 90% of total events. Each formed cluster was not dominated by individual agonists indicating a commonality to these resulting phenotypes. Comparative expression heatmaps with hierarchical dendrograms within the clustering indicated that cluster 2 was characterised by lower expression of CD31, CD63 and CLEC-2, whilst platelets in cluster 3 had lower FSC values, lower levels of GPVI and higher CD107a expression. (Figure 6D).

Discussion

The levels of individual platelet surface receptors, basally and following activation, have been well-characterised. However, improvements in flow cytometry and mass cytometry technology now permit much larger antibody panels for simultaneous measurement of receptors on individual platelets. Through greater immunophenotyping of platelets in health and in disease it is increasingly possible to address the question of whether, and what, platelet sub-populations exist. Here we report a 16-parameter phenotypic approach utilising spectral flow cytometry with computational analysis, including machine learning, to phenotype platelets.

Our selection of surface proteins was based on the phenotypic panel used by Blair *et al*.[8] for mass cytometry. First, we set out to measure the effect of different activation pathways

of platelet activation on the expression of surface markers. A spectral cytometry-based approach, compared to mass cytometry in terms of antibody costs, machine running costs and processing time, permits an expansion in experimental conditions that can be performed in each run. Moreover, we established this assay could be applied to either PRP or whole blood samples. Across the range of conditions tested (agonists, sample type) there were significant increases in PAC-1, CD62P, CD63, CD107a, CD61, and CD29 expression, decreases in CD42b expression, and no changes in CLEC-2 expression. These patterns of surface protein changes were consistent with the current understanding of the field and echo those presented by Blair *et al*.[8] and Hindle *et al*.[38].

These types of initial analyses rely on median fluorescence scores derived from measurements of 10,000 platelets per subject, and comparative variation determined across donors. Dimensionality reduction (tSNE) and clustering provides for greater analytical weight for each individual platelet across all replicates. It also allows for the visualisation of the contribution of each individual subject to the interpreted expression patterns[39]. This approach validated that these observed patterns are shared across donors, confirming these effects were driven by functional responses and not by potential batch variation. Whilst sex has recently been associated with some increased platelet reactivity[40], this shared pattern indicates that there were no obvious sex-based differences in our small group of younger, healthy individuals.

Unlike previous reports[38], based on P-selectin expression and PAC-1-binding, k-means clustering silhouette analysis within our dataset did not support the presence of 4 (or more) distinct sub-populations. We did observe that across all agonists, on activation the cluster of platelets that had the largest drop in GPVI or CD42a/b were not always the same

cluster of platelets that had the highest increase in classic activation markers (CD62P/PAC-1). In our whole blood stimulation experiments, the three common clusters that emerged across all agonists had comparable levels of CD62P expression and PAC-1 binding but differed due to loss of GPVI or CD31. This implies that susceptibility to activation-induced 'shedding' may be a constituent factor to pre-destined sub-populations. However, further characterisation of these potential sub-populations would be required to determine their composition pre-activation.

We next turned to machine learning to construct unbiased algorithms to uncover potential classifications at an individual platelet level and to compare predictive capacities between vehicle- and agonist-treated datasets. Consistent with platelet biology and our traditional user-led population analysis, CD62P expression and PAC-1 binding were consistently identified with agonist stimulation, as was loss of CD42b expression. Interestingly, the weighted importance of CD107a, lysosomal-associated membrane protein 1 (LAMP-1), was only high for PAR-1 activation using TRAP-6. Using the computed weightings, machine learning correctly identified over 89% of platelets within the training data set. When applied to an unseen data set, correct identification was predictably lower but maintained greater than 77% accuracy, rising to 91-92% for PAR-stimulated platelets. The difference in predictive capabilities is likely due to U46619 and ADP being comparatively weaker secondary agonists and therefore producing a less uniform pattern of response[41–44].

We next applied the same machine learning approach to differentiate between young and old platelets within mRNA stain-based, flow-sorted platelet samples which were subsequently vehicle- or agonist-stimulated. Strikingly, there is remarkable consistency

in the highly-weighted parameters across the unstimulated and stimulated samples with FSC-A, CD41, SSC-A, GPVI, CD61, and CD42b featuring prominently.

A significant change in CD41 with platelet age is consistent with previous work from our group looking at the proteomics and transcriptomics of young and old platelets[33], and from others looking at their thrombotic potential[19]. Similarly, an association between GPVI levels and young platelets have been recently reported by Veninga *et al*.[45] in human platelets and by us using a temporal labelling approach in mice[12].

Interestingly, FSC-A, which is considered an approximate indicator of size in flow cytometry, is also highly important in our machine learning algorithm when distinguishing between young and old platelets. One point of contention concerning the field of platelet ageing is whether platelet size changes with age. Although there is some evidence linking mean platelet volume to thrombotic risk[46,47], there is also contradictory data suggesting that the two variables are independent of one another[12,48,49]. The hypothesis that young platelets are larger was originally proposed in the 1960s[50,51], however by the mid-1980s several studies were published reporting no correlation between platelet age and size[49]. Regardless, a caveat of using SYTO-13 dye as a surrogate marker of platelet age is that larger platelets may have more mRNA, and subsequently take up more dye and appear brighter, skewing sorting and subsequent analysis[52]. However, previous work from our group using a similar mRNA dye noted variations in megakaryocyte-derived mRNAs such that young and old platelets would have to vary 32-64-fold in size for such a relationship to hold.[33] Similarly, it is also notable that SSC-A, generally taken as an indicator of internal complexity (i.e. granularity) of a cell, is a highly important discriminator of young versus old platelets. This finding is directly in line with

our observations that platelets lose approximately 50% of their total protein content and mitochondria as they age in the circulation[33]. Similarly, in turn this parameter may also reflect the potential density of a platelet, a measure that studies have suggested is an accurate indicator of platelet age[53–55].

For unseen samples, machine learning-based identification of young and old platelets was 76% accurate in vehicle control samples. This accuracy was maintained in platelets treated with PAR-4 or U46619 but decreased to 67% in platelets treated with ADP. Notably, identification of ADP-stimulated platelets was primarily determined by 4 parameters, whilst for all other conditions 6-7 parameters were used, indicating that perhaps the composition of this panel could be altered for improved accuracy. Conversely, predictive accuracy was higher in samples treated with TRAP-6 or CRP-XL, at 83% and 84%, respectively. In addition, CD62P was more highly-weighted as an additional discriminatory parameter in these samples, which reflects the well-described greater thrombotic potential of young platelets and their higher CD62P degranulation[12,56].

Finally, we applied our machine learning-based identification of young and old platelets back to our healthy platelet populations. This analysis demonstrated that young platelets, as defined by machine learning in each condition, were consistently associated with platelet clusters carrying higher levels of activation markers, in accordance with our earlier reports[12,33]. This suggests that machine learning algorithms employing FSC-A and SSC-A, in addition to surface markers, can be used to further discriminate platelet sub-populations in healthy individuals based on circulatory age.

Existing flow cytometry technology is starting to be replaced by machines incorporating full spectrum measures and algorithm-based unmixing of contributing fluorophores. This

will make larger multi-parameter measurements of samples increasingly accessible to researchers. However, a key question will be the relevance and feasibility for clinical use as multi-parametric measurements of platelets or aggregates potentially offer the capability of more accurate diagnoses or monitoring of platelet disorders and thrombotic risk. As we demonstrate here, this phenotypic approach is suitable for staining and analysing very low volumes of whole blood which we believe are important advantages to overcome the logistical and economical hurdles to more routine clinical use. An alternative approach to achieve this goal is the development of label-free analysis of samples, often based on image analysis[57]. Both approaches offer the capability to identify and differentiate subtle features. Indeed, with companies already developing spectral based imaging flow cytometers it will most likely become possible to apply the two approaches simultaneously to analyse samples. At the heart of each approach is the application of machine learning, which is most powerful when informed by large numbers of samples to account for natural physiological variation. To achieve this will require acquisition of samples across multiple sites which will also necessitate the development of standardised protocols.

In conclusion, we present a 16-parameter, flow cytometry-based assay coupled to powerful bioinformatic approaches to undertake unparalleled, deep phenotyping of platelets and their functionality. The incorporation of machine learning into this workflow provides impartial analysis and predictive capability at a by-platelet and by-individual level. We posit that this approach combining surface and physical markers will be highly valuable in phenotyping platelet sub-populations and studying platelet populations in disease or pathological states.

Funding information

Acknowledgements

Authorship contributions

AV and PCA designed the research, performed the assays, collected data, analysed and interpreted data, performed statistical analysis, and wrote the manuscript. JB, ND and HEA designed the research, analysed and interpreted data, performed statistical analysis, and revised the manuscript. JB performed bioinformatic analyses. CAM and TDW designed the research, analysed and interpreted data, and revised the manuscript.

Disclosure of conflicts of interest

The authors declare no conflicts of interest.

1	Leeksma CHW, Cohen JA. Determination of the Life of Human Blood Platelets using Labelled Diisopropylfluorophosphonate. *Nature*; 1955; **175**: 552–3.

3	Lebois M, Josefsson EC. Regulation of platelet lifespan by apoptosis. *Platelets*; 2016; **27**: 497–504.

3	Quach ME, Chen W, Li R. Mechanisms of platelet clearance and translation to improve platelet storage. *Blood* 2018; **131**: 1512–21.

4	Rivera J, Lozano ML, Navarro-Núñez L, Vicente V. Platelet receptors and signaling in the dynamics of thrombus formation. *Haematologica* 2009; **94**: 700–11.

5	Gibbins JM. Platelet adhesion signalling and the regulation of thrombus formation. *Journal of Cell Science* 2004; **117**: 3415–25.

6	Baaten CCFMJ, ten Cate H, van der Meijden PEJ, Heemskerk JWM. Platelet populations and priming in hematological diseases. *Blood Reviews* 2017; **31**: 389–99.

7	den Dekker E, van Abel M, van der Vuurst H, van Eys GJJM, Akkerman J-WN, Heemskerk JWM. Cell-to-cell variability in the differentiation program of human megakaryocytes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2003; **1643**: 85–94.

8	Blair TA, Michelson AD, Frelinger AL. Mass Cytometry Reveals Distinct Platelet Subtypes in Healthy Subjects and Novel Alterations in Surface Glycoproteins in Glanzmann Thrombasthenia. *Sci Rep* 2018; **8**: 10300.

9	Södergren AL, Ramström S. Platelet subpopulations remain despite strong dual agonist stimulation and can be characterised using a novel six-colour flow cytometry protocol. *Sci Rep* 2018; **8**: 1441.

10	Machlus KR, Italiano JE Jr. The incredible journey: From megakaryocyte development to platelet formation. *Journal of Cell Biology* 2013; **201**: 785–96.

11	Cecchetti L, Tolley ND, Michetti N, Bury L, Weyrich AS, Gresele P. Megakaryocytes differentially sort mRNAs for matrix metalloproteinases and their inhibitors into platelets: a mechanism for regulating synthetic events. *Blood* 2011; **118**: 1903–11.

12	Armstrong PC, Allan HE, Kirkby NS, Gutmann C, Joshi A, Crescente M, Mitchell JA, Mayr M, Warner TD. Temporal in vivo platelet labeling in mice reveals age-dependent receptor expression and conservation of specific mRNAs. *Blood Advances* 2022; **6**: 6028–38.

13	Kienast J, Schmitz G. Flow cytometric analysis of thiazole orange uptake by platelets: a diagnostic aid in the evaluation of thrombocytopenic disorders. *Blood* 1990; **75**: 116–21.

14 Hoffmann JJML. Reticulated platelets: analytical aspects and clinical utility. *Clinical Chemistry and Laboratory Medicine (CCLM)* 2014; **52**: 1107–17.

15 Imperiali CE, Arbiol-Roca A, Sanchez-Navarro L, Dastis-Arias M, Lopez-Delgado JC, Cortes-Bosch A, Sancho-Cerro A, Dot-Bach D. Reference interval for immature platelet fraction on Sysmex XN haematology analyser in adult population. *Biochem Med* 2018; **28**: 0–0.

16 Angénieux C, Maître B, Eckly A, Lanza F, Gachet C, Salle H de la. Time-Dependent Decay of mRNA and Ribosomal RNA during Platelet Aging and Its Correlation with Translation Activity. *PLOS ONE* 2016; **11**: e0148064.

17 Allan HE, Vadgama A, Armstrong PC, Warner TD. What can we learn from senescent platelets, their transcriptomes and proteomes? *Platelets* **34**: 2200838.

18 Thompson CB. Selective consumption of large platelets during massive bleeding. *Br Med J* 1985; **291**: 95–6.

19 McBane RD, Gonzalez C, Hodge DO, Wysokinski WE. Propensity for young reticulated platelet recruitment into arterial thrombi. *J Thromb Thrombolysis* 2014; **37**: 148–54.

20 Allan HE, Vadgama A, Armstrong PC, Warner TD. Platelet ageing: A review. *Thrombosis Research* 2023; **231**: 214–22.

21 Kakouros N, Rade JJ, Kourliouros A, Resar JR. Platelet Function in Patients with Diabetes Mellitus: From a Theoretical to a Practical Perspective. *International Journal of Endocrinology* 2011; **2011**: e742719.

22 Cesari F, Marcucci R, Gori AM, Caporale R, Fanelli A, Casola G, Balzi D, Barchielli A, Valente S, Giglioli C, Gensini GF, Abbate R. Reticulated platelets predict cardiovascular death in acute coronary syndrome patients. Insights from the AMI-Florence 2 Study. *Thromb Haemost* 2013; **109**: 846–53.

23 Ibrahim H, Schutt RC, Hannawi B, DeLao T, Barker CM, Kleiman NS. Association of Immature Platelets With Adverse Cardiovascular Outcomes. *Journal of the American College of Cardiology* 2014; **64**: 2122–9.

24 Schofield H, Rossetto A, Armstrong P, Allan H, Warner T, Brohi K, Vulliamy P. Immature platelet dynamics are associated with clinical outcomes after major trauma. *Journal of Thrombosis and Haemostasis* 2023; in press.

25 Bongiovanni D, Schreiner N, Gosetti R, Mayer K, Angiolillo DJ, Sibbing D, Holdenrieder S, Anetsberger A, von Scheidt M, Schunkert H, Laugwitz K-L, Schüpke S, Kastrati A, Fegers-Wustrow I, Bernlochner I. Immature Platelet Fraction Predicts Adverse Events in Patients With Acute Coronary Syndrome: the ISAR-REACT 5 Reticulated Platelet Substudy. *Arterioscler Thromb Vasc Biol* 2023; **43**: e83–93.

26 Armstrong PC, Hoefer T, Knowles RB, Tucker AT, Hayman MA, Ferreira PM, Chan MV, Warner TD. Newly Formed Reticulated Platelets Undermine Pharmacokinetically Short-Lived Antiplatelet Therapies. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2017; **37**: 949–56.

27 Guthikonda S, Alviar CL, Vaduganathan M, Arikan M, Tellez A, DeLao T, Granada JF, Dong J-F, Kleiman NS, Lev EI. Role of Reticulated Platelets and Platelet Size Heterogeneity on Platelet Activity After Dual Antiplatelet Therapy With Aspirin and Clopidogrel in Patients With Stable Coronary Artery Disease. *Journal of the American College of Cardiology* 2008; **52**: 743–9.

28 Bernlochner I, Goedel A, Plischke C, Schüpke S, Haller B, Schulz C, Mayer K, Morath T, Braun S, Schunkert H, Siess W, Kastrati A, Laugwitz K-L. Impact of immature platelets on platelet response to ticagrelor and prasugrel in patients with acute coronary syndrome. *Eur Heart J* 2015; **36**: 3202–10.

29 Perl L, Lerman-Shivek H, Rechavia E, Vaduganathan M, Leshem-Lev D, Zemer-Wassercug N, Dadush O, Codner P, Bental T, Battler A, Kornowski R, Lev EI. Response to prasugrel and levels of circulating reticulated platelets in patients with ST-segment elevation myocardial infarction. *J Am Coll Cardiol* 2014; **63**: 513–7.

30 Frelinger AL, Spurgeon BEJ. Clinical Cytometry for Platelets and Platelet Disorders. *Clinics in Laboratory Medicine* 2023; **43**: 445–54.

31 Njemini R, Onyema OO, Renmans W, Bautmans I, De Waele M, Mets T. Shortcomings in the Application of Multicolour Flow Cytometry in Lymphocyte Subsets Enumeration. *Scandinavian Journal of Immunology* 2014; **79**: 75–89.

32 Spurgeon BEJ, Frelinger III AL. Platelet Phenotyping by Full Spectrum Flow Cytometry. *Current Protocols* 2023; **3**: e687.

33 Allan HE, Hayman MA, Marcone S, Chan MV, Edin ML, Maffucci T, Joshi A, Menke L, Crescente M, Mayr M, Zeldin DC, Armstrong PC, Warner TD. Proteome and functional decline as platelets age in the circulation. *Journal of Thrombosis and Haemostasis* 2021; **19**: 3095–112.

34 Dai Y, Xu A, Li J, Wu L, Yu S, Chen J, Zhao W, Sun XJ, Huang J. CytoTree: an R/Bioconductor package for analysis and visualization of flow and mass cytometry data. *BMC Bioinformatics* 2021; **22**: 1–20.

35 Ellis B, Haal P, Hahne F, Meur NL, Gopalakrishnan N, Spidlen J, Jiang M, Finak G, Granjeaud S. flowCore: flowCore: Basic structures for flow cytometry data. Bioconductor version: Release (3.18); 2023.

36 Ashhurst TM, Marsh-Wakefield F, Putri GH, Spiteri AG, Shinko D, Read MN, Smith AL, King NJC. Integration, exploration, and analysis of high-dimensional single-cell cytometry data using Spectre. *Cytometry Part A* 2022; **101**: 237–53.

37 Kuhn M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 2008; **28**: 1–26.

38 Hindle MS, Spurgeon BEJ, Cheah LT, Webb BA, Naseem KM. Multidimensional flow cytometry reveals novel platelet subpopulations in response to prostacyclin. *Journal of Thrombosis and Haemostasis* 2021; **19**: 1800–12.

39 Spurgeon BEJ, Michelson AD, Frelinger AL. Platelet mass cytometry: Optimization of sample, reagent, and analysis parameters. *Cytometry A* 2021; **99**: 170–9.

40 Chan MV, Chen M-H, Thibord F, Nkambule BB, Lachapelle AR, Grech J, Schneider ZE, Wallace de Melendez C, Huffman JE, Hayman MA, Allan HE, Armstrong PC, Warner TD, Johnson AD. Factors that modulate platelet reactivity as measured by 5 assay platforms in 3429 individuals. *Research and Practice in Thrombosis and Haemostasis* 2024; **8**: 102406.

41 Bednar B, Condra C, Gould RJ, Connolly TM. Platelet aggregation monitored in a 96 well microplate reader is useful for evaluation of platelet agonists and antagonists. *Thrombosis Research* 1995; **77**: 453–63.

42 Nakahata N. Thromboxane A2: Physiology/pathophysiology, cellular signal transduction and pharmacology. *Pharmacology & Therapeutics* 2008; **118**: 18–35.

43 Armstrong PCJ, Truss NJ, Ali FY, Dhanji AA, Vojnovic I, Zain ZNM, Bishop-bailey D, Paul-clark MJ, Tucker AT, Mitchell JA, Warner TD. Aspirin and the in vitro linear relationship between thromboxane A2-mediated platelet aggregation and platelet production of thromboxane A2. *Journal of Thrombosis and Haemostasis* 2008; **6**: 1933–43.

44 Li Z, Zhang G, Le Breton GC, Gao X, Malik AB, Du X. Two Waves of Platelet Secretion Induced by Thromboxane A2 Receptor and a Critical Role for Phosphoinositide 3-Kinases*. *Journal of Biological Chemistry* 2003; **278**: 30725–31.

45 Veninga A, Handtke S, Aurich K, Tullemans BME, Brouns SLN, Schwarz SL, Heubel-Moenen FCJI, Greinacher A, Heemskerk JWM, van der Meijden PEJ, Thiele T. GPVI expression is linked to platelet size, age, and reactivity. *Blood Advances* 2022; **6**: 4162–73.

45 Brækkan SK, Mathiesen EB, NjøLstad I, Wilsgaard T, StøRmer J, Hansen JB. Mean platelet volume is a risk factor for venous thromboembolism: the Tromsø Study, Tromsø, Norway. *Journal of thrombosis and haemostasis* 2010; **8**: 157–62.

47 Thattaliyath B, Cykowski M, Jagadeeswaran P. Young thrombocytes initiate the formation of arterial thrombi in zebrafish. *Blood* 2005; **106**: 118–24.

47 Thompson CB, Love DG, Quinn PG, Valeri CR. Platelet size does not correlate with platelet age. *Blood* 1983; **62**: 487–94.

48 Thompson CB, Jakubowski JA, Quinn PG, Deykin D, Valeri CR. Platelet size and age determine platelet function independently. *Blood* 1984; **63**: 1372–5.

49 Karpatkin S. Heterogeneity of human platelets. I. Metabolic and kinetic evidence suggestive of young and old platelets. *The Journal of clinical investigation* 1969; **48**: 1073–82.

51 Karpatkin S. Heterogeneity of human platelets. II. Functional evidence suggestive of young and old platelets. *The Journal of clinical investigation* 1969; **48**: 1083–7.

51 Joutsi-Korhonen L, Sainio S, Riikonen S, Javela K, Teramo K, Kekomäki R. Detection of reticulated platelets: estimating the degree of fluorescence of platelets stained with thiazole orange. *European journal of haematology* 2000; **65**: 66–71.

53 Mezzano D, Hwang K, Catalano P, Aster RH. Evidence that platelet buoyant density, but not size, correlates with platelet age in man. *Am J Hematol* 1981; **11**: 61–76.

54 van Oost BA, Timmermans AP, Sixma JJ. Evidence that platelet density depends on the alpha-granule content in platelets. *Blood* 1984; **63**: 482–5.

55 Corash L, Shafer B, Perlow M. Heterogeneity of human whole blood platelet subpopulations. II. Use of a subhuman primate model to analyze the relationship between density and platelet age. *Blood* 1978; **52**: 726–34.

56 Peng J, Friese P, Heilmann E, George JN, Burstein SA, Dale GL. Aged Platelets Have an Impaired Response to Thrombin as Quantitated by P-Selectin Expression. *Blood* 1994; **83**: 161–6.

57 Zhou Y, Yasumoto A, Lei C, Huang C-J, Kobayashi H, Wu Y, Yan S, Sun C-W, Yatomi Y, Goda K. Intelligent classification of platelet aggregates by agonist type. Janus ED, Walczak AM, editors. *eLife* eLife Sciences Publications, Ltd; 2020; **9**: e52938.

Tables

Table 1: Predictive efficacy of machine learning of stimulated platelets.

	ADP	TRAP-6	U46619	PAR-4	CRP-XL
Training set accuracy	0.92	0.94	0.89	0.95	0.95
Unseen set accuracy	0.80	0.90	0.78	0.87	0.89

Table 2: Predictive efficacy of machine learning of young and older platelets.

	Vehicle	TRAP-6	PAR-4	CRP-XL	ADP	U46619
Training set accuracy	0.86	0.87	0.86	0.86	0.90	0.86
Unseen set accuracy	0.76	0.78	0.77	0.76	0.76	0.74

Figure legends

Figure 1: Changes in individual surface marker expression in resting and activated platelets. Platelets were incubated with vehicle (PBS), ADP (30µM), TRAP-6 (10µM), U46619 (10µM), PAR-4 (100µM) or CRP-XL (3µM) for 20 minutes at 37ºC. Results expressed as: (A) raw MFI ±SEM were analysed using a one-way ANOVA with a Dunnett test to correct for multiple comparisons; (B) a heatmap based on MFI values; and (C) a heatmap based on MFI log fold-changes. All data obtained from n=20-21 volunteers.

Figure 2: T-distributed Stochastic Neighbour Embedding (tSNE) visualisation of concatenated samples. Plots calculated for combined datasets (n=20-21) of platelets incubated with (columns) vehicle (PBS), ADP (30µM), TRAP-6 (10µM), U46619 (10µM), PAR-4 (100µM) or CRP-XL (3µM) for 20 minutes at 37ºC. Each plot individually coloured to reflect intensity (low=blue, high=red) for each of the measured 16 parameters.

Figure 3: Marker weighting from machine learning predications to distinguish between platelets incubated with vehicle and agonists. (A) ADP (30µM), (B) TRAP-6 (10µM), (C) U46619 (10µM), (D) PAR-4 (100µM), and (E) CRP-XL (3µM). (Left) Visual comparison of logicle transformed expression data and (right) marker weighting listed in order of importance with gradient overlay representing most important=red, least important=white. (N=20-21 for all).

Figure 4: Marker weighting from machine learning predications to differentiate between "young" and "old" platelets incubated (20 minutes at 37ºC) with (A) vehicle (PBS), (B) ADP (30µM), (C) TRAP-6 (10µM), (D) U46619 (10µM), (E) PAR-4 (100µM), and (F) CRP-XL (3µM). (Left) Visual comparison of logicle transformed expression data and (right) marker weighting listed in order of importance with gradient overlay representing most important=red, least important=white. (N=8 for all).

Figure 5: Interrogation of machine learning predicted 'young' and 'old' platelets within PRP samples incubated with vehicle or agonists. Platelets incubated (20 minutes at 37ºC) with (A) vehicle (PBS), (B) ADP (30µM), (C) TRAP-6 (10µM), (D) U46619 (10µM), (E) PAR-4 (100µM), and (F) CRP-XL (3µM). (Left) tSNE division of concatenated datasets coloured with predicted 'young' (teal) or 'old' (red). (Middle panel) Violin plot of probability value for each event divided in clusters per condition, as identified by markers weighted by machine learning at above importance. (Right) Heatmap breakdown of relative marker expression within each cluster (low=white, high=purple). (N=20-21 for all).

Figure 6: Identification of common sub-populations within whole blood incubated with agonists (20 minutes at 37ºC) whole blood. tSNE visualisation of concatenated data coloured by (A) donor or (B) identified cluster. (C) Relative proportion of clusters and contribution by agonist (red: ADP; olive: CRP-XL; green: PAR-4; blue: TRAP-6; purple: U46619). (D) Comparative expression heatmaps with hierarchical dendrograms of identified clusters. (N=6 for all).