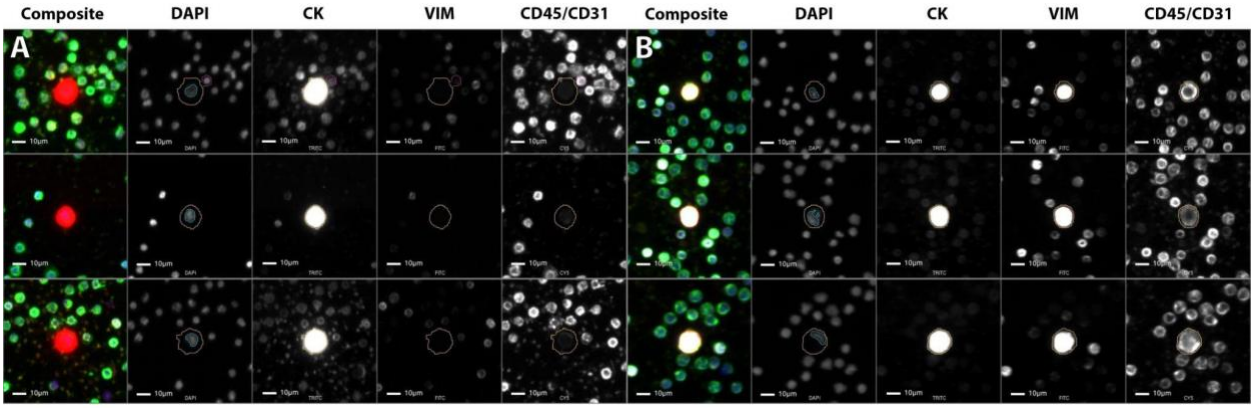
Supplemental Information

Supplemental Figure 1: Representative images of control cell lines used for reagent validation.	2
Preliminary CD41 Stain	2
Supplemental Figure 2: Gallery of rare acellular events found in the PB of Post-COVID patients using the CD41 platelet stain.	2
Image Derived Liquid Biopsy Event Space and Centroids	3
Classification Model Training Phase	3
Supplemental Table 1: Classification model performance for Post-COVID, Post-SYMP COVID and NDs after removing all cancer patients.	3

Supplemental Figure 1: Representative images of control cell lines used for reagent validation. Normal blood donor peripheral blood samples spiked with A) SKBR3 (ATCC HTB-30) or B) HPAEC (ATCC PCS-100-022) cells. Contrived samples were used to confirm performance accuracy and reproducibility of the IF assays in each batch stained. Images provided are taken at 100x magnification. Composite image consists of Blue: DAPI, Red: Cytokeratin, White: Vimentin, Green: CD45/CD31.



Preliminary CD41 Stain

We hypothesised that the CD31/45 positive acellular events found using the Landscape assay constitute platelets. As described in the Methods section, an additional patient slide (PACS-C005) was stained with a custom stain, replacing Vimentin with CD41, a platelet specific marker, and removing CD31. The results found elongated acellular events that are CD41 positive and CD45 negative (Supplemental Figure 2). The morphology both in the channel positivity and the brightfield images are similar to that observed in the Landscape assay (Figure 2B). These preliminary results support the hypothesis that the elongated acellular events consist of platelets and may be platelet structures.

Supplemental Figure 2: Gallery of rare acellular events found in the PB of Post-COVID patients using the CD41 platelet stain. Blue: DAPI, Red: CK, White: CD41, Green: CD45. Images taken at 100x magnification. Scale bar = 10 µm.

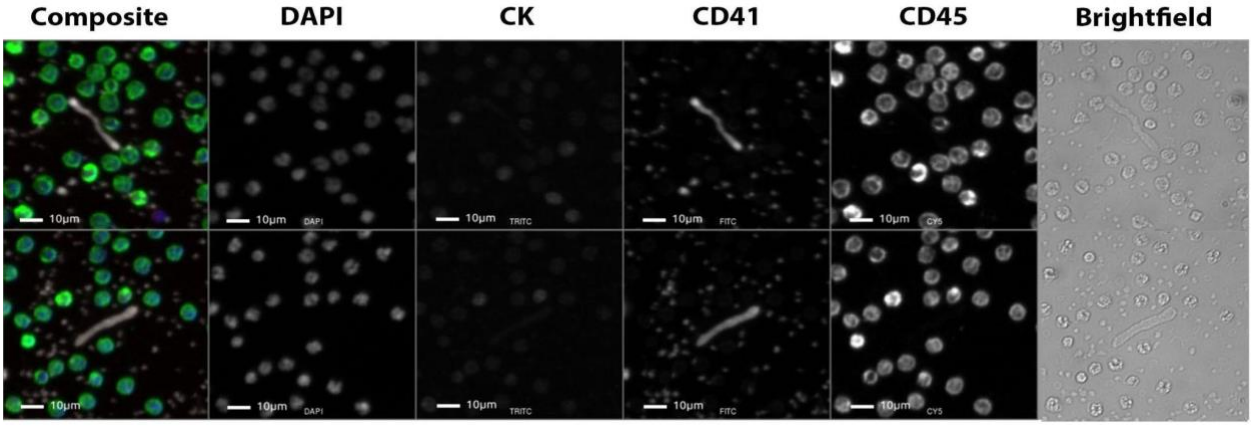


Image Derived Liquid Biopsy Event Space and Centroids

All cells from 1870 previously scanned slides were segmented using the "EBImage" R package (EBImage_4.12.2) and 761 morphometrics were obtained for each cell. Then, principal component analysis was performed on the morphometrics and the top 350 components, according to explained variance, were kept. Subsequently, a clustering algorithm was employed on the reduced morphometric representation of all the cells to identify the centroids. First, for each cell, the number of neighboring cells within a PCA distance of 12 was calculated. Then, cells with the highest number of neighboring cells were identified as centroids and their neighboring cells were removed from the centroid cell candidates. This process is repeated until all cells belong to a centroid or constitute a centroid. With the parameters used, the algorithm results in ~30,000 centroid cells in the space. The centroids in this space are detected cells and cells in their vicinity share characteristics as they have similar morphological parameters. An identical procedure was applied for constructing the image derived event space and identifying centroids for acellular events. This resulted in ~22,000 acellular centroids.

Utilising the morphometric derived space and the generated centroids, a centroid can be algorithmically assigned to any event of interest. This is done by obtaining the morphometrics and then using the PCA transform, used to generate the space, in order to bring the cell to its respective representation, or position. Then, the centroid, which is closest to the event's position, is assigned as a label to the event. This label is a way to discreetly characterise the morphological parameters of an event and constitutes a nearest centroid classification model. In this way, we expand beyond manual event classifications and are able to automatically assign a classification to events, providing a reproducible and scalable framework for characterisation of liquid biopsy analytes.

Classification Model Training Phase

The training of the classifier algorithm consists of multiple steps. First, the number of clusters that the cell centroids will be grouped into has to be determined. To that end, a classifier is built and evaluated on the training set for a variable number of clusters. The number of clusters is chosen such that the misclassification error is minimised. This is done separately for DAPI positive and DAPI negative events. After the number of clusters of centroid cells is determined, the input data is converted to the cluster number format for DAPI positive and DAPI negative events respectively. Subsequently, the input data is combined and set as input to a new random forest model. This is done by taking the combined model, ordering the centroid cell clusters or features according to importance, and assessing its error rate for considering a variable number of features. To quantify the feature importance, the mean decrease in impurity (MDI) metric is used. This process starts from the highest importance feature and ends on the least important feature or centroid cluster. Ultimately, the final model is built by using the most important centroid clusters both from the DAPI positive and DAPI negative events in a single random forest classification model.

Supplemental Table 1: Classification model performance for Post-COVID, Post-SYMP COVID and NDs after removing all cancer patients. Averages and standard errors are reported for model accuracy, sensitivity, and specificity. Due to small dataset sizes, only cohort comparisons with more than 19 samples in each class are reported.

Cohort 1	Cohort 2	Accuracy	Sensitivity	Specificity
Normal Donors	Post-COVID	92·5%±1·6%	93·7%±1·8%	80·0%±8·2%
Normal Donors	Post-SYMP COVID	89·7%±1·0%	90·4%±1·5%	85·0%±6·7%
Normal Donors	PACS Suspected	90·9%±1·5%	92·6%±1·6%	82·5%±6·5%