ToF-SIMS and machine learning for single-pixel molecular discrimination of an acrylate polymer microarray

Wil Gardner[1,2,3], Andrew L. Hook[4], Morgan R. Alexander[4], Davide Ballabio[5], Suzanne M. Cutts[2], Benjamin W. Muir[3] and Paul J. Pigram[1*].

1   Centre for Materials and Surface Science and Department of Chemistry and Physics, La Trobe University, Melbourne, Victoria, Australia

2   La Trobe Institute for Molecular Sciences, La Trobe University, Melbourne, Victoria, Australia

3   CSIRO Manufacturing, Clayton, Victoria, Australia

4   Advanced Materials and Healthcare Technologies, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK

5   Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano-Bicocca,

        P.zza della Scienza 1, 20126, Milano, Italy

* Corresponding Author: p.pigram@latrobe.edu.au

ABSTRACT: Combinatorial approaches to materials discovery offer promising potential for the rapid development of novel polymer systems. Polymer microarrays enable the high-throughput correlation of physical and chemical properties, such as surface chemistry, with polymer functionality, such as cell or protein adsorption. A limitation to this approach is the ability to accurately discriminate between highly similar polymers or identify heterogeneities within each individual polymer spot. Time-of-flight secondary ion mass spectrometry (ToF-SIMS) offers unique potential in this regard, capable of describing the chemistry associated with the outermost layer of a sample with high spatial resolution and chemical sensitivity. However, this comes at the cost of generating large scale, complex hyperspectral imaging data. We have demonstrated previously that machine learning is a powerful tool for interpreting ToF-SIMS images, describing a method for color-tagging the output of a self-organizing map (SOM). This reduces the entire hyperspectral data set to a single reconstructed color similarity map, in which the spectral similarity between pixels is represented by their color similarity. Here, we apply the same methodology to a ToF-SIMS image of a printed polymer microarray. We report complete, single-pixel molecular discrimination of the 70 unique polymer spots in the array, while also identifying intra-spot heterogeneities thought to be related to polymer orientation. In this way, we show that the SOM can identify layers of similarity and clusters in the data, both with respect to polymer backbone structures and their individual side groups. Finally, we relate the output of the SOM analysis with fluorescence data from polymer-protein adsorption studies, highlighting how polymer functionality can be visualized within the context of the global topology of the data set.

Time-of-flight secondary ion mass spectrometry (ToF-SIMS) is an attractive technique for studying surface chemistry at the molecular scale, with a depth resolution of a few nanometers and excellent spatial resolution. ToF-SIMS has a broad range of applications for both organic and inorganic materials, including, chemical mapping of cells and tissue[1-9], studying stress corrosion cracking in metal alloys and flexible optoelectronic devices[10, 11] and characterizing chemically similar polymeric materials. [12-18] The data produced by contemporary ToF-SIMS instruments is inherently hyperspectral, characterized by both x and y spatial information, with a mass spectrum associated with each pixel. When spatial information is not relevant to the analyst, the data can be summed to produce a single 1D spectrum containing the cumulative mass spectral information from across the analysis area. Alternatively, spatial coordinates can be used to produce 2D chemical maps of the surface, in which the distribution of intensities from a single peak, mass segment or group of peaks is visualized as an image. The latter approach, typically termed ToF-SIMS imaging, can provide unique insights into how particular components are distributed on the surface. For example, in the field of bioimaging, a recent study investigated the distribution of zinc in the hippocampi of healthy and brain injured rats. [8] Other recent work has focused on imaging nanoparticles in cells and tissues, both organic[5] and inorganic. [9, 19]

The size and complexity of ToF-SIMS imaging data are significant challenges impeding complete and robust interpretation. A single ToF-SIMS spectrum can contain hundreds or sometimes thousands of identifiable mass peaks. This equates to the same number of individual molecular ion maps. Each map will typically contain tens of thousands of pixels, each of which could potentially contain analytically important information. Since it is not feasible to compare hundreds of images, or to analyze thousands of individual spectra, the use of

multivariate analysis (MVA) is essential and has proven invaluable in the accurate interpretation of ToF-SIMS hyperspectral images.

Our previous work[20-25] has focused on applying ToF-SIMS and MVA in the study of printed polymer microarrays. Hook et al. [26] provide a comprehensive review of the development, principles, applications and advantages of polymer microarray systems for high-throughput materials discovery. Briefly, polymer microarrays are prepared by printing a series of polymer spots onto glass slides, with diameters typically in the range of hundreds of micrometers. [20, 21, 24, 26] Such arrays have been successfully used to discover novel materials with useful biological properties, such as improved stem cell attachment[27-29] or resistance to bacterial biofilm formation. [30, 31] Once printed, the polymers can be subjected to a suite of characterization techniques, such as XPS, Raman spectroscopy, ToF-SIMS and water contact angle (WCA) measurements. [26] The microarray thereby provides a platform for the high-throughput characterization of polymer materials. Beyond materials discovery, the large number of material-biological interactions that can be assessed using this sample format makes it possible to identify structure-function relationships by correlating properties such as surface chemistry with functionality; for example protein adsorption or polymer wettability. ToF-SIMS, in combination with MVA, has been particularly valuable in this regard. [22-24]

Initially, MVA was used as a tool for evaluating the surface chemistries of different polymer and copolymer materials, based on their summed 1D ToF-SIMS spectra. For example, Urquhart et al. [22] used principal components analysis (PCA) to analyze ToF-SIMS spectra from a 576 spot copolymer microarray. The study demonstrated important insights into the similarities between many of the polymers, highlighting specific fragments that were identified as commonalities. The authors also demonstrated the use of partial least-squares (PLS) regression analysis for correlating ToF-SIMS spectra with surface wettability, via WCA measurements. Similar results were obtained by Celiz et al. [20] who, in addition to predicting the WCA of polymers based on their surface chemistry, identified correlations between ToF-SIMS data and stem cell adhesion. Analogous findings are reported by Yang et al. [23] regarding the adhesion of human embryoid body cells to acrylate polymers with differing surface chemistries.

In addition to analyzing 1D spectra, recent research has focused on using MVA to evaluate ToF-SIMS hyperspectral images of polymer microarrays. Hook et al.[24] investigated the surface chemistry of 70 different poly(meth)acrylate spots in a printed microarray, using ToF-SIMS imaging and multivariate curve resolution (MCR). Similar to the studies involving 1D spectra, MCR revealed correlations between polymers using their associated spectra. However, in this case spectra were extracted from individual pixels, such that spatial information was retained. MCR scores were then used to reconstruct images of the surface, providing visualization of the spectral similarities between different spots in the array. Thereby, correlations related to both the polymer backbones and the extending side groups were revealed.

Recently, we have demonstrated the power of machine learning for interpreting ToF-SIMS data, using a particular artificial neural network architecture known as Kohonen[32] self-organizing maps (SOMs). The design and function of the SOM in general is described elsewhere[32-34] and in our previous works. [13-17, 35] For brevity, a detailed mathematical description of the underlying algorithm of the SOM is omitted here.

SOMs are designed to reduce the dimensionality of a high dimensional dataset, usually producing a 2D model of the data. A key feature of the SOM is the ability to preserve the *topology* of the data during the non-linear mapping to a lower dimensional space. SOMs are supposed to mimic the action of biological networks of neurons, which share information and thus evolve over time. Neurons are typically organized in a square 2D space and associated with a vector of weights. These weights are then updated iteratively on the basis of the input samples. Similar samples activate topologically close neurons in the network, and consequently SOMs are expected to learn and model patterns from the input data.

We employ a toroidal topology for the SOM, in which neurons located on the edges of the lattice are neighbors to those on the opposite edge, forming a continuous surface in all directions. This is topologically equivalent to the surface of a toroid. Critically, toroidal SOMs have been shown to be superior in performance to planar SOMs.[36] Specifically, planar SOMs suffer from boundary effects associated with neurons on the edges of the lattice, due primarily to these neurons having fewer neuron neighbors than those in the center of the lattice. [37, 38]

SOMs are better suited to handling non-linear ToF-SIMS data than traditional linear approaches, such as PCA or MCR. [14-17] Recently, our group has presented a novel approach for applying toroidal SOMs in the analysis of ToF-SIMS hyperspectral images. [35] The output of the SOM is a 2D topology-preserving map. In order to visualize this topology intuitively as a reconstructed image, it is useful to overlay a coloring scheme onto the SOM. In this way, the neurons are colored according to their positions, such that adjacent neurons are assigned a similar color. The color-tagging is designed to match the toroidal topology, such that the color change is continuous across the boundaries of the SOM. Pixels in the original image can then be colored according to their winning neurons on the SOM, which enables the reconstruction of a single, color similarity map of the analysis area in which pixels are colored according to their spectral similarity. This process is described in more detail in the Experimental Methods section, and in our recent work. [35]

Here, we present the application of color-tagged SOMs for the analysis of a ToF-SIMS hyperspectral image of a polymer microarray. Specifically, we use the same data set used for MCR analysis by Hook et al.,[24] albeit processed differently. Results from this work are discussed and evaluated with two primary outcomes. First, we focus on the technique itself, demonstrating the power and robustness of the SOM for visualizing complex ToF-SIMS hyperspectral imaging data. Second, we provide strong justification for the use of ToF-SIMS and color-tagged SOMs in the analysis of polymer microarrays, specifically. In this regard, we show how ToF-SIMS and SOMs together offer unique potential for the continued advancement of combinatorial materials discovery using polymer microarrays.

EXPERIMENTAL METHODS

Microarray printing. Polymer microarray printing has been described elsewhere generally, and for this array specifically. [24] Briefly, epoxy-functionalized glass slides (Genetix) were dip coated in 4% (w/v) poly(hydroxy ethylmethacrylate) (pHEMA) (Sigma) in ethanol. Polymerization solution – composed of 75% (v/v) monomer (Sigma, dissolved in dimethylformamide
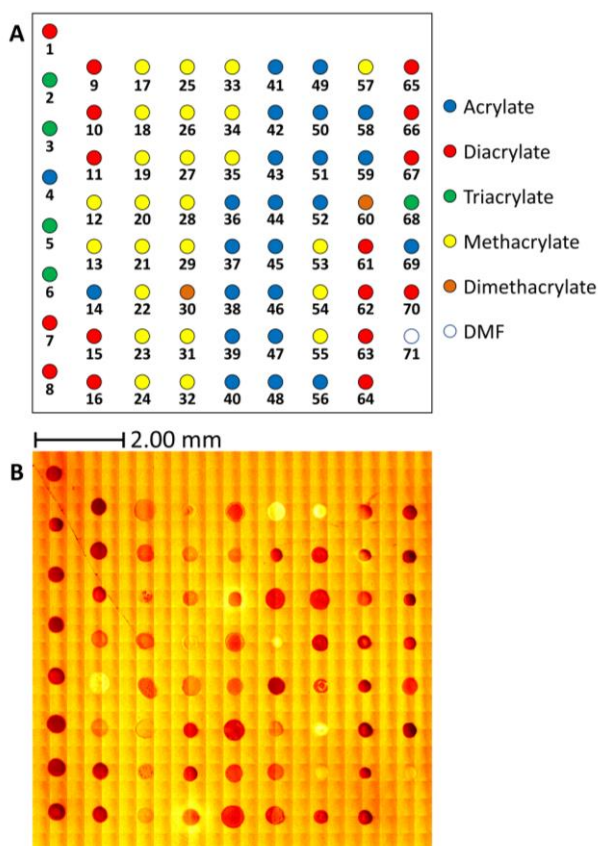
Figure 1. (A) Schematic showing the layout of the 70 polymers in the microarray, in which each polymer is colored according to the broadly defined polymer groupings provided on the right. Spot 71 contained only the solvent, DMF. (B) Total ion image of the 9.2×9.2 mm analysis area obtained using negative ion mode. Figure adapted, with permission, from Hook et al. [24]

(DMF)) and 1% (w/v) photoinitiator 2,2-dimethoxy-2-phenylacetophenone – was printed onto the pHEMA-coated slides, using an XYZ3200 dispensing workstation (Biodot). Slides were irradiated with a long wave UV source for 30 s after printing each material, then for a further 10 min following completion of the array. Finally, the slides were vacuum extracted at <50 mTorr for 7 days. The CAS number and chemical name for each of the 71 spots is included in the Supplementary Information (Table S-1).

Protein adsorption and fluorescence microscopy. Protein adsorption and fluorescence microscopy are described in Hook et al. [24] Briefly, polymer microarrays were immersed in a solution of 25 μg/ml tetramethylrhodamine isothiocyanate labeled albumin (Sigma) in phosphate buffered saline (Gibco; pH 7.4), then incubated for 1 h at 37 °C under stagnate conditions. Arrays were then washed for 1 min with ultrapure water, blotted and dried overnight. A Genepix fluorescence scanner was used for imaging, with a 532 nm laser and 5 μm pixel size. Images were captured before and after protein adsorption, to account for background fluorescence. The same procedure was applied to a control array – without any added protein – to quantify fluorescence. A total of eight replicates were measured.

ToF-SIMS experimental. Hook et al. [24] describe the protocol used to acquire the ToF-SIMS data using an IONTOF (GmbH) TOF.SIMS 4 instrument. In brief, a 25 keV Bi_3^+ primary ion source was scanned over a 9.2×9.2 mm analysis area, employing the macroraster scanning functionality. Negative ion mode was used, and a single scan performed using 15 pulses per pixel, with a pixel size of 10×10 μm. A low energy (~20 eV) electron flood gun was used for charge compensation.

Data preprocessing, export and analysis using SOMs. Peaks were selected automatically using the spatially-summed spectrum with the peak search function in the SurfaceLab6 software package. A minimum threshold of >100 counts was used over the entire mass range, producing a peak list comprising 717 mass peaks. Individual ion images were then exported as text files, compiled into a single data matrix and normalized to total ion intensity per pixel using in-house MATLAB scripts, as described previously.[35] SOM training was performed using the Kohonen and CP-ANN Toolbox[33, 34] in MATLAB. A squared, toroidal topology was used for all analyses, as well as the batch training approach, whereby all samples are introduced to the SOM simultaneously. The analysis-specific data scaling procedures, as well as SOM parameters used, are described in more detail in the Results and Discussion section.

SOM visualization and analysis. The SOM output was visualized using the color-tagging approach described in our previous work. [35] All steps were performed using in-house MATLAB functions. Briefly, the neurons of the toroidal SOM were mapped to the surface of a 3D torus from their positions within the 2D plane. The equations describing this mapping are presented in Gardner et al. [35] Neurons were then colored using a red-green-blue (RGB) color scheme according to their x-, y- and z-coordinates on the torus, respectively. For image reconstruction, pixels were assigned the color of their winning neuron, producing a similarity map in which similar pixels (according to their mass spectra) appear as a similar color.

RESULTS AND DISCUSSION

Background subtraction. SOM analysis of the ToF-SIMS image was split into two parts: background pixel subtraction and sample training. Background subtraction was implemented due to the large size of the data set (920×920 pixels, with 717 mass peaks), with the goal of reducing computation time by the removal of irrelevant data. The background subtraction workflow is illustrated by the flow chart in Fig 2 and visualized using the polymer microarray image in Fig 3. A 10 x 10 neuron rectangular SOM was initialized using the eigenvalue approach, whereby the neuron weights were assigned based on the eigenvectors associated with the first two principal components of the data set. [34] With this approach, data were initially clustered based on a linear PCA solution. Hence, we opted to first scale the ToF-SIMS data using Poisson scaling, which has been shown to improve PCA performance by accounting for Poisson noise. [39, 40] This scaling attempts to account for heteroscedastic noise by transforming the data into an alternate space in which the uncertainty in the data is more uniform. [39, 40] The SOM was trained for a nominal single epoch following initialization to produce the SOM output. Pixels representative of the background were then selected with regions of interest (ROIs), using the similarity map produced by the color-tagged SOM as a
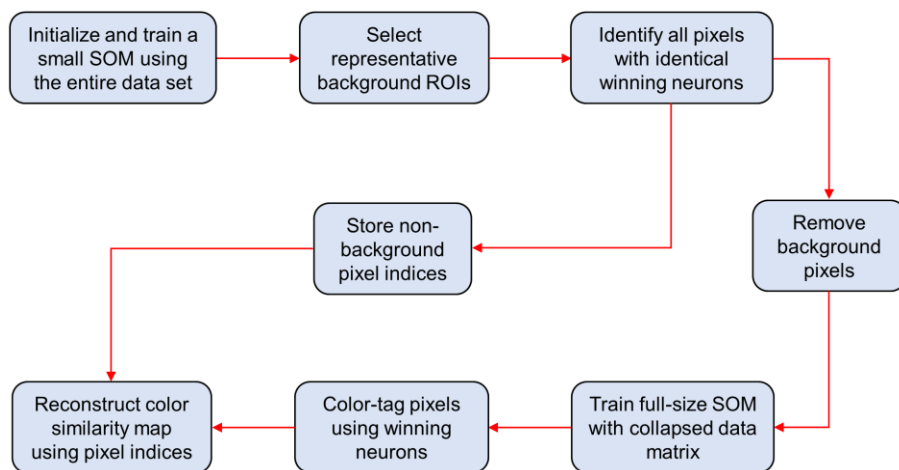
Figure 2. Flowchart outlining the data processing and analysis workflow. Note that in this work eigenvalue initialization of the small SOM was used in the first step. This was sufficient to distinguish background pixels.
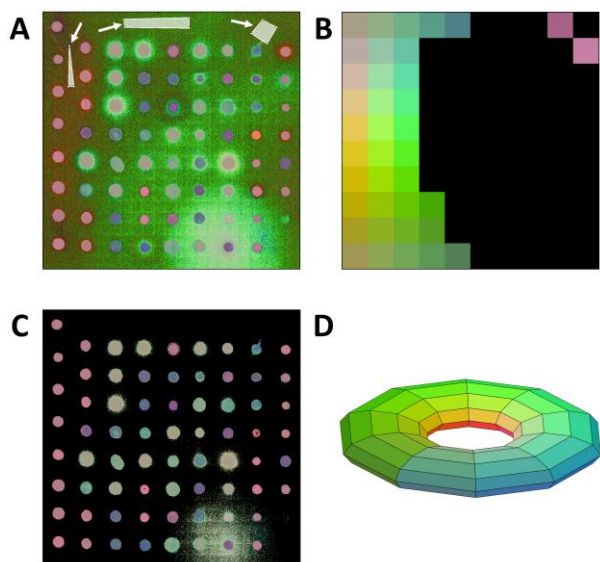

Figure 3. Applying the data processing workflow presented in Figure 2 to the ToF-SIMS image of the polymer microarray. (A) Selection of three polygon ROIs representative of the entire set of background pixels, indicated by the white arrows. (B) The localization of the winning neurons associated with the pixels selected in (A) on the SOM, shown in black. This represents the cluster of neurons that are modelling the set of background pixel spectra. (C) Subtraction of background pixels sharing the winning neurons displayed in (B). (D) The origin of the neuron colors, which are based on their positions in the toroidal SOM. RGB colors are proportional to the x-, y- and z-coordinates of the neurons on the 3D toroidal surface, respectively.

guide (Fig 3A). Pixels within the ROIs were surveyed for their winning neurons, producing a list of neurons corresponding to the substrate (represented in black on the SOM in Fig 3B). Finally, all pixels sharing these neurons were identified in the original image and removed from the data matrix. Matrix indices corresponding to the location of non-background pixels were stored for future image reconstruction. Fig 3C shows the reconstructed image after background subtraction, while Fig 3D

highlights the origin of the neuron colors according to their position on a 3D toroid.

This workflow has the same analytical goals as the manual selection of ROIs for background removal, however is more objective, automated, data-oriented and not restricted to shapes that can be drawn manually. For example, background pixels were identified dispersed within the chemical leaching visible from spot 56, which would be impossible to locate and remove manually. This has the advantage of removing experimenter bias and avoiding the loss of important information or interesting pixels. The identification of background pixels is determined by their spectral similarity, and is based on the complete mass spectrum at each pixel rather than the intensity of an individually selected peak, set of peaks or total ion intensity. Here, we opted to use only the eigenvalue initialization of the SOM for background identification, as the polymer materials were clearly linearly separable from the substrate and this was the least time-consuming approach. However, it would be equally feasible to train the SOM for the required number of epochs to reach convergence, which would be necessary when there is a high level of spectral similarity between the sample and substrate. This would still reduce computation time, as a much smaller SOM can be used to complete a broad classification of background and sample pixels. As we demonstrate, a larger SOM can then be trained using the collapsed data matrix. It should be noted that caution is required during background subtraction to avoid removing useful data by the misclassification of pixels. Hence, it is recommended that this approach should only be used when the approximate layout of the image is known. In this case the dimensions and layout of the microarray were clearly defined.

SOM visualization and analysis. Following background subtraction, the remaining pixels were used to train a new 40×40 neuron SOM, in this instance for the 200 training epochs required to reach convergence. Using the stored background pixel indices, it was then possible to reconstruct the original image using the SOM output. Specifically, non-background pixels were assigned the RGB colors of their winning neurons, then reinserted into their original positions in the image according to their stored pixel indices. Background pixels were colored
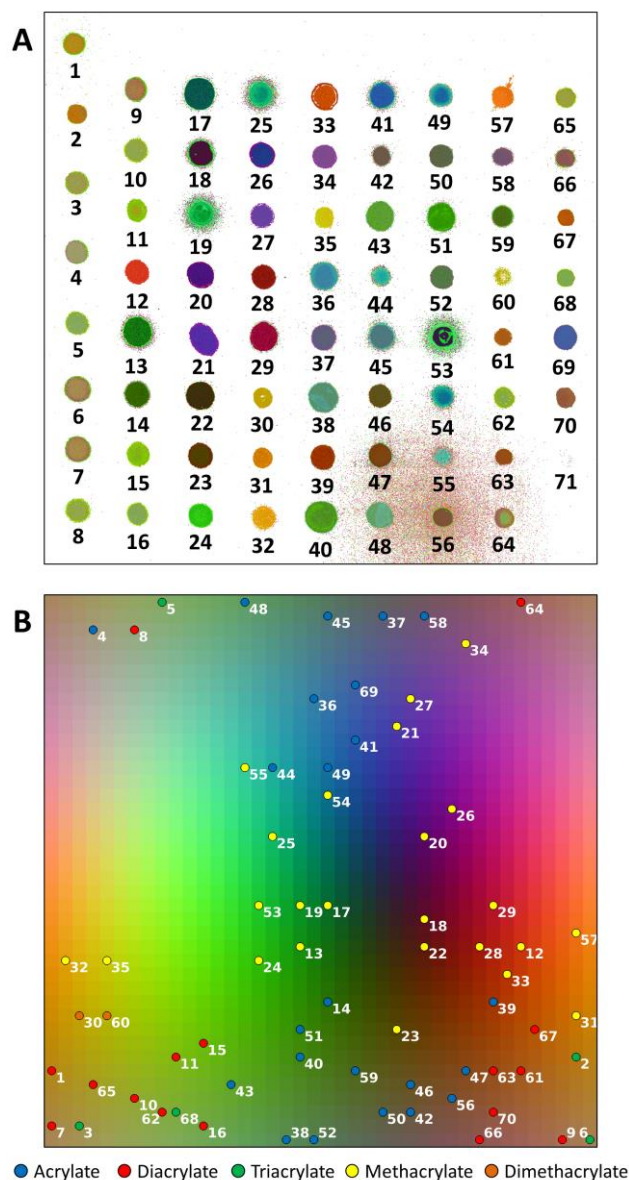
🔵 Acrylate 🔴 Diacrylate 🟢 Triacrylate 🟡 Methacrylate 🟠 Dimethacrylate

Figure 4. (A) Reconstructed RGB similarity map after background subtraction and subsequent training with a 40×40 SOM. Pixels are colored by their winning neuron colors, such that spectral similarity between pixels is captured by their color similarity. White pixels represent background-subtracted pixels. (B) Localization of the most abundant winning neuron from each of the 70 polymers on the color-tagged SOM. Numbers correspond to the polymer numbers shown in (A), and the datapoint colors represent to the broad polymer grouping scheme shown below.

white, producing a color similarity map of non-background pixels only. This image is presented in Fig 4A. Note that Poisson scaling was not used for this SOM. Instead, we used the same preprocessing steps that we have reported previously, [13-17, 35] which have consistently delivered positive analytical outcomes from the resulting SOM.

Qualitatively, it is clear from Fig 4A that the SOM successfully discriminated many of the polymers based on their mass spectra, as evidenced by their respective coloring. Furthermore, most of the polymers exhibited high color uniformity,

suggesting insensitivity to spectral noise and chemical homogeneity within each spot. Polymer spots showing variation in assigned color were those in which there appeared to be real heterogeneities across the spot (for example, spots 38, 48 and 53), likely associated with drying effects, incomplete/inconsistent polymerization, or some other phenomenon. The SOM also revealed chemical leaching from several spots, especially from polymer 56. Note that the DMF spot (71) was almost entirely absent after background subtraction, suggesting minimal difference in composition relative to the substrate, at least according to the PCA-based separation. Hence this spot was excluded from further chemical analyses.

Figure 4B shows the locations of the most abundant winning neurons from each of the 70 polymers, plotted on the color-tagged SOM. Briefly, each polymer was selected individually using the polygon ROI select tool in MATLAB (polygon selections are displayed in Fig S-1 in Supporting Information). For each polymer, a histogram of winning neuron frequency was then calculated, and the most abundant neuron was selected and plotted on the color-tagged SOM. Each data point was then colored by the polymer backbone categories, namely acrylates, diacrylates, triacrylates, methacrylates and dimethacrylates. These categories are shown in Fig 4B, as well as in Fig 1A.

Figure 4B clearly illustrates the capabilities of the SOM in distinguishing each of the 70 polymers, at least with regards to their most abundant neurons. Only the top neurons are shown so that global performance can be visualized in a single image. However, the individual winning neuron compositions of all the polymers are also presented separately in Fig S-2 in the Supporting Information. Together, these figures demonstrate that the SOM distinguished most, if not all, of the polymers on the basis of their underlying surface chemistry revealed by ToF-SIMS.

It is important to reiterate that this result was achieved using an entirely unsupervised approach, including the peak selection process. The only information that was required was the approximate layout of the array, for background subtraction. Hence, Figs 4A-B demonstrate the separation of 70 unique polymer structures without any prior knowledge of their surface chemistry. This exemplifies the power of the SOM in interpreting complex ToF-SIMS image data with minimal user input or intervention. According to Fig 4B, polymers in the array were generally well clustered with regard to their backbone structure. In particular, the acrylates, methacrylates and dimethacrylates appeared separately clustered to well defined regions on the SOM, whereas the diacrylates and triacrylates were clustered together. There also appeared to be ordering on the SOM reflecting the side group moieties present within each polymer. For example, spots 21, 27, 34, 37, 42, 58 and 69 all contained aromatic functional groups (specifically phenol or benzyl) and are all adjacent on the SOM in Fig B. The only other polymer in the array in this category is spot 63 which, despite not being adjacent, also appeared in a similar region. This is strong evidence that the SOM recognized the chemical commonality with the side groups and weighted the neurons accordingly. Looking more closely at the chemical structure of each polymer and their corresponding weights reveals this explicitly.

Fig S-3 in the Supporting Information shows the mean weights (normalized to total peak weighting) associated with each of these 8 polymer spots. The weights can be used to infer which mass peaks were most important in distinguishing each

polymer, according to the SOM algorithm. Note that the corresponding normalized mean spectra from each polymer are included in Fig S-4 in the Supporting Information. The $C_6H_6O^-$ ion, relating to the phenol group, was the top weighted fragment in spots 27, 34 and 37, and ranked 8th in 58 (in which the top-ranking weight is the similar phenolate ion, $C_6H_5O^-$). These polymers all contain phenol moieties situated at the end of their side groups, comprising a phenyl group with attached oxygen atom. Hence, the release of either of these ions is expected given this structure. [41] Similarly, the top weighted peak for both 21 and 69 was $C_7H_8O^-$, which is the fragmented benzyloxy group characteristic of the side groups in these polymers. The benzoate ion, $C_7H_5O_2^-$, was the 2nd highest ranked weight for spot 63, which again is in direct agreement with the structure of this polymer. Finally, the phenyl group in spot 42 is part of a larger nonylphenyl functional group, containing a characteristic 9-carbon tail. The corresponding nonylphenol fragment ion, $C_{15}H_{24}O^-$, was ranked as the 15th highest weight for this polymer. This lower ranking is consistent with the increased size of this moiety, which is expected to fragment more readily into its constituents.

It is remarkable that, despite the different aromatic fragment ions produced by these eight polymers, they were still mostly grouped together on the SOM. This suggests that there were other, more subtle similarities in their spectra that correlated with aromatic groups. Interestingly, this does not seem to be the phenyl anion itself, $C_6H_5^-$, which was ranked at 6th, 203rd, 129th, 103rd, 174th, 149th, 3rd and 7th for spots 21, 27, 34, 37, 42, 58, 63 and 69, respectively. It is more likely that the similarities detected by the SOM are distributed across a range of mass peaks – including many that do not exclusively originate from the aromatic groups – which would otherwise be a significant challenge or practically impossible to identify manually. As a reference, table S-2 in the Supporting Information shows 25 top weighted mass fragments corresponding to each spot.

The ability of the SOM to detect subtle difference between individual pixels is exemplified in Fig 5. Here, it is clear that the SOM identified two distinct regions within spot 38, as highlighted in Fig 5A. This image was used as a guide to select a set of representative pixels from each region (Fig 5B). After ROI selection, pixels were surveyed for their winning neurons, similar to the process described earlier used to produce Fig 4. However, rather than only considering the most abundant neuron in each ROI, the neurons accounting for > 95% of the pixel count were identified and included in the analysis. This approach was used to avoid including any outlier pixels in the selection, which could have skewed the mean weights. Fig 5C shows the distribution of pixels with these same winning neurons across the entire analysis area, demonstrating the precision of the SOM clustering with almost all of these pixels localized within polymer 38. This is further emphasized in the close-up of the spot, which reveals that the SOM clearly distinguished ROI 1 from ROI 2, with minimal shared neurons. This is explored further in Fig 5D, in which the SOM has been shaded in greyscale according to neuron abundance in each ROI, based on the corresponding pixel histograms. Visualizing the distribution on the SOM is useful to gauge the relatedness of the two ROIs.

Here it is obvious that, while clearly distinct, ROI 1 and ROI 2 were highly similar with slight overlaps. This is also visibly evident in the weights extracted from each ROI (Fig 5E and Table 1).

Table 1 shows the top 16 weights extracted from each ROI – according to their normalized weightings – and their respective peak assignments. Also shown are the weights in the adjacent region, as well as the ratio between the two regions. The top weights provide insight into the mass peaks that best represented the underlying chemical composition of each region. In ROI 1, all of the weights shown correspond well with the chemical structure of the polymer, isodecyl acrylate. Further, the ratios of these weights between ROI 1 and ROI 2 were relatively consistent, varying between 0.72 and 1.36. This is also true for 13 of the top 16 weights in ROI 2. However, the remaining three weights were much higher than in ROI 2, exhibiting ratios between 4.00 to 9.24. Furthermore, these fragment ions appear highly related and are seemingly indicative of the acrylate backbone. This suggests that the SOM identified differing orientations of the monomers between ROI 1 and ROI 2. Specifically, an increase in signal from the polymer backbone implies that the side groups in ROI 2 were hidden in the polymer bulk. This would indicate phase separation between the two regions, likely arising due to hydrophobic interactions between the long aliphatic chains of the monomers. This is further supported by the decreased abundance of highly weighted oxygen-containing fragments in ROI 1 and the corresponding increase in hydrocarbons, which together provide strong evidence that the aliphatic side group was more exposed in ROI 1 than in ROI 2. The differences between ROI 1 and ROI 2 in this regard are subtle, and only become apparent when considering the set of weights together. Nevertheless, by visualizing these differences using the color-tagged SOM, the two regions are clearly distinguishable by their unique color compositions. This exemplifies the ability of this approach to reveal minor chemical difference between highly similar regions. More importantly, this is achieved using an entirely unsupervised workflow, requiring no prior knowledge of the sample composition.

Previous work has demonstrated how polymer microarrays can be used for high-throughput materials discovery, by correlating multivariate ToF-SIMS data with univariate functionality metrics, such as protein adsorption, surface wettability or cell proliferation rates. [20, 23, 24, 41] These studies have been successful in identifying individual mass fragments that either positively or negatively correlated with a given indicator, using partial least squares (PLS) regression. However, the relationship between ToF-SIMS data *topology* and functionality has not been thoroughly explored. This was the motivation for Fig 6, which presents protein adsorption levels for each of the 70 polymers in the microarray in the context of their topological arrangement. Fig 6 was constructed in a similar fashion to Fig 4A, by considering the most abundant neuron associated with each polymer. However, rather than coloring the plot using the polymer categories, a heat map was used to display protein fluorescence and hence adsorption levels. Exploring the relationship between data topology and functionality in this way is
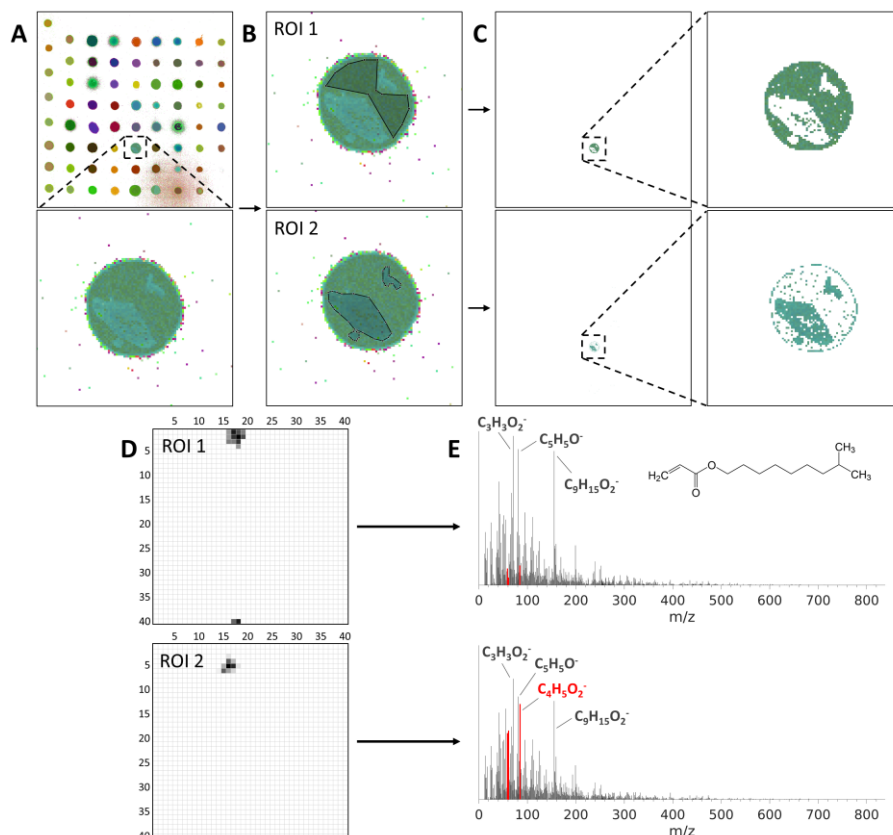
Figure 5. Identification of intra-spot surface chemistry heterogeneity using the color-tagged SOM. (A) Reconstructed RGB similarity map of the polymer microarray, with a close-up of polymer spot 38 (isodecyl acrylate) illustrating the presence of two distinct regions. (B) Polygon selection of ROI 1 and ROI 2. (C) Distribution of the winning neurons in ROI 1 and ROI 2 across the entire analysis area, indicating localization within spot 38. Also shown is a close-up of spot 38, highlighting the separation of ROI 1 and ROI 2. (D) SOM localization for winning neurons located with ROI 1 and ROI 2, colored using a grey scale that represents relative neuron abundance within the ROI (with white equal to zero). (E) Weights extracted from each ROI using the winning neurons presented in (D), presented as a weighted mean (based on neuron abundance) and normalized to total weighting for each ROI. The labels show some of the prominently weighted fragments originating from the polymer structure, and the red bars refer to three of the major differences in peak weightings between the two ROI.

Table 1. Top weighted peaks from each ROI in spot 38 (isodecyl acrylate) with peak assignments and weighting ratios.

| Region	Peak m/z	Assignment	Weight in ROI 1 (%)	Weight in ROI 2 (%)	Ratio	Region	Peak m/z	Assignment	Weight in ROI 2 (%)	Weight in ROI 1 (%)	Ratio
	71.01	$C_3H_3O_2^-$	2.91	2.35	1.24		71.01	$C_3H_3O_2^-$	2.35	2.91	0.81
	81.04	$C_5H_5O^-$	2.65	2	1.32		81.04	$C_5H_5O^-$	2.00	2.65	0.76
	155.09	$C_9H_{15}O_2^-$	2.62	1.92	1.36		155.09	$C_9H_{15}O_2^-$	1.92	2.62	0.73
	41.01	C_2HO^-	2.02	1.69	1.19		85.04	**$C_4H_5O_2^-$**	1.86	0.38	**4.92**
	67.02	$C_4H_3O^-$	1.8	1.42	1.27		55.02	$C_3H_3O^-$	1.77	1.28	1.38
	95.05	$C_6H_7O^-$	1.42	1.16	1.22		41.01	C_2HO^-	1.69	2.02	0.84
	43.02	$C_2H_3O^-$	1.39	1.27	1.1		67.02	$C_4H_3O^-$	1.42	1.8	0.79
ROI 1	51.02	$C_4H_3^-$	1.36	1.16	1.17	ROI 2	61.04	**$C_2H_5O_2^-$**	1.34	0.15	**9.24**
	111.05	$C_6H_7O_2^-$	1.33	1.17	1.14		59.02	**$C_2H_3O_2^-$**	1.29	0.32	**4.00**
	69.04	$C_4H_5O^-$	1.32	1.05	1.25		43.02	$C_2H_3O^-$	1.27	1.39	0.91
	55.02	$C_3H_3O^-$	1.28	1.77	0.72		111.05	$C_6H_7O_2^-$	1.17	1.33	0.88
	49.01	C_4H^-	1.19	1.13	1.06		95.05	$C_6H_7O^-$	1.16	1.42	0.82
	38.02	$C_3H_2^-$	1.06	1.00	1.06		49.01	C_4H^-	1.13	1.19	0.95
	79.02	$C_5H_3O^-$	1.04	0.91	1.14		69.04	$C_4H_5O^-$	1.05	1.32	0.8
	13.01	CH^-	1.03	0.89	1.16		25.01	C_2H^-	1.04	1.22	0.85
	53.04	$C_4H_5^-$	1.00	0.79	1.27		45.00	CHO_2^-	1.00	0.83	1.2
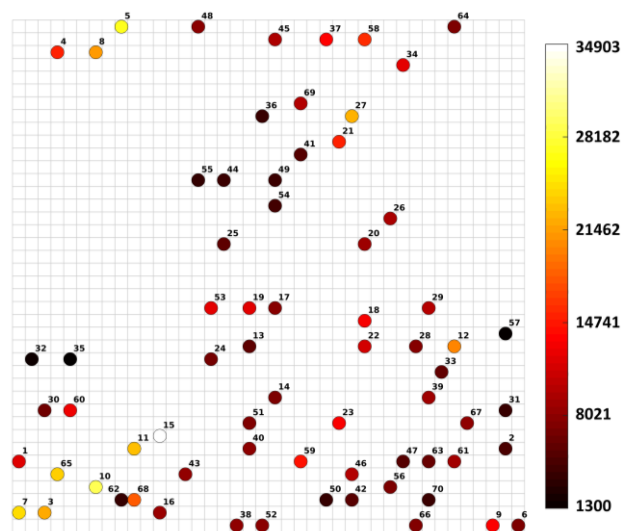
Figure 6. Localization of the most abundant winning neuron from each of the 70 polymers on the color-tagged SOM, colored with a heat map that represents the mean protein binding (n = 8) for each polymer. The color bar on the right shows the arbitrary fluorescence scale.

useful, as it reveals global trends in the influence of surface chemistry. This contrasts with techniques such as PLS, which provides information regarding which *specific* variables correlate with functionality. To demonstrate this, consider the bottom and top left of Fig 6. This region appears to be heavily populated with highly functional polymers, which suggests common underlying mechanisms or properties that aid in protein binding. It is also useful to investigate polymers that appear to be highly similar according to their mass spectra, yet exhibit vastly different protein binding. For example, polymers 10, 68 and 62 are adjacent on the SOM, yet were correlated with high, moderate and low protein binding, respectively.

The advantages of the SOM are complementary to those of PLS, and hence combining these two approaches in the analysis is likely to provide unique insights into structure-function correlations. For example, applying PLS to this same dataset, Hook et al. [24] showed that $C_8H_3O_4^-$ and $C_6H_3O^-$ ions correlated with high protein adsorption. It was then argued that these ions originated from the acrylate backbone and therefore that the backbone itself could promote protein adsorption. With regards to the SOM analysis, for spots 3, 5, 7, 10, 11, and 65 – which all exhibited moderate to high protein adsorption and are all clustered on the SOM – $C_8H_3O_4^-$ was ranked as the 2nd, 3rd, 5th, 2nd, 17th and 2nd highest weighted fragment, respectively. These results are consistent with the PLS results.

It is intriguing that the same fragment ion is also ranked as the 5th highest weight in the adjacent spot 62, which is correlated with poor protein adsorption. The weights associated with spot 62, comprising ethylene (glycol) diacrylate, are indeed very similar to those of the adjacent high-adsorbing spots. Using PLS analysis, Hook et al. presented the top fragment ions that correlated – either positively or negatively – with protein adsorption. We considered how highly these fragments were weighted in spot 62, in comparison to the high-adsorbing spots. Interestingly, none of these peaks distinguished spot 62 from the other spots, either alone or in combination. This suggests that spot 62 did not fit well with the PLS results, which is not surprising considering its location on the SOM.

We also considered the weights of spot 27, which exhibited moderate protein adsorption yet was distinctly separated from the other high-adsorbing spots on the SOM (Fig 6). Contrary to the other spots, $C_8H_3O_4^-$ was ranked as the 207th highest weight for spot 27. However, $C_6H_3O^-$ was ranked as 11th, whereas this fragment ranks between 17th – 34th for the other spots. Together, these results suggest that, while polymers 3, 5, 7, 10, 11, 27 and 65 all exhibited moderate to high protein adsorption, the binding mechanism for polymer 27 may be different than that of the other six polymers.

It is noteworthy that polymer 27 contains a phenol group, which distinguishes it from the other high-adsorbing polymers. Furthermore, there appears to be a cluster of moderate protein binding associated with most of the other polymers containing aromatic functional groups (21, 34, 37, 42 and 58, as discussed earlier). Hence, this could suggest a weak correlation between the presence of an aromatic functional group and protein adsorption, although this is only a qualitative assessment. Regardless, it is clear that the SOM provides a unique perspective for investigating correlations between groups of polymers and their functionality that is not possible with PLS.

CONCLUSION

The study reported here builds on our previously published work, demonstrating the ability of color-tagged SOMs to interpret complex ToF-SIMS hyperspectral images in an unsupervised manner. The advantages in using a polymer microarray for this purpose are twofold: first, this type of system offers a unique opportunity for investigating the capabilities of the SOM. A broad range of surface chemistries can be included on a single chip, and hence captured within a single ToF-SIMS image. This removes uncertainties associated with preparing multiple samples, as well as when acquiring data under slightly different instrument conditions. Further, the underlying composition of each spot is known, which enables direct comparison between ground truth information and the output from the SOM. As we have demonstrated, this is useful to clarify whether the SOM is revealing accurate information about the polymer surface chemistries, and hence demonstrate the suitability of this approach for more complex samples with an unknown chemical distribution.

Color-tagged SOMs were demonstrably able to produce unique analytical outcomes, with respect to other multivariate approaches. The methodology described offers an intuitive approach for visualizing complex relationships between the spectra of individual pixels. We have confirmed the accuracy of the output by considering not only how the polymers were clustered with regards to their polymer backbone, but also with respect to their individual side groups. Specifically, each of the 70 polymers were successfully discriminated which, considering the high similarity in the spectra produced by many of the polymers, is a nontrivial task. Further, polymers were clearly clustered on a global scale on the basis of their backbone structures, while side groups were influencing their local arrangement.

One of the most important applications of polymer microarrays is the identification of correlations between surface chemistry and polymer functionality. This has been demonstrated using PLS paired with ToF-SIMS. [22, 23, 25, 43] We have shown how SOMs can be used to visualize the functionality of individual polymers within the context of the topological relatedness of their surface chemistries. This provides a unique perspective that is complementary to the information provided by PLS and similar techniques – rather than focusing on specific fragment

ions, the SOM can be used to identify entire subspaces in high dimensional space that correlate with high or low functionality. Further, polymers with surface chemistries that are adjacent topologically, yet exhibit entirely different properties, can be identified and studied. Coupled with PLS, this can provide the necessary information to unravel the complicated relationship between surface chemistry and functionality.

ASSOCIATED CONTENT

Supporting Information

Chemical names and CAS numbers for the polymer spots; ROIs used for analyzing spots; individual localization and abundance of winning neurons on the SOM for each spot; normalized intensities extracted from polymers with aromatic functional groups; and the top rankings from the same aromatic-containing polymers (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: p.pigram@latrobe.edu.au.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

ACKNOWLEDGMENT

REFERENCES

1. Colliver, T. L.; Brummel, C. L.; Pacholski, M. L.; Swanek, F. D.; Ewing, A. G.; Winograd, N., Atomic and Molecular Imaging at the Single-Cell Level with TOF-SIMS. *Anal Chem* **1997,** *69* (13), 2225-2231.

2. Pacholski, M. L.; Cannon Jr., D. M.; Ewing, A. G.; Winograd, N., Static time-of-flight secondary ion mass spectrometry imaging of freeze-fractured, frozen-hydrated biological membranes. *Rapid Communications in Mass Spectrometry* **1998,** *12* (18), 1232-1235.

3. Brulet, M.; Seyer, A.; Edelman, A.; Brunelle, A.; Fritsch, J.; Ollero, M.; Laprevote, O., Lipid mapping of colonic mucosa by cluster TOF-SIMS imaging and multivariate analysis in cftr knockout mice. *Journal of lipid research* **2010,** *51* (10), 3034-45.

4. Passarelli, M. K.; Winograd, N., Lipid imaging with time-of-flight secondary ion mass spectrometry (ToF-SIMS). *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2011,** *1811* (11), 976-990.

5. Kokesch-Himmelreich, J.; Woltmann, B.; Torger, B.; Rohnke, M.; Arnhold, S.; Hempel, U.; Muller, M.; Janek, J., Detection of organic nanoparticles in human bone marrow-derived stromal cells using ToF-SIMS and PCA. *Analytical and bioanalytical chemistry* **2015,** *407* (16), 4555-65.

6. Passarelli, M. K.; Newman, C. F.; Marshall, P. S.; West, A.; Gilmore, I. S.; Bunch, J.; Alexander, M. R.; Dollery, C. T., Single-Cell Analysis: Visualizing Pharmaceutical and Metabolite Uptake in Cells with Label-Free 3D Mass Spectrometry Imaging. *Anal Chem* **2015,** *87* (13), 6696-6702.

7. Castellanos, A.; Ramirez, C. E.; Michalkova, V.; Nouzova, M.; Noriega, F. G.; Fernndez-Lima, F., Three dimensional secondary ion mass spectrometry imaging (3D-SIMS) of Aedes aegypti ovarian follicles. *J. Anal. At. Spectrom.* **2019,** *34* (5), 874-883.

8. Dowlatshahi Pour, M.; Ren, L.; Jennische, E.; Lange, S.; Ewing, A. G.; Malmberg, P., Mass spectrometry imaging as a novel approach to measure hippocampal zinc. *J. Anal. At. Spectrom.* **2019,** *34* (8), 1581-1587.

9. Henss, A.; Otto, S.-K.; Schaepe, K.; Pauksch, L.; Lips, K. S.; Rohnke, M., High resolution imaging and 3D analysis of Ag nanoparticles in cells with ToF-SIMS and delayed extraction. **2018,** *13* (3), 03B410.

10. Sierros, K. A.; Morris, N. J.; Ramji, K.; Cairns, D. R., Stress–corrosion cracking of indium tin oxide coated polyethylene terephthalate for flexible optoelectronic devices. *Thin Solid Films* **2009,** *517* (8), 2590-2595.

11. Zhou, L. F.; Liu, Z. Y.; Wu, W.; Li, X. G.; Du, C. W.; Jiang, B., Stress corrosion cracking behavior of ZK60 magnesium alloy under different conditions. *International Journal of Hydrogen Energy* **2017,** *42* (41), 26162-26174.

12. Hook, A. L.; Scurr, D. J., ToF-SIMS analysis of a polymer microarray composed of poly(meth)acrylates with C6 derivative pendant groups. **2016,** *48* (4), 226-236.

13. Madiona, R. M. T.; Winkler, D. A.; Muir, B. W.; Pigram, P. J., Optimal machine learning models for robust materials classification using ToF-SIMS data. *Applied Surface Science* **2019,** *487*, 773-783.

14. Madiona, R. M. T.; Bamford, S. E.; Winkler, D. A.; Muir, B. W.; Pigram, P. J., Distinguishing Chemically Similar Polyamide Materials with ToF-SIMS Using Self-Organizing Maps and a Universal Data Matrix. *Anal Chem* **2018,** *90* (21), 12475-12484.

15. Madiona, R. M. T.; Welch, N. G.; Russell, S. B.; Winkler, D. A.; Scoble, J. A.; Muir, B. W.; Pigram, P. J., Multivariate analysis of ToF-SIMS data using mass segmented peak lists. *Surface and Interface Analysis* **2018,** *50* (7), 713-728.

16. Madiona, R. M. T.; Winkler, D. A.; Muir, B. W.; Pigram, P. J., Effect of mass segment size on polymer ToF-SIMS multivariate analysis using a universal data matrix. *Appl. Surf. Sci.* **2019,** *478*, 465-477.

17. Welch, N. G.; Madiona, R. M. T.; Payten, T. B.; Jones, R. T.; Brack, N.; Muir, B. W.; Pigram, P. J., Surface Adsorbed Antibody Characterization Using ToF-SIMS with Principal Component Analysis and Artificial Neural Networks. *Langmuir* **2016,** *32* (34), 8717-8728.

18. Madiona, R. M. T.; Alexander, D. L. J.; Winkler, D. A.; Muir, B. W.; Pigram, P. J., Information content of ToF-SIMS data: Effect of spectral binning. *Applied Surface Science* **2019,** *493*, 1067-1074.

19. Leo, B. F.; Fearn, S.; Gonzalez-Cater, D.; Theodorou, I.; Ruenraroengsak, P.; Goode, A. E.; McPhail, D.; Dexter, D. T.; Shaffer, M.; Chung, K. F.; Porter, A. E.; Ryan, M. P., Label-Free Time-of-Flight Secondary Ion Mass Spectrometry Imaging of Sulfur-Producing Enzymes inside Microglia Cells following Exposure to Silver Nanowires. *Analytical chemistry* **2019,** *91* (17), 11098-11107.

20. Celiz, A. D.; Hook, A. L.; Scurr, D. J.; Anderson, D. G.; Langer, R.; Davies, M. C.; Alexander, M. R., ToF-SIMS imaging of a polymer microarray prepared using ink-jet printing of acrylate monomers. *Surface and Interface Analysis* **2013,** *45* (1), 202-205.

21. Scoutaris, N.; Hook, A. L.; Gellert, P. R.; Roberts, C. J.; Alexander, M. R.; Scurr, D. J., ToF-SIMS analysis of chemical heterogenities in inkjet micro-array printed drug/polymer formulations. *Journal of Materials Science: Materials in Medicine* **2012,** *23* (2), 385-391.

22. Urquhart, A. J.; Taylor, M.; Anderson, D. G.; Langer, R.; Davies, M. C.; Alexander, M. R., TOF-SIMS Analysis of a 576

Micropatterned Copolymer Array To Reveal Surface Moieties That Control Wettability. *Analytical Chemistry* **2008,** *80* (1), 135-142.

23. Yang, J.; Mei, Y.; Hook, A. L.; Taylor, M.; Urquhart, A. J.; Bogatyrev, S. R.; Langer, R.; Anderson, D. G.; Davies, M. C.; Alexander, M. R., Polymer surface functionalities that control human embryoid body cell adhesion revealed by high throughput surface characterization of combinatorial material microarrays. *Biomaterials* **2010,** *31* (34), 8827-8838.

24. Hook, A. L.; Williams, P. M.; Alexander, M. R.; Scurr, D. J., Multivariate ToF-SIMS image analysis of polymer microarrays and protein adsorption. **2015,** *10* (1), 019005.

25. Epa, V. C.; Hook, A. L.; Chang, C.; Yang, J.; Langer, R.; Anderson, D. G.; Williams, P.; Davies, M. C.; Alexander, M. R.; Winkler, D. A., Modelling and Prediction of Bacterial Attachment to Polymers. **2014,** *24* (14), 2085-2093.

26. Hook, A. L.; Anderson, D. G.; Langer, R.; Williams, P.; Davies, M. C.; Alexander, M. R., High throughput methods applied in biomaterial development and discovery. *Biomaterials* **2010,** *31* (2), 187-198.

27. Celiz, A. D.; Smith, J. G.; Patel, A. K.; Hook, A. L.; Rajamohan, D.; George, V. T.; Flatt, L.; Patel, M. J.; Epa, V. C.; Singh, T.; Langer, R.; Anderson, D. G.; Allen, N. D.; Hay, D. C.; Winkler, D. A.; Barrett, D. A.; Davies, M. C.; Young, L. E.; Denning, C.; Alexander, M. R., Discovery of a Novel Polymer for Human Pluripotent Stem Cell Expansion and Multilineage Differentiation. *Advanced materials (Deerfield Beach, Fla.)* **2015,** *27* (27), 4006-12.

28. Mei, Y.; Saha, K.; Bogatyrev, S. R.; Yang, J.; Hook, A. L.; Kalcioglu, Z. I.; Cho, S. W.; Mitalipova, M.; Pyzocha, N.; Rojas, F.; Van Vliet, K. J.; Davies, M. C.; Alexander, M. R.; Langer, R.; Jaenisch, R.; Anderson, D. G., Combinatorial development of biomaterials for clonal growth of human pluripotent stem cells. *Nature materials* **2010,** *9* (9), 768-78.

29. Zhang, R.; Mjoseng, H. K.; Hoeve, M. A.; Bauer, N. G.; Pells, S.; Besseling, R.; Velugotla, S.; Tourniaire, G.; Kishen, R. E. B.; Tsenkina, Y.; Armit, C.; Duffy, C. R. E.; Helfen, M.; Edenhofer, F.; de Sousa, P. A.; Bradley, M., A thermoresponsive and chemically defined hydrogel for long-term culture of human embryonic stem cells. *Nature Communications* **2013,** *4* (1), 1335.

30. Hook, A. L.; Chang, C. Y.; Yang, J.; Atkinson, S.; Langer, R.; Anderson, D. G.; Davies, M. C.; Williams, P.; Alexander, M. R., Discovery of novel materials with broad resistance to bacterial attachment using combinatorial polymer microarrays. *Advanced materials (Deerfield Beach, Fla.)* **2013,** *25* (18), 2542-7.

31. Hook, A. L.; Chang, C. Y.; Yang, J.; Luckett, J.; Cockayne, A.; Atkinson, S.; Mei, Y.; Bayston, R.; Irvine, D. J.; Langer, R.; Anderson, D. G.; Williams, P.; Davies, M. C.; Alexander, M. R., Combinatorial discovery of polymers resistant to bacterial attachment. *Nature biotechnology* **2012,** *30* (9), 868-875.

32. Kohonen, T., Self-organized formation of topologically correct feature maps. *Biological Cybernetics* **1982,** *43* (1), 59-69.

33. Ballabio, D.; Consonni, V.; Todeschini, R., The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems* **2009,** *98* (2), 115-122.

34. Ballabio, D.; Vasighi, M., A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems* **2012,** *118*, 24-32.

35. Gardner, W.; Cutts, S. M.; Muir, B. W.; Jones, R. T.; Pigram, P. J., Visualizing ToF-SIMS Hyperspectral Imaging Data Using Color-Tagged Toroidal Self-Organizing Maps. *Analytical chemistry* **2019,** *91* (21), 13855-13865.

36. Mount, N.; Weaver, D., Self-organizing maps and boundary effects: quantifying the benefits of torus wrapping for mapping SOM trajectories. *Pattern Analysis and Applications* **2011,** *14* (2), 139-148.

37. Andreu, G.; Crespo, A.; Valiente, J. M. In *Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recognition*, Proceedings of International Conference on Neural Networks (ICNN'97), 12-12 June 1997; 1997; pp 1341-1346 vol.2.

38. Schmidt, C.; Rey, S.; Skupin, A., Effects of Irregular Topology in Spherical Self-Organizing Maps. *International Regional Science Review* **2011,** *34*, 215-229.

39. Keenan, M. R.; Kotula, P. G., Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surface and Interface Analysis* **2004,** *36* (3), 203-212.

40. Keenan, M. R.; Kotula, P. G., Optimal scaling of TOF-SIMS spectrum-images prior to multivariate statistical analysis. *Appl. Surf. Sci.* **2004,** *231-232*, 240-244.

41. Hook, A. L.; Scurr, D. J., ToF-SIMS analysis of a polymer microarray composed of poly(meth)acrylates with C6 derivative pendant groups. *Surface and interface analysis : SIA* **2016,** *48* (4), 226-236.

42. Tyler, B. J.; Peterson, R. E., Dead-time correction for time-of-flight secondary-ion mass spectral images: a critical issue in multivariate image analysis. *Surf Interface Anal* **2013,** *45* (1), 475-478.

43. Da Silva, I. N.; Hernane Spatti, D.; Andrade Flauzino, R.; Liboni, L. H. B.; Dos Reis Alves, S. F., *Artificial Neural Networks: A Practical Course*. Cham: Springer International Publishing: Cham, 2017.

For Table of Contents Only