Journal of Materials Chemistry

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

PAPER

www.rsc.org/xxxxxx

Modelling human embryoid body cell adhesion to a combinatorial library of polymer surfaces

V. Chandana Epa,[a] Jing Yang,[e] Ying Mei,[b,c,d] Andrew L. Hook, Robert Langer,[b,c,d] [e] Daniel G. Anderson [b,c,d], Martyn Davies,[e] Morgan R. Alexander,[e] David A. Winkler[f,g]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Designing materials to control biology is an intense focus of biomaterials and regenerative medicine research. Discovering and designing materials with appropriate biological compatibility or active control of cells and tissues is being increasingly undertaken using high throughput synthesis and assessment methods. We report a relatively simple but powerful machine-learning method of generating models that link microscopic or molecular properties of polymers or other materials to their biological effects. We illustrate the potential of these methods by developing the first robust, predictive, quantitative, and purely computational models of adhesion of human embryonic stem cell embryoid bodies (hEB) to the surfaces of 496-member polymers.

1. Introduction

Culture of multipotent cells such as haematopoietic stem cells (HSCs) and induced pluripotent stem cells is a major research focus in regenerative medicine. Present methods to culture them and expand their population rely upon animal-derived products now increasingly under scrutiny. Much research effort is focused on designing chemically defined, serum-free, feeder-free synthetic substrates and media to support robust self-renewal of pluripotent cells. Changes in cellular properties such as adhesion, morphology, motility, gene expression and differentiation are influenced by surface properties of the materials on which cells have been cultured. Important surface properties that have been identified include surface chemistry,[1] surface wettability,[2] topography,[3] and elastic modulus.[4] Additionally, it is clear that proteins adsorbed onto material surfaces strongly influence the biological responses to the surfaces.[5, 6] High throughput methods employing large polymer libraries and rapid screening methods can play an important role in discovery of materials for culture and expansion of stem cells.[7] High throughput surface characterisation has been developed that allows surface structure-property relationships to be investigated.[8-10] Working together, these techniques allow a much larger part of materials property space to be explored than has been possible in the past. However, as the dimensionality of materials property space is too large to be explored by even high throughput methods, computational modelling provides an effective means of leveraging the limited and expensive experimental data into a larger portion of materials property space.

Consequently, high throughput synthesis and characterization technologies are complementary to computational modelling tools that analyse large data sets and provide interpretation and prediction of new, improved materials. Robust machine learning methods can extract useful information on design and optimization of new materials from many types of existing data. They can identify which physical, process, and chemical properties of polymers and other materials will have the greatest influence on cell and tissue response. They can also reduce the dimensionality of complex synthesis and processing procedures by identifying the subset of these parameters that have little effect on biological outcomes and may be ignored.[11] Machine learning methods are simple to apply, broad in application, and particularly well suited to data from high throughput experiments.[12]

Recently Yang et al.[13] reported the first relationship between surface chemistry and structure of a polymer microarray and the adhesion of partially differentiated stem cells: human embryonic stem cell embryoid bodies (hEB). The large library of materials in the microarray was characterized experimentally by wettability, surface topography, surface chemistry, and indentation elastic modulus properties. These studies employed high-throughput synthesis and characterization methods to explore the polymer property space supporting stem cell growth. They identified materials that, with a fibronectin pre-treatment, could support hEB adhesion. The adhesion of human stem cells is critical for cellular activities such as proliferation and differentiation. Multivariate analysis of time of flight secondary ion mass spectrometry (ToF-SIMS) data was used to identify relationships between surface chemistry and cell attachment.[14] Yang et al.[13] used these TOF-SIMS data and other experimentally-derived polymer properties to generate a model of hEB adhesion. This approach has since been applied to other cell characteristics such as pluripotency.[15] Their methodology provided a general paradigm for the combinatorial development of synthetic substrates for stem cell culture that has recently been extended to developing materials with reduced bacterial pathogen attachment.[16]

We investigated whether advanced machine-learning methods

coupled with efficient mathematical descriptions of molecular properties could model and predict hEB adhesion to this large library of polymers. Our aim was to determine how well we could predict experimental hEB adhesion of the polymer library using computational descriptors alone, not using any experimental data such as contact angle, TOF-SIMS spectra, or mechanical properties. Purely computational methods of modelling high throughput materials data will clearly accelerate new materials discovery by reducing the need for additional experimental measurements to characterize the microscopic, bulk, or surface chemistry properties of large materials libraries.

2. Experimental

We employed partially differentiated hEB cells rather than undifferentiated human embryonic stem cells (hES) cells because fully dissociated hES cells tend to undergo cell death during plating. hEB cells are substantially more robust, while maintaining high differentiation potential. The hEBs were cultured for 8 days, as described in Yang et al.[13] hEBs were subsequently trypsinized and cultured on fibronectin (Fn) pre-conditioned polymer arrays for 16 hrs to test their initial adhesion. Polymer arrays were washed with PBS, fixed with Accustain (Sigma) solution for 30 min, permeabilized with 1% Triton X-100 in PBS for 10 min, and then stained with Cyto 24 (Invitrogen) for 1 h. The arrays were gently washed with PBS and deionised water to remove buffer salts and air dried before imaging by laser-scanning cytometry, and cell number quantification.

The polymer library was synthesized and characterized as previously described by Yang et al.[13] It consisted of 496 polymers synthesized by mixing 22 monomers at various ratios, for which hEB adhesion on the surface had been measured. Surface contact angle, elastic modulus of the polymers in the library, and the surface roughness were measured, and surface chemistry parameters were characterized using ToF-SIMS. These experimental measurements had been used by Yang at al. to model the growth of hEB on the library polymers. However, we generated models of this biological property that employed only molecular descriptors that could be calculated from the monomer structures (no experimental measurements required).

For computational modelling we partitioned the data set into a training and test set. The training set was used to generate the models and contained 80% of the data (397 polymers). The remaining 20% of the data (99 polymers) constituted an independent test set used to estimate how well the models could predict data not used to generate the model. The splitting of training and test sets was achieved by using k-means cluster analysis. We generated 68 molecular descriptors (mathematical objects that capture the molecular properties of polymers) using Dragon v. 5.5[17] and Adriana v. 2.2[18] software. Descriptors were chosen to be chemically interpretable and a large number of more complex potential descriptors were not used. The QSPR models were generated using multiple linear regression with sparsity imposed by an expectation maximization algorithm.[19] Nonlinear models used three layer neural networks with the same number of input nodes as descriptors used, a variable but small number of hidden layer nodes, and a single output node corresponding to the property (e.g. hEB adhesion) being modelled (Figure 1).

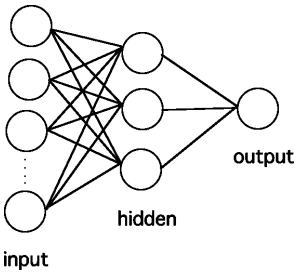


Figure 1. Structure of the neural networks. The input nodes receive the molecular descriptors, the hidden layer (2-3 nodes) does the computation, and the output node generates the predicted response variable (hEB adhesion or roughness).

The logarithm of the properties being modelled was used, as is usual practice in these types of machine learning models. The complexity of the neural network models was controlled using Bayesian regularization that employed either a Gaussian prior (BRANNGP)[20] or a sparsity-inducing Laplacian prior (BRANNLP)[21]. The maximum of the Bayesian evidence for the model was used to stop training of the neural network. Both neural network methods effectively prune the number of weights in the network to a number that is substantially smaller than the number of weights in a fully connected network. This reduced number of weights is called the number of effective weights, and is one of the reasons why Bayesian regularized neural networks are relatively immune to overfitting. The BRANNLP neural network also prunes less relevant descriptors from the model, depending on the sparsity setting chosen. Details of the three modelling algorithms have been published previously.[19-21] No outliers were removed from the models.

3. Results and discussion

3.1 Stem cell embryoid body adhesion models

We modelled the adhesion of hEBs to the entire 496-member polymer library in several ways. We used linear modelling methods with increasing levels of sparsity to model the EB adhesion in order to identify the molecular features most relevant to the biological activity of the polymers. Optimally sparse models have the greatest ability to predict the properties of new polymers. We also used nonlinear modelling methods to generate models for EB adhesion to determine whether interactions between the relevant molecular features, or nonlinear relationships between these features and the adhesion were important. We generated models of EB adhesion that employed only calculated molecular descriptors for the polymer components. The quality of prediction of the EB adhesion generated by both linear and nonlinear models was relatively high.

The linear hEB adhesion model (MLR) predicted the training set with an r^2 value of 0.68 (i.e. the model accounted for 68% of the variance in the data), and a standard error of estimation (SEE) of 0.163 logEB (predicted hEB binding within a factor of ±1.5). This model successfully predicted the hEB adhesion on polymers in the test set with an r^2 value of 0.66, and a standard error of prediction (SEP) of 0.145 logEB. The similarity between the training and test set results suggests the model is robust and not overfitted. These results were similar to those for a partial least

squares (PLS) model of hEB adhesion that used experimental ToF-SIMS peaks as descriptors reported by Yang et al.[13] They reported a training set r^2 value of 0.74 and test set r^2 of 0.62 for their model (training and test partitioning were different to our study). No standard errors were reported.

The two nonlinear Bayesian neural network models were substantially better than the linear model at predicting training and test sets. The quality of both neural network models was similar to each other. The Bayesian neural network using a Gaussian prior (BRANNGP) with two nodes in hidden layer predicted the hEB adhesion of the training set polymers with an r^2 value of 0.81 (i.e. the model explained 81% of the variation in the data), and an SEE=0.108 logEB (the model could predict the EB binding to within a factor of ±1.3). The model predicted the hEB adhesion for test set polymers with an r^2 value of 0.80, and an SEP of 0.107 logEB (predicted EB binding within a factor of ±1.3). This model had 28 effective weights in the neural network, considerably fewer than the number of polymers in the training set and similar to the number of monomers from which the library was generated. The Bayesian neural network with sparse Laplacian prior (BRANNLP) also employed two nodes in the hidden layer. It predicted hEB adhesion for training set polymers with an r^2 value of 0.80, and an SEE=0.113 logEB (predicted EB binding within a factor of ±1.3). This model predicted hEB adhesion of test set polymers with very similar fidelity to the BRANNGP model with an r^2 of 0.82, and an SEP of 0.101 logEB (predicted EB binding within a factor of ±1.3) (Figure 2). This model used twenty-three molecular descriptors. The BRANNLP method automatically prunes out the least relevant molecular descriptors and network weights. The majority of molecular descriptors were pruned from the model. The twenty-three most relevant descriptors used in the model are summarized in Table 1, together with a description of the type of information these descriptors encode.

Table 1. Description of parameters used in the hEB adhesion model

Parameter	Description
HAcc_N	Number of H-bond acceptors on nitrogen
XlogP	Log octanol/water partition coefficient
Dipole	Molecular dipole moment
LogS	Log aqueous solubility
NRotBond	Number of rotatable bonds
NViolationsRo5	Number of Lipinski's rule of 5 violations
NStereo	Number of tetrahedral stereo centres
Complexity	Molecular complexity parameter
RComplexity	Ring complexity
Rgyr	Radius of gyration
Aspheric	Molecular asphericity
nCs	Number of secondary C(sp3)
nCrs	Number of ring secondary C(sp3)
nCar	Number of aromatic C(sp2)
nR=Cp	Number of terminal primary C(sp2)
nR=Cs	Number of aliphatic secondary C(sp2)
nRCOOR	Number of esters (aliphatic)
C-004	Number of atom-centred fragments CR_4
C-006	Number of atom-centred fragments CH_2RX
C-015	Number of atom-centred fragments $=CH_2$
C-026	Number of atom-centred fragments R--CX--R
H-047	Number of H attached to C1(sp3)/CO(sp2)
O-059	Number of aliphatic ether atom-centred fragments

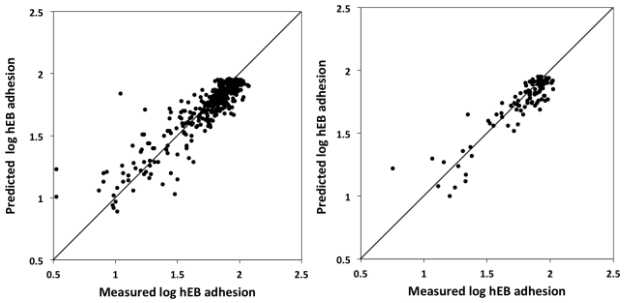


Figure 2. Predictions of the log hEB adhesion on the polymers for the training (left) and test (right) sets for the nonlinear Bayesian (BRANNLP neural net model.

The two neural network models had substantially higher predictive power than the PLS models using experimentally determined parameters reported by Yang et al. This suggests that there is some nonlinearity in the relationships between polymer structure and hEB adhesion, or that some of the descriptors used interact with each other in the models. The similarity between the training and test set statistics also strongly suggests that all models are quite robust with no overtraining or overfitting occurring. Earlier PLS models of hEB adhesion reported by Yang et al.[13] indicated that hEB adhesion correlated with ions identified in the ToF-SIMS experiments corresponding to the following polymer environments: hydrocarbons, esters, cyclic structures, tertiary amines, propylene glycol, tertiary butyl)[22]. The most relevant descriptors used in our models are in very good agreement with these conclusions. The logP octanol/water and water solubility, (XlogP, logS) and hydrocarbon indicator variable (nCs, nCrs, nCar , nR=Cp, nR=Cs) descriptors are describing molecular surface chemistry properties similar to those of the hydrocarbon ToF-SIMS peaks. The descriptor for the number of esters (nRCOOR) contains information similar to that of ions assigned in the ToF-SIMS to esters from the monomer structures that correlated with hEB adhesion. The cyclic structures ToF-SIMS peak is mimicked to some extent by the molecular complexity (Complexity, RComplexity), radius of gyration (Rgyr), and molecular sphericity (Aspheric molecular) descriptors. Finally the tertiary amine and propylene glycol ToF-SIMS peaks contain similar information on hydrogen bonding interactions to that of the number of hydrogen bond acceptors on nitrogen (HAcc_N) and dipole moment (Dipole).

As the polymers were pretreated with Fn, it was possible that it is the presence of this protein that modulates the hEB adhesion, rather than the polymers directly. Therefore, we calculated the correlation between Fn adhesion to the polymer library and that of hEB. Surprisingly, the correlation was only 0.05, with the correlation between the log transformed values modelled below being slightly higher at 0.18. This poor correlation between Fn binding and hEB adhesion suggests that the relationship between surface chemistry and properties and hEB adhesion is quite complex. Recent work by Szott and Horbett indicates that it is protein conformation, not the amount that modulates cell adhesion.[23] Polymers in the library are therefore influencing hEB adhesion indirectly via their effect on Fn conformation. The modelling of Fn adhesion to this polymer library will be reported elsewhere.

To understand how the calculated descriptors could substitute

for experimentally measured properties in modelling hEB adhesion on polymer surfaces, we additionally generated machine-learning models of surface roughness that also employed calculated molecular descriptors solely.

3.2 Surface Roughness models

Although the modelling and prediction of the adhesion of hEBs on polymers was the primary focus of our work, we also constructed models of the experimentally measured surface roughness because this appeared to impact on the adhesion of hEBs. It was not intuitively obvious that surface roughness could be modelled computationally, as this material property may have more to do with sample preparation than the chemical structure of monomers and polymers. However, it is likely that materials properties will have *some* influence on polymer surface roughness. We have previously observed that certain combinations of monomer chemistries (e.g. mixed hydrophobic and hydrophilic) produce specific nanotopographies with associated changes in roughness[24]. In some cases this results from phase separation prior to polymerization.

The statistics of the prediction of the training and test sets were similar to each other (Table 2), but compared to the hEB model, the statistical quality of the surface roughness models was lower. The moderate values for the r^2 value of the non-linear models for the test sets in particular suggests that the models have some degree of useful predictive power. Clearly other factors such as how the samples are prepared may indeed have a substantial impact on the surface roughness, as might be intuitively expected. The best nonlinear models account for 60% of the variance in the data, the remainder we suggest is largely due to experimental factors. There were twenty-two indices with nonzero weights in the most parsimonious BRANNLP model. These corresponded to descriptors for hydrophilic properties (number of H-bond acceptors on nitrogen, dipole moment, number of primary alcohols) and hydrophobic properties (number of tetrahedral stereo centres; ring complexity; first principal moment of inertia; molecular asphericity; number of secondary sp3 carbon atoms; number of total quaternary carbon atoms; number of secondary sp3 carbon atoms in a ring; number of substituted benzene carbon atoms; number of terminal primary sp2 carbon atoms; number of aliphatic secondary sp2 carbon atoms; number of aliphatic ethers; number of aromatic ethers; number of atom centred fragments CR4, CH_2RX, CHR_2X, =CHR, R—CX—R, aliphatic-O-aliphatic, and aliphatic-O-aliphatic/aromatic-O-aromatic/R-O-R/R-O-C=X). These descriptors were consistent with phase separation playing a role in surface topography. As surface roughness had previously been identified as an important factor in hEB adhesion, the fact that it can be modelled numerically reasonably well provides an explanation as to why we can model hEB adhesion without requiring this measured polymer surface property.

The relative performance of the three methods in modelling roughness is summarized in Table 2. The MLR model performs poorly compared to the neural network models.

Table 2. Summary of surface roughness model statistics

Model	r^2_{train}	SEE	r^2_{test}	SEP	$N_{effective}$
MLR	0.44	0.199	0.51	0.259	69
BRANNGP (3 nodes in hidden layer)	0.66	0.134	0.63	0.212	47
BRANNLP (2 nodes in hidden layer)	0.61	0.143	0.64	0.209	22

The quality of the prediction of the BRANNGP models for training and test set is illustrated in Figure 3. The models have modest although statistically significant predictivity in contrast to the lack of correlation of experimental ToF-SIMS data with the polymer roughness reported by Hook et al.[22] This lack of correlation may be due to artefacts in the estimation of the surface roughness reported that were subsequently removed in the data modelled here.

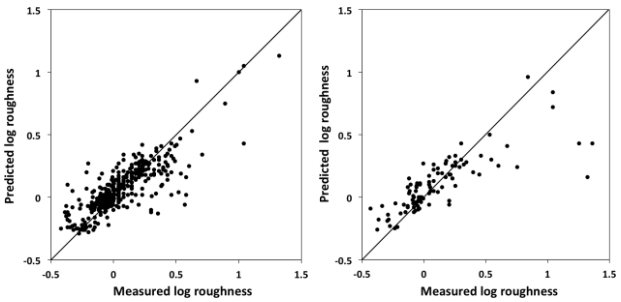


Figure 3. Predictions of surface roughness for the training (left) and test (right) sets for the nonlinear Bayesian (BRANNGP) neural net models.

4. Conclusions

We found that the stem cell hEB adhesion on polymeric surfaces could be modelled well by our approach using only calculated molecular descriptors. These models provide a compact summary of a large amount of numerical data, and some interpretation of the role of surface chemistry in hEB adhesion. These models allow experimental data to be leveraged into a larger portion of materials property space by predicting polymers with improved properties. In addition, surface roughness can also be modelled moderately well using molecular descriptors. This suggests that surface roughness, important for hEB adhesion, may have at least a partial molecular basis, most likely phase separation. Our analysis and the descriptors that we use are amenable to systematic 'reverse engineering' by predicting the properties of larger virtual libraries of plausible polymer candidates and by allowing chemical interpretation of the relevant polymer molecular descriptors. These robust modelling methods that require only computed materials descriptors are a valuable complement to high throughput synthesis and characterization methods. They will allow more of materials property space to be accessed than by experimental methods alone.

Notes and references

[a]*CSIRO Materials Science & Engineering, Parkville, Australia*
[b]*Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA*
[c] *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA*
[d] *Division of Health Science Technology, Massachusetts Institute of Technology, Cambridge, MA*
[e]*University of Nottingham, Nottingham, UK.*
[f]*CSIRO Materials Science & Engineering, Clayton, Australia*
[g]*Monash Institute of Pharmaceutical Sciences*

1. A. S. Curtis, J. V. Forrester, C. McInnes and F. Lawrie, *J. Cell Biol.*, 1983, **97**, 1500-1506.
2. K. L. Menzies and L. Jones, *Optom. Vis. Sci.*, 2010, **87**, 387-399.
3. H. V. Unadkat, M. Hulsman, K. Cornelissen, B. J. Papenburg, R. K. Truckenmuller, G. F. Post, M. Uetz, M. J. Reinders, D. Stamatialis, C. A. van Blitterswijk and J. de Boer, *Proc. Natl. Acad. Sci. U S A*, 2011, **108**, 16565-16570.
4. K. G. Robinson, T. Nie, A. D. Baldwin, E. C. Yang, K. L. Kiick and R. E. Akins, Jr., *J. Biomed. Mater. Res. A*, 2012, **100**, 1356-1367.
5. Y. Mei, S. Gerecht, M. Taylor, A. J. Urquhart, S. R. Bogatyrev, S.-W. Cho, M. C. Davies, M. R. Alexander, R. S. Langer and D. G. Anderson, *Adv. Mater.*, 2009, **21**, 1-6.
6. C. J. Wilson, R. E. Clegg, D. I. Leavesley and M. J. Pearcy, *Tissue Eng.*, 2005, **11**, 1-18.
7. D. Anderson, S. Levenberg and R. Langer, *Nature Biotech. Lett.*, 2004, **22**, 863-866.
8. M. Taylor, A. J. Urquhart, D. G. Anderson, R. Langer, M. C. Davies and M. R. Alexander, *Surf. Interf. Anal.*, 2009, **41**, 127-135.
9. A. Urquhart, M. Taylor, D. Anderson, R. Langer, M. Alexander and M. C. Davies, *Adv. Mater.*, 2007, **19**, 2486–2491.
10. A. J. Urquhart, M. Taylor, D. G. Anderson, R. Langer, M. C. Davies and M. R. Alexander, *Anal. Chem.*, 2008, **80**, 135-142.
11. D. A. Winkler and F. R. Burden, *Mol. Biosys.*, 2012, **8**, 913-920.
12. T. C. Le, V. C. Epa, F. R. Burden and D. A. Winkler, *Chem. Rev.* , 2012, **ASAP**.
13. J. Yang, Y. Mei, A. L. Hook, M. Taylor, A. J. Urquhart, S. R. Bogatyrev, R. Langer, D. G. Anderson, M. C. Davies and M. R. Alexander, *Biomat.*, 2010, **31**, 8827-8838.
14. M. Taylor, A. J. Urquhart, D. G. Anderson, P. M. Williams, R. Langer, M. R. Alexander and M. C. Davies, *Macromol. Rapid Comm.*, 2008, **15**, 1298-1302.
15. Y. Mei, K. Saha, S. R. Bogatyrev, J. Yang, A. L. Hook, Z. I. Kalcioglu, S. W. Cho, M. Mitalipova, N. Pyzocha, F. Rojas, K. J. Van Vliet, M. C. Davies, M. R. Alexander, R. Langer, R. Jaenisch and D. G. Anderson, *Nature Mater.*, 2010, **9**, 768-778.
16. A. L. Hook, Y. Mei, J. Yang, S. Atkinson, C.-Y. Chang, D. Irvine, R. Bayston, R. Langer, D. Anderson, G., P. Williams, M. C. Davies and M. R. Alexander, in *Nature Biotech.*, 2012.
17. A. Mauri, V. Consonni, M. Pavan and R. Todeschini, *Match-Comm. Math Co.*, 2006, **56**, 237-248.
18. J. Gasteiger, *J. Med. Chem.*, 2006, **49**, 6429-6434.
19. F. R. Burden and D. A. Winkler, *QSAR Comb. Sci.*, 2009, **28**, 645-653.
20. F. R. Burden and D. A. Winkler, *J. Med. Chem.*, 1999, **42**, 3183-3187.
21. F. R. Burden and D. A. Winkler, *QSAR Comb. Sci.*, 2009, **28**, 1092-1097.
22. A. L. Hook, J. Yang, X. Chen, C. J. Roberts, Y. Mei, G. G. Anderson, R. Langer, M. R. Alexander and M. C. Davies, *Soft Matter*, 2011, **7**, 7194.
23. L. M. Szott and T. A. Horbett, *Curr. Opin. Chem. Biol.*, 2011, **15**, 677-682.
24. A. L. Hook, J. Yang, X. Chen, C. J. Roberts, Y. Mei, D. G. Anderson, R. Langer, M. R. Alexander and M. C. Davies, *Soft Matter*, 2011, **7**, 7194-7197.

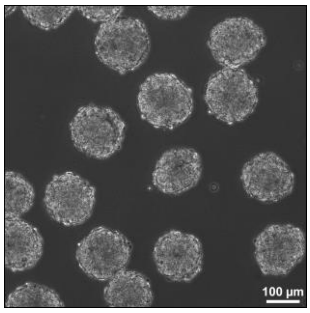
Journal of Materials Chemistry

Cite this: DOI: 10.1039/c0xx00000x

PAPER

www.rsc.org/xxxxxx

TOC entry



Modern, sparse machine learning methods allow accurate *in silico* prediction of stem cell embryoid body adhesion to large polymer libraries.