scientific reports

OPEN

On the correlation between material-induced cell shape and phenotypical response of human mesenchymal stem cells

Aliaksei S. Vasilevich[1], Steven Vermeulen[1,2], Marloes Kamphuis[2], Nadia Roumans[2], Said Eroumé[2], Dennie G. A. J. Hebels[2], Jeroen van de Peppel[3], Rika Reihs[2], Nick R. M. Beijer[2], Aurélie Carlier[2], Anne E. Carpenter[4], Shantanu Singh[4] & Jan de Boer[1]✉

Learning rules by which cell shape impacts cell function would enable control of cell physiology and fate in medical applications, particularly, on the interface of cells and material of the implants. We defined the phenotypic response of human bone marrow-derived mesenchymal stem cells (hMSCs) to 2176 randomly generated surface topographies by probing basic functions such as migration, proliferation, protein synthesis, apoptosis, and differentiation using quantitative image analysis. Clustering the surfaces into 28 archetypical cell shapes, we found a very strict correlation between cell shape and physiological response and selected seven cell shapes to describe the molecular mechanism leading to phenotypic diversity. Transcriptomics analysis revealed a tight link between cell shape, molecular signatures, and phenotype. For instance, proliferation is strongly reduced in cells with limited spreading, resulting in down-regulation of genes involved in the G2/M cycle and subsequent quiescence, whereas cells with large filopodia are related to activation of early response genes and inhibition of the osteogenic process. In this paper we were aiming to identify a universal set of genes that regulate the material-induced phenotypical response of human mesenchymal stem cells. This will allow designing implants that can actively regulate cellular, molecular signalling through cell shape. Here we are proposing an approach to tackle this question.

Cells adapt to new situations by perceiving information, and transforming it into changes in their gene and protein expression repertoire[1]. Typical examples are diffusible morphogens[2] that control stem cell differentiation during development and regeneration[3,4], UV exposure resulting in the synthesis of pigment skin melanocytes[5], lack of oxygen initiating the HIF1 mediated hypoxia response[6] and shear forces in blood vessels regulating endothelial cell physiology[7]. Adaptation to the environment can also materialize as change in cell shape[8]. Elongated multinuclear muscle cells are particularly well suited to exert stretching forces[9], whereas the globular shape of lipid cells is optimized for volume-efficient fat storage[10].

Cell shape thus follows cell function, but is cell shape also important to control cell function? Does cell function follow cell shape? Loss of physiological cell shape is sometimes accompanied by pathology[11]. For instance, in tendinopathies loss of extracellular matrix integrity in tendon and ligaments is preceded by a change from the typically elongated tenocyte to a more rounded shape[12]. Similarly, cancer cell malignancy is often preceded by a change in cell morphology[13], suggesting that shape changes may be the cause rather than the consequence of the functional change. Experimental evidence for "function follows shape" comes from in vitro experiments where cell shape can be controlled using micropatterning techniques. Basic cellular decisions such as differentiation, apoptosis or metabolic rate can be controlled by merely controlling cell shape[14], suggesting that cell shapes generate signals which are transduced into the nucleus and result in changes in gene and protein expression[15].

[1]BIS-Biointerface Science in Regenerative Medicine, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. [2]Department of Cell Biology-Inspired Tissue Engineering, MERLN Institute for Technology-Inspired Regenerative Medicine, Maastricht University, Maastricht, The Netherlands. [3]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. [4]Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ✉email: j.d.boer@tue.nl

An important source of cell shape signaling is the actin cytoskeleton which changes upon shape change, leading to changes in transcriptional activity of transcription factors such as MRTF, SRF, YAP and EGR[16–18]. However, nature has more mechanosensitive proteins in store. Stretch-activated ion channels open upon stretching of the plasma membrane resulting in the influx of ions which trigger several signalling cascades[19], for example Zhang et al. previously provided evidence that potassium ion channel can be a key player to control MSC proliferation[20]. Many proteins contain mechanosensitive protein domains, e.g. the focal adhesion protein talin[21], and some proteins contain a so-called BAR domain which is able to sense the curvature of a membrane and transduce that into a signal[22]. The generation of cell signalling events under the control of shape can be clustered under the term mechanotransduction[23], and an open question in this field is how many different mechanobiological signals can be picked up by how many molecular circuits. The analogy is that of the canonical signalling pathways, where a dozen well-described families of diffusible molecules (hormones, metabolites, etc.) bind to and activate their cognate families of receptors and initiate signalling cascades.

We aim to identify how many mechanobiological pathways exist and how they are activated by cell shape parameters. How specific are the mechanobiological pathways and can we correlate shape features to gene expression and changes in phenotype? To answer these questions, we[24] and others[25] have used surface topography to control cell shape and direct cell function, because shape mediated control can be applied to functionalize the surface of medical devices. In this study, we have generated a library of topography-induced cell shapes and were able to correlate shape induced cell signalling to several different phenotypes, including molecular and morphological. We used bone marrow-derived mesenchymal stem cells as a conventional cell model system. These cells are multipotent, their progeny has very distinct differences in cell shape and there is ample evidence for shape-directed differentiation.

Results

Cell shape diversity can be captured by feature-based clustering of a library of cell shapes.
To systematically explore the relation between cell shape and phenotypic response, we exposed human bone marrow-derived mesenchymal stem cells (MSC) to a library of 2176 unique topographies (TopoChip) and observed a plethora of different cell shapes[26] which were distinct from the MSCs cultured on the flat substrate (Fig. 1a,b). For example, we observed cells with an elongated narrow cell body without any protrusions which resemble the shape of tenocytes (Fig. 1b, bottom-right cell) or had the cuboidal shape typical for bone lining cells; still, others appeared neuronal. Thus, topographies can induce cell shapes, some of which reminisce different mesenchymal cells in vivo.

To capture shape diversity, we generated a morphological fingerprint for each cell on each topography using 11 shape descriptors including Area, Perimeter, Eccentricity and Form Factor[27] (Supplementary Tables 1, 2). We pruned the cell database to remove outliers and imaging artefacts and focused on topographies that consistently induce the same cell shape (Supplementary Materials and Methods). The resulting selection of 851 cell shapes was plotted by Principal Components Analysis (PCA, Fig. 1c), which reduces data dimensionality and allows identifying cells with similar shapes. We observed a continuum of cell shapes with some densely and sparsely occupied areas, demonstrating that some shape features are more abundant in our shape collection than others. This was confirmed by hierarchical clustering, where the tree structure displays four distinct branches, which are further divided until it reaches single cell shapes (Fig. 1d). For further phenotypic experiments, we sampled the medoid surfaces from 28 clusters, which together span the whole range of shape feature range, and reproduced them on 12 mm polystyrene disks. Visual inspection of the MSCs again shows a large diversity of cell shape, which is further highlighted by their cell shape fingerprints (Fig. 1e). The 28 surfaces are thus representing cell shape diversity induced by the TopoChip library (Fig. 1f) and further permits well plate compatible biological assays.

Shape features of mesenchymal stem cells correlate to their phenotypic response.
Cell shape is related to cell function[28–30], so we probed the relation between cell shape features and a series of basic cell phenotypes, i.e. proliferation, apoptosis, protein biosynthesis, migration and differentiation, in MSCs exposed to the 28 groups of topographically-patterned surfaces and to flat control surfaces. We found that topographies induced profound differences in all phenotypic assays, demonstrating that basic cell biological processes are influenced by surface topographical cues. For instance, we observed a big difference in the rate of proliferation, ranging from 12 to 59% of EdU-positive cells, between the low and high scoring surfaces (Fig. 2a). Differentiation was also affected, with, e.g. a six-fold difference in lipid production under adipogenic conditions between the lowest and the highest performing surface (Fig. 2b). None of the surfaces exclusively affected one biological property, yet a comparison of the magnitude of the phenotypic response showed that each surface induces a distinct phenotypic fingerprint (Fig. 2c, Supplementary Fig. 1). This is further illustrated in Fig. 2d, in which we clustered the surfaces based on phenotypic response. Flat surfaces formed a distinct cluster, demonstrating that cells on topographies share part of their phenotypic response which is different from the behaviour of cells on a flat surface (Fig. 2d). A cluster was formed by surfaces 2063 and 1130, located between the flat surface and the other topographies. Interestingly, the topographies on these surfaces are composed of sparsely located pillars, which resemble the flat surface, inducing relatively similar cell shape and similar phenotypic scores.

We assume that the phenotypic fingerprint is the result of the specific combination of multiple cellular signals induced by the topographies, some of which are directly related to cell shape, but others may not. For instance, it is conceivable that the direct effect of topography on nuclear shape influences gene expression and thus phenotype[31], whereas we did not include nuclear shape features in our analysis. To assess the hypothesis that phenotypic responses are directly related to cell shape, we constructed computational models using the Lasso algorithm in which cell shape features were correlated to each specific phenotype.
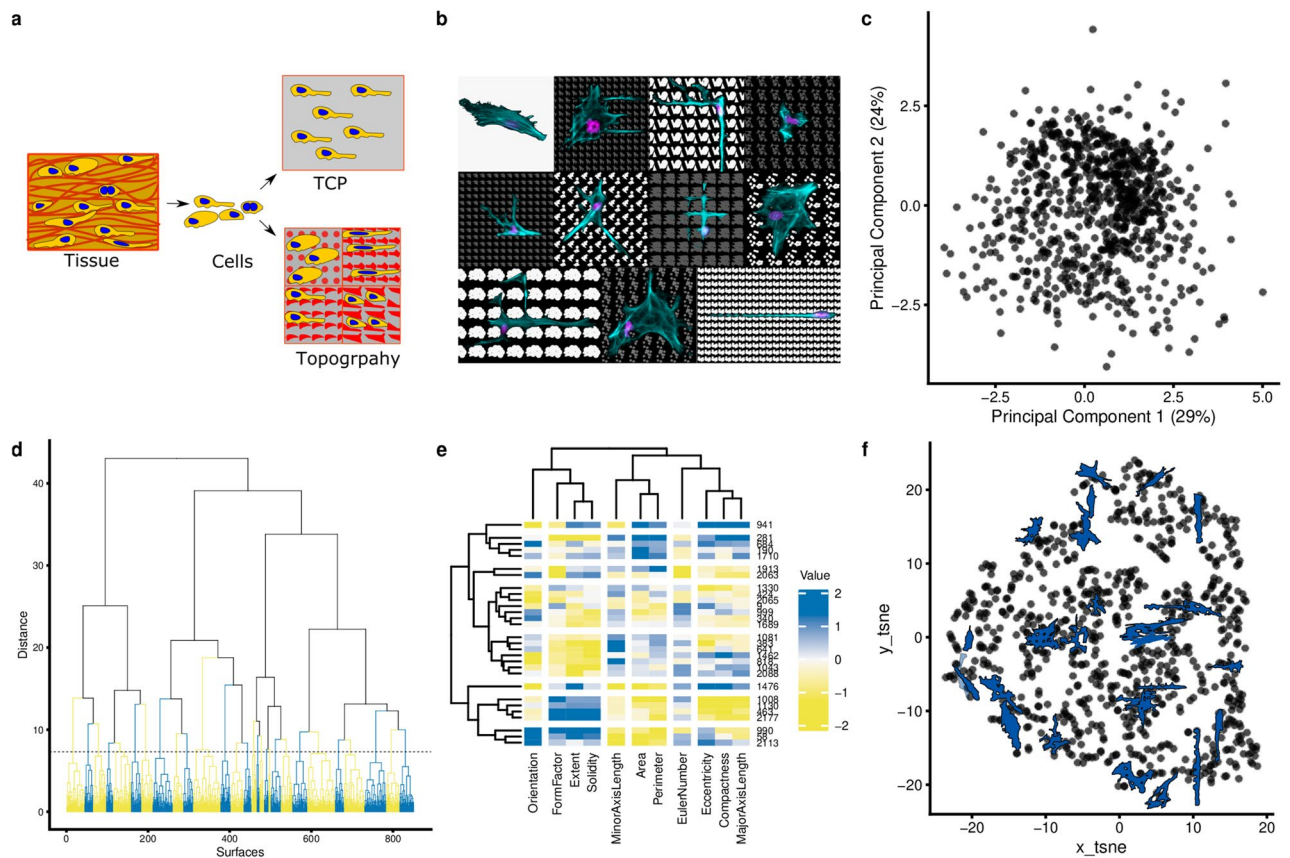
Figure 1. Clustering of cell shape diversity (**a**) Surface topography as a model system to investigate cell shape-induced cells response. TCP: tissue culture polystyrene. The schematic was prepared in the Inkscape 0.92.4. (**b**) The montage image that represents the diverse cell of hMSCs seeded on different topographies and flat surface (upper, left). Nuclei were stained with DAPI (magenta), actin was stained with phalloidin (cyan) The collage was prepared in the GIMP v 2.10.14. (**c**) Principal components analysis of all cell shapes on the TopoChip. The first two principal components are shown. Note that no noticeable clusters are observed. (**d**) Cluster gram representing a clustering of TopoChip cell shape data. Blue and yellow colours distinguish different clusters. (**e**) Heatmap representing the median values of cell shape features on the chosen medoid topographies (**f**) tSNE (t-distributed stochastic neighbor embedding) visualization of the obtained clusters. Cells shape silhouettes on 28 selected surfaces are represented.

The most accurate computational model was able to predict alkaline phosphatase (ALP) expression, a marker for osteogenic differentiation, with high accuracy (goodness of fit of 0.82, Fig. 2e). The shape parameters Form-Factor and Minor Axis length positively correlate with ALP expression, whereas Extent showed a negative correlation (Fig. 2f). Thus, a typical ALP positive cell has an elongated, irregular shape with many protrusions, which is consistent with previous work[26] (Fig. 3a).

The second most accurate computational model predicted protein biosynthesis with goodness of fit 0.68, in which Area shows a positive correlation, whereas Euler Number (the number of the "holes" in the object, which can appear when the cell is forming a protrusion around a pillar), negatively correlates to protein synthesis (Supplementary Fig. 2). We were also able to produce predictive models for cell proliferation (goodness of fit 0.63) and adipogenesis (0.37) but were unable to construct a model to predict apoptosis and migration (Fig. 2g). This suggests that parameters other than explored shape descriptors are responsible for differences in those readouts.

To summarise, surfaces selected based on the diverse cell shapes also induced diverse cell response in functional assays. Investigated cell shape descriptors could partially model the phenotypic variance. Different shape features account for different phenotypes, this suggests that shape affects the phenotype through various molecular mechanisms.

Topographies induce distinct but overlapping gene expression profiles in MSCs. To assess the landscape of genes under the control of surface topography, we exposed MSCs to seven topographies that span the phenotypic space and to flat control surfaces and assessed the transcriptome after 24 h. We observed between 46 to 258 differentially expressed genes with a fold-change > 1.5 as compared to flat (Fig. 3b). Each topography induced a unique fingerprint at the gene expression level (Fig. 3c), but with various degrees of overlap. Differential gene expression data was used to identify molecular pathways which were activated. Lists of Differentially Expressed Genes (DEGs) per topography were queried in the Connectivity Map (CMap)[32], a library of gene
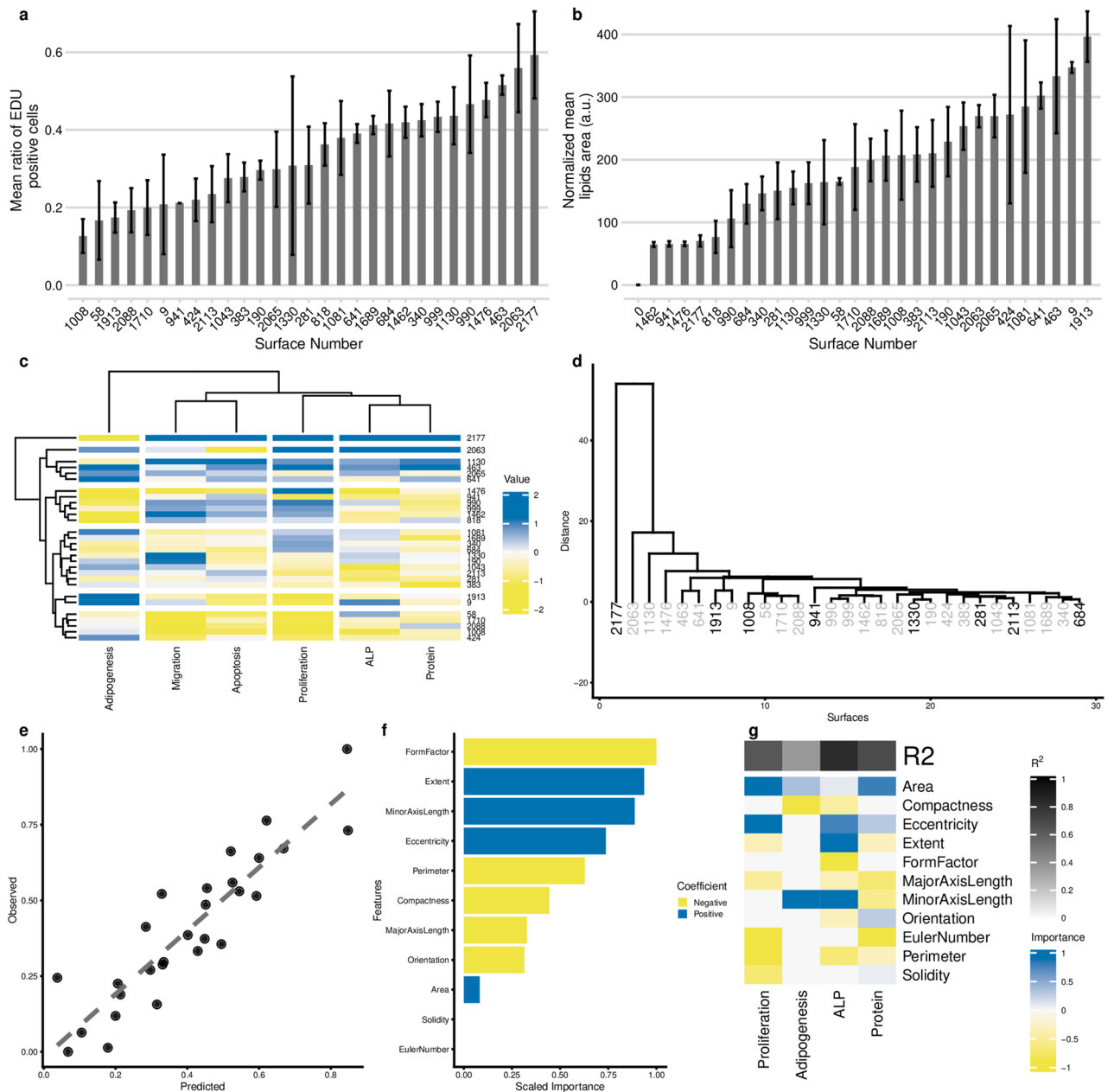
3

Figure 2. Profiling of phenotypic responses of hMSCs to topographies. (**a**) Ranking of surfaces based on the percentage of EdU staining. hMSCs were starved for 24 h, seeded on topographies overnight, and EdU was added 48 h prior to fixation. The bar represents mean of up to 3 replicas; error bar represents standard deviation. (**b**) Ranking of surfaces based on area covered by lipids. hMSCs were cultured on topographies in adipogenic medium for three weeks, and lipids were stained with Oil red O. Bar represents mean of up to 3 replicas, error bar represents standard deviation. Surface number 0 represents hMSCs on a flat surface in control media. (**c**) Heatmap represents the performance of the 28 selected topographies in 6 functional assays. (**d**) Clustering of the surfaces based on the readout in the phenotypic assays. Surfaces highlighted with black colour were selected for microarray analysis. On top, goodness of Fit R2 values reported. (**e**) Performance of the Lasso regression model on predicting ALP induction based on the 11 cell shape features. (**f**) Cell shape features that are important for predicting ALP induction. (**g**) Importance of the features for predicting the outcome of the phenotypic assays obtained with Lasso model.

expression signatures induced by chemical compounds or genetic interference (perturbants), and connectivity scores between the topography-induced DEGs and Connectivity Map DEGs were retrieved. We focused on perturbant classes (PCL), which are groups of signatures in which members belong to the same gene family or are targeted by the same compound (Fig. 4a). Some PCLs were shared by most topographies such as perturbants that induce *Cell cycle inhibition gain of function* or known inhibitors such as *JAK inhibitor* and *IKK inhibitors*, demonstrating that some signaling events are always induced when an MSC encounters a topography. In agreement
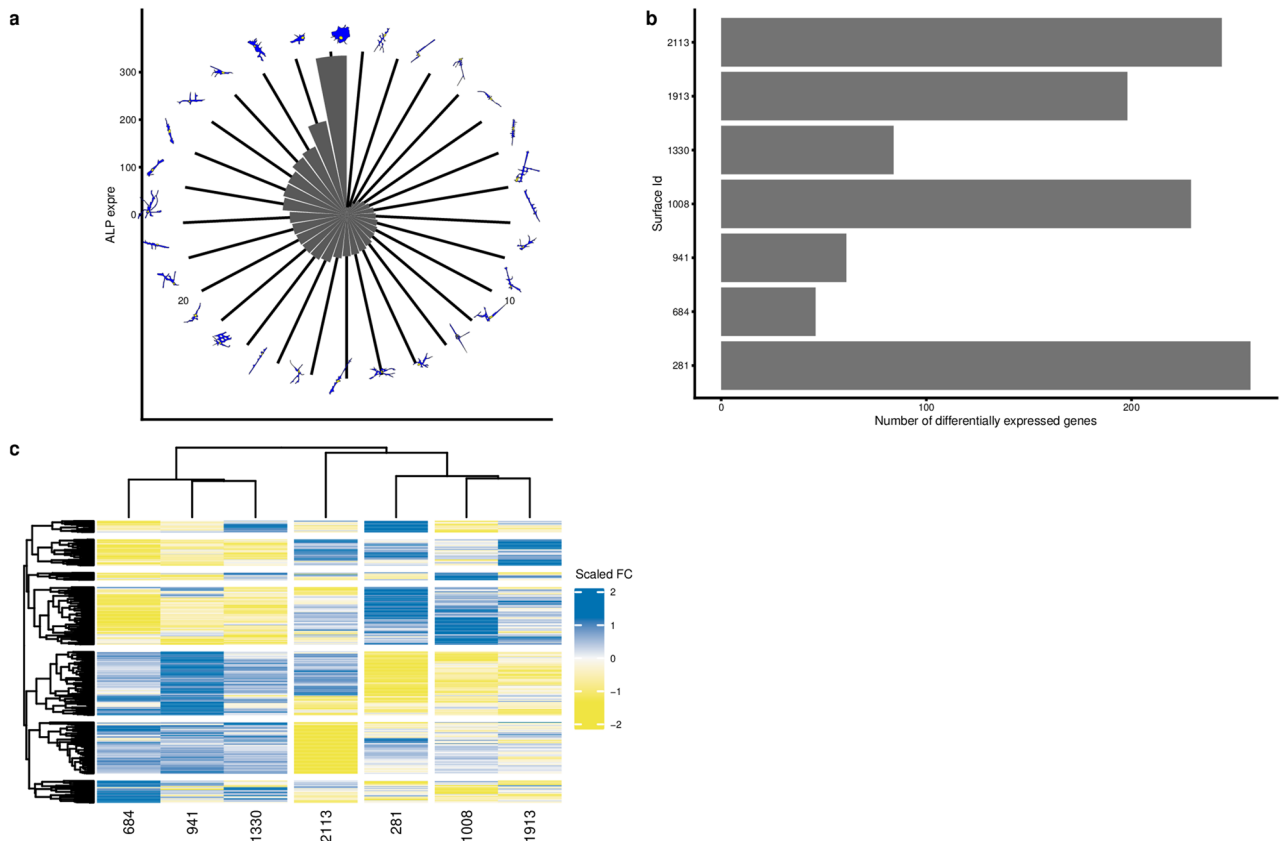
Figure 3. Gene expression on selected topographies. (**a**) Polar Bar plot that represents the level of ALP expression on different topographies, ordered from lowest to highest. On top, representative cell shapes silhouettes of the hMSCS cultured on the surface topographies are shown. (**b**) A number of differentially expressed genes per topography. For identification of differentially expressed genes (DEGs), we used data from a flat surface as a reference. (**c**) Heatmap represents a unique gene expression signature of cells on each topography. To express surface-specific DEGs Fold Change (FC) values were scaled by subtracting mean and dividing by sd of DEG FC for seven surfaces. (standard scaling approach).

with this, proliferation is impaired on all topographies relative to flat. Other PCLs were present in only one or a few topographical gene lists such as *LIM class homeoboxes GOF* or *glycogan synthase kinase inhibitor*, suggesting that some topographies induce distinctive signaling events.

To investigate how the DEGs collaborate to elicit a phenotypic response, we selected all DEGs absolute value with fold change > 1.9 from all seven topographies and created a gene network (Fig. 4b). Genes, represented by nodes, are connected by edges based on binary protein interactions described in public resources (such as KEGG, Reactome, and WikiPathways) using ConsensusPathDB software[33]. 51 out of 91 genes were placed in this topography-induced gene network, suggesting an extensive overlap in functionality. Based on Gene Ontology and manual literature research, we noticed three main clusters of cellular processes, i.e. related to actin, p53 and NF-kB signaling, respectively. Further literature research revealed that many DEGs in our network are specifically expressed during the transition from the G2 to M phase of the cell cycle (e.g. Aurora Kinase and CDC20) and are repressed in cells grown on topographies. We noticed a significant overlap in topographical DEGs and those associated with the p53/DREAM complex, which regulates quiescence (Supplementary Fig. 3). We previously reported on topography-induced quiescence as a typical topography-induced phenotype[34]. Whether actin, NF-kB and proliferation are in the same signalling network or represent separate signalling pathways induced by different cell shapes remains to be investigated.

Shape-dependent phenotypic responses show a strong association with specific genes. To determine the relation between gene expression, cell shape and phenotype, we calculated pairwise Spearman correlations between cell shape descriptors, gene expression, and phenotype score levels on all topographies. Many genes correlate to shape features such as Solidity and Perimeter (Fig. 5a), whereas far fewer genes correlate to Compactness or EulerNumber. We also noted that relatively few genes correlated with protein synthesis (Fig. 5b), ALP expression was positively correlated with many genes, and Adipogenesis negatively correlated with gene expression. We were particularly interested in situations where the Gene-Shape-Phenotype triad of correlations was robust and consistent (Fig. 6a). For example, Spearman correlation between VCAM-1 gene expression and hMSC proliferation was − 0.75 (Fig. 6b), and at the same time, Spearman correlation between cell Area and proliferation was − 0.71 (Fig. 6d), as the result the Spearman correlation between Cells Area and
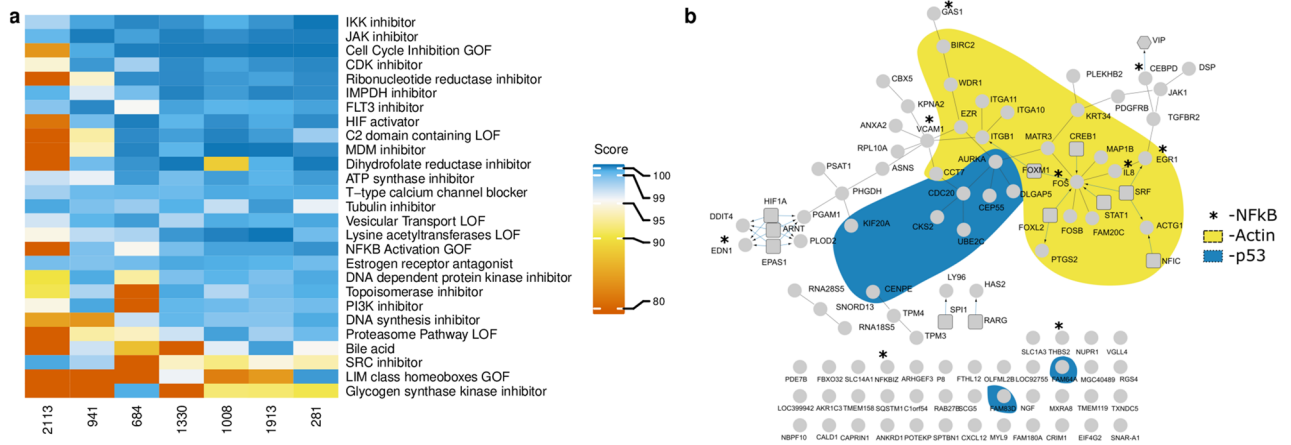
Figure 4. Network of genes and cell biological processes regulated on seven distinct topographical surfaces. (**a**) Gene expression in hMSCs on seven different topographies was compared to hMSCs on a flat surface and differentially expressed genes were compared in the Connectivity Map, a gene expression database with more than 1 million profiles. All processes that have absolute score value above 99 at least for one condition are depicted, with 0 low and 100 as high similarity. Biological processes have been ranked based on the number of topographies that can affect the process (specificity). (**b**) Gene expression in hMSCs on seven different topographies was compared to hMSCs on a flat surface, and 437 differentially expressed genes with an adjusted p-value of less than 0.05 were detected. The list was uploaded in ConsensusPathDB, which retrieves correlations between genes and proteins from a number of databases. Circular nodes in the graph represent the genes; rectangles are genes retrieved from a transcription factor database, thus transcriptional control. Edges represent reported relationships. The yellow and grey shading represent two clusters of genes with published involvement in actin related-processes and p53 respectively, based on literature study. Each circular node is a pie chart indicating in which topographies the gene is differentially expressed. . The image was prepared in the Inkscape 0.92.4, the network was drawn in the CytoScape v.3.5.1[66].
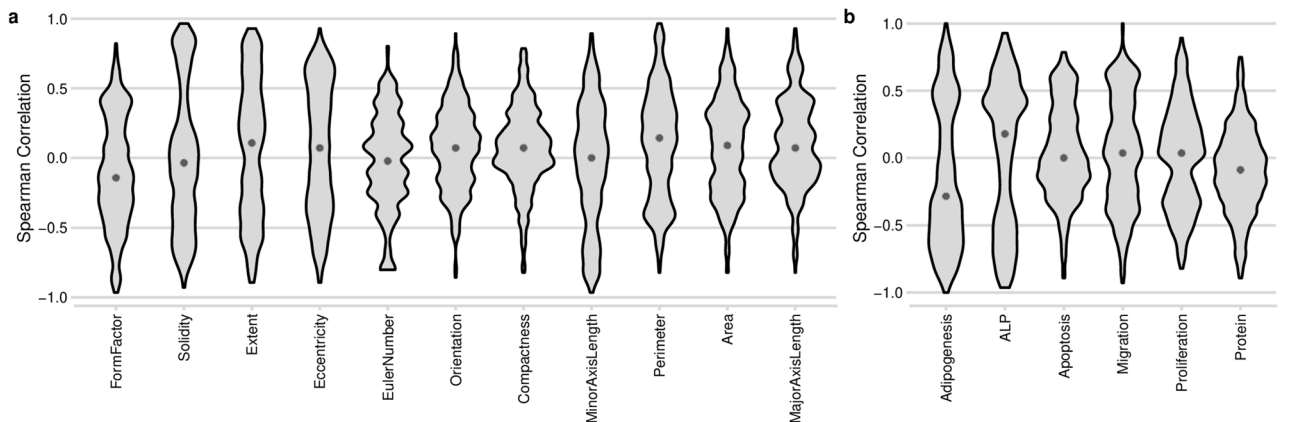

Figure 5. Correlation of gene expression to shape features or phenotype. Spearman calculated for all different samples a combined in 1 plot in which the distribution of outcomes can be observed. Violin plot representing the distribution of genes' Spearman correlation to either a shape feature (**a**) or a phenotype (**b**). Each gene was plotted against the respective feature and correlation presented as a Spearman score. Solidity has many high and low scoring correlation values, meaning that Solidity correlates strongly with gene expression. Most Spearman score for Euler number is close to zero, meaning that few genes' expression strongly correlate to this feature.

VCAM-1 was 0.71 (Fig. 6f). This implies that topographies that promote large cell areas are associated with a high level of proliferation and the expression of the VCAM-1 gene. Another clear example is the strong correlation between expression of the transforming growth factor receptor 2 (TGFBR2), ALP protein expression and Extent with Spearman correlation values, correspondingly, − 0.96, 0.67, − 0.75 (Fig. 6c,e,g). Interestingly, a strong association between ALP protein levels and TGFBR2 are reported in the literature[35]. Similar to what we found here, in a previous study we discovered a strong association between the amount of protrusions and ALP protein level[26].

Browsing the list of 59 unique genes with an absolute correlation value above 0.5 contains 124 combinations and contains many interesting correlations that connect cell shape with phenotype and potentially causal gene (available as Supplementary Information). To further corroborate this, we uploaded the Gene-Shape Correlation
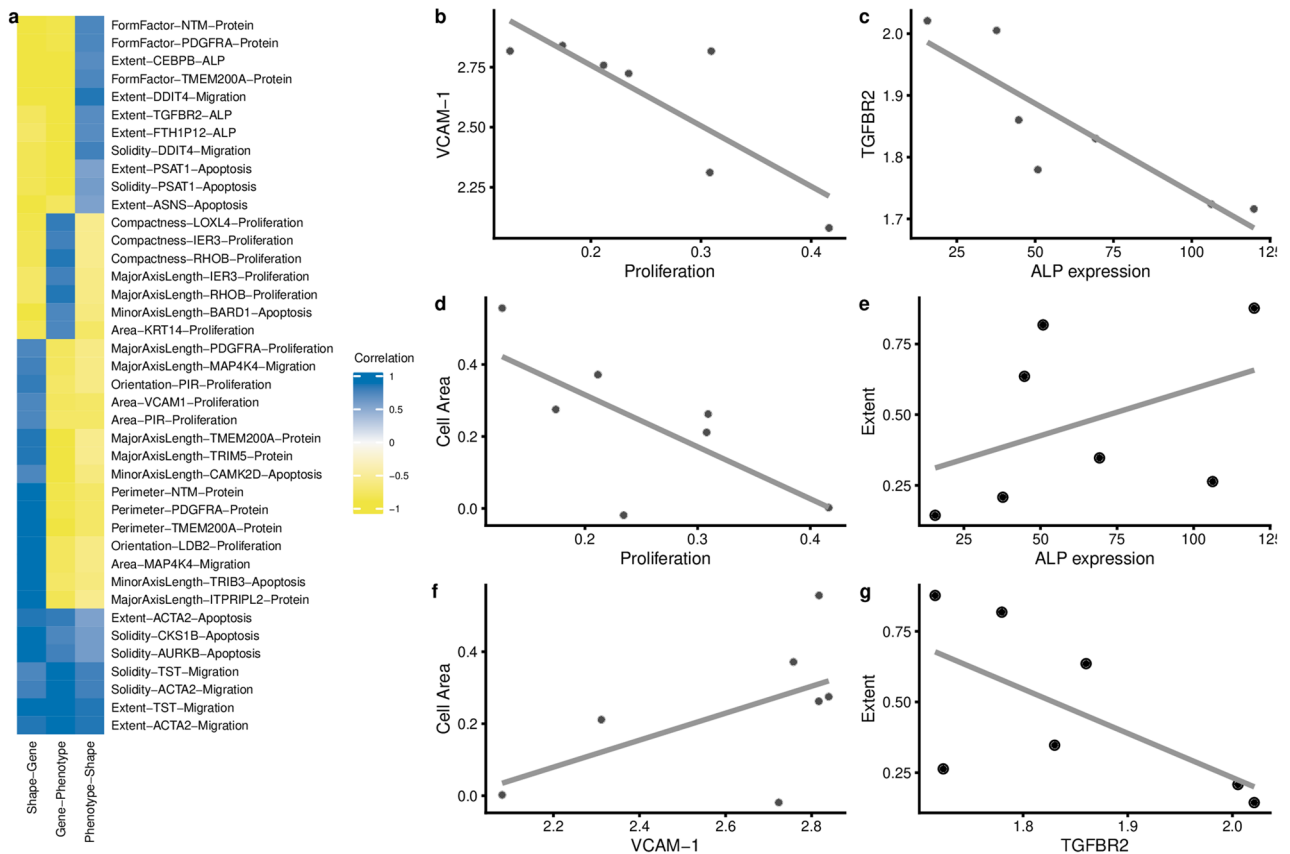
Figure 6. Correlation between shape feature-gene expression and outcome in the phenotypic assay. (**a**) Gene expression, cell shape and phenotypic data from hMSCs grown on the seven selected surfaces were compared by Spearman correlation between the 11 shape features, 437 genes and six phenotypes. Top three correlations per unique combination Shape parameter- phenotypic assay with the highest positive or negative values are shown. (**b**–**g**) Shows corresponding scatter plots with raw data.

list into Cmap and retrieved the PCLs (Fig. 7a,b). Surprisingly no PCL signatures above absolute value 0.99 were associated with Area. In contrast, Minor Axis length has the largest number of associated signatures. Moreover, we have observed that PCLs *HIF activator*, *CDK inhibitor* and *Cell Cycle Inhibition GOF* were the most common signatures with the highest score. Interestingly, *ATPase inhibitor* was exclusively linked to shape parameter Euler Number, which was the most important for the prediction of the protein biosynthesis. At the same time, Euler Number has a very strong signature of the PCL *Protein Synthesis Inhibitor*. This example demonstrates that we can independently connect gene signalling and outcome in the phenotypic assay via cell shape features.

Comparison of different databases reveals a list of universal shape related genes. To investigate the broader relevance of our list of cell shape-correlated genes, we compared its overlap with two other data sets. In one study, whole transcriptome gene expression and related cell shape changes were induced by chemical compounds[36]. In a second study, cell shapes were under the control of adhesive islands, and gene expression was assessed[37]. Overlap of all topographically-induced genes, 437 in total, and the two other gene sets yielded a list of only 12 genes (Fig. 8a) and all the genes showed a strong Pearson correlation cell shape features (Fig. 8b). Of the 12 genes found to be shape-predictable in all three studies, Minor Axis Length and Compactness correlated to eleven of them; Extent correlated to seven of them. As expected, Cell Orientation did not correlate to gene expression (Fig. 8c). The above results demonstrate that filtering genes based on correlation to cell shape descriptors is a powerful method to find associations between gene expression, cell shape, and phenotype and that genes on the list of 275 genes can be considered as candidate genes directly influenced by cell shape. Indeed, of the twelve genes, seven have previously been directly linked to changes in cell shape:

- BIRC5 (Yap transcriptional target)[38],
- EGR1[39],
- FOS[40],
- VGLL4 (YAP/TAZ inhibitor)[41],
- ALDOA[42],
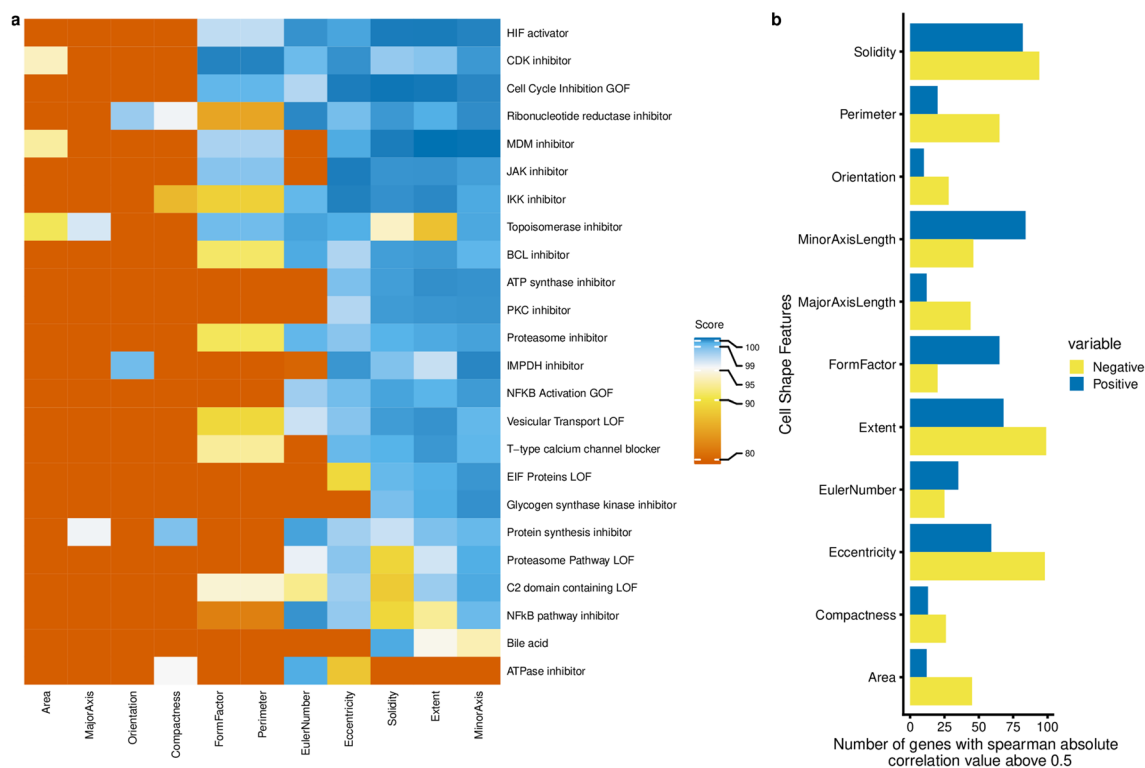- SQSTM1 (cytoskeleton remodeling via autophagy)[43].

Figure 7. Cell biological processes related to cell shape features. (**a**) Spearman correlation was calculated between gene expression and cell shape features. Genes with absolute Spearman correlation above 0.5 per condition (either per phenotype or per cell features) were used as input in the Connectivity Map, a gene expression database with more than 1 million profiles. All processes that have absolute score value above 99 at least for one condition are depicted, with 0 low and 100 as a high similarity. Biological processes have been ranked based on the number of conditions that can affect the process (specificity). (**b**) Number of genes that were used for the analysis per shape feature.

Three additional genes,

- CDK1[44],
- GADD45B[45] and
- CCNB2[46].

have been associated with proliferation. CDC 20 linked to both cell shape (Rho Signaling Protein)[47], and cells proliferation[48].

Discussion and conclusion

The molecular mechanisms connecting cell shape to basic cell functions and phenotype maintenance are important and yet remain largely unknown. Using high content imaging, transcriptomics and machine learning we were able to identify strong relationships between cell shape, molecular signalling and cellular phenotype.

In order to correlate the datasets (imaging, phenotypical assays and transcriptomics), all experiments were performed with cells from one donor and at one passage number because both parameters are known to affect the quantitative response of MSCs, as we have described in earlier work[49,50]. It is, therefore, likely that the ranking of the features described in this manuscript will be different when investigated in cells from different donors. Validation of specific parts of the data set is underway and was already achieved with the TGBFR2-EGR1 transcriptional response, which was observed in a different human MSC cell line and even in cells of different origin such as skin and muscle[51]. We also want to emphasize that differences in the transcriptomics response to external stimuli are obvious, as we have described earlier in a cohort of 62 donors, but we also observed a strikingly strong quantitative similarity between donors[52].

Some signal transduction pathways are particularly strongly correlated to topography-induced changes in cell shape. Changes in cell cycle-related signalling, and in the JAK, IKK and HIF pathways were observed in cells grown on most topographies. Other molecular signalling signatures were more specific, such as SRC, ATPase inhibition and Glycogen synthesis kinase which were affected in MSCs grown on specific topographies. This confirms earlier findings by Nassiri and McCall who identified signatures of DNA damage and proliferation signalling changes in cell shape[36], and work of Jael et al. who noticed changes in NF-kB[53].
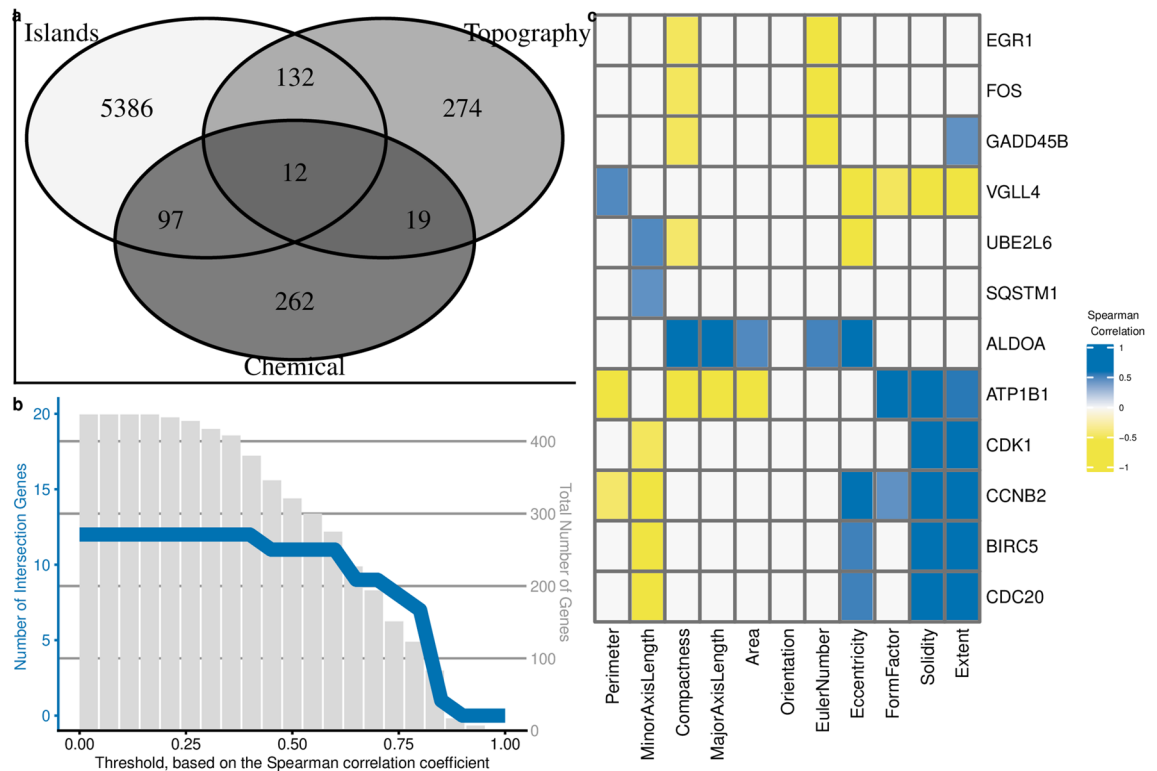
Figure 8. Genes related to shape are enriched in shape-based transcriptomics data sets. (**a**) Venn diagram representing the overlap between genes differentially expressed on different adhesive islands, genes related to chemically induced shape changes and the 437 shape-based genes differentially expressed on the seven topographies with a fold change above 1.5. (**b**) Filtering of the shape-specific genes based on the Spearman correlation score between the gene and at least on of the cell shape parameters. The red line and Y-axis at the left represents a number of selected shape-related genes with specified Spearman correlation threshold value (X-axis). Y-axis on the right represents the total number of genes that have Spearman correlation value above the specified threshold (X-axis). (**c**) Heatmap that represents the correlation value between shape specific genes and shape parameters. All Spearman correlations with an absolute value below 0.4 are depicted for clarity.

A classical mechano-transduction pathway which responds to changes in surface topography[54] but also material stiffness[55] and extracellular matrix remodeling links integrins to proteins in focal adhesions such as talin and vinculin, which subsequently affects Rho-ROCK signaling, actin remodeling and thus leading to activation of the YAP/TAZ transcription factor and downstream genes[56]. In a recently published manuscript[51], we show that surface topographies from our TopoChip library are able to stimulate TGFbeta signaling, leading to the transcriptional activation of the scleraxis gene. Under these circumstances, TGFbeta2 signaling can be mitigated by blocking Rho/ROCK and actin mediated signaling. On the other hand, small molecules that activate PKC, which is also associated with actin remodeling, can mimic the synergistic effect of topographies on TGFbeta2 induced signaling. In the current manuscript, we note that PCK is one of the small molecules whose gene expression profiles correlate to that of the cell shape features Extend, Solidity and Minor Axis. Additional evidence for the role of the integrin/Rho-ROCK-actin pathway is seen in Fig. 4b, where we see a distinct actin-associated network of genes with genes such as FOS, EGR1 and ACTG1, all of which are SRF target genes. SRF and MRTF-A are transcription factors whose activity can be directly regulated by actin. These observations suggest that at least part of the transcriptional response is related to signaling that relates to the integrin/Rho-ROCK/actin axis. Striking, we do not see a YAP signaling signature either in GO analysis or the CMap data.

Moreover morphological change of MSC was tightly associated with phenotypic functions, particularly in immunomodulation. Morphological changes of MSCs are related to cell skeletal re-arrangement, that is tightly involved in cytokine stimulus i.e. INF-r stimuli[57], TNF-a stimuli[58], Rap1/NFkb signaling[59], the formation of tunnelling nanotubes (TNT), and mitochondrial transfer[60].

Further support for the relevance of the genes and pathways discovered here in shape-related cell signalling comes from the overlap of 12 genes between our data set and those of two independent shape-related transcriptomics studies, and the known role of these genes. Thus, our systematic approach allowed us to confirm previous findings and to identify many novel unreported signatures. The data we have produced and publicly shared can be regarded as the first compendium of shape-gene expression and will likely lead to many new discoveries. The correlations can also be used to guide bio-instructive surfaces. For instance, more genes are affected by cell shape features such as Solidity and Extent then, for example, Compactness and some shape features are strongly

Functional experiment	Duration of experimental conditions
Adipogenesis	3 weeks
ALP induction	7 days
Cell proliferation	48 h
Apoptosis	66 h
Migration	66 h
Protein synthesis	90 min

Table 1. Duration of experimental conditions for functional experiments.

correlated with certain signalling pathways. Thus, cell shape features can be used to manipulate cells' molecular signatures with bio-active precision.

In this manuscript, we obtained proof of principle for the correlation between image-based cell features and cell signalling and 11 basic cell shape features. There is likely an opportunity for mechanistic information within cell organelle shape and associated correlations with phenotype. Although a thorough investigation of this was beyond the scope of this work, we did discover a strong negative correlation between nuclear area and the expression of histone H3, suggesting a link between nuclear shape and epigenetic regulation (data not shown). The approach can be taken much further and feed into the creation of large-scale computational models which describe biomaterial surface design, cell shape and cell phenotype. This approach is referred to as Quantitative Structure–Activity Relationships (QSAR) and is widely used to model the biological activity of pharmaceutical compounds[61]. It allows predicting the biological effect of untested compounds and thus narrowing down the search space.

To improve our shape-based models, we need more data which can be obtained by, e.g. upscaling the number of topographies from 28 in this manuscript to thousands, more extensive phenotypic characterisation of the cells for instance by using Cell Painting protocols (Bray et al. Nat Protocols) or other stains that display relevant biological processes and more extensive transcriptomics analysis on different cell types. As a start, data obtained in our project is freely available via our cBIT[62] repository and can already be used to design a biomaterial that inhibits cell proliferation, for example, to reduce growth of the fibrosis tissue on the interface of the tissue-implant. The gene signatures related to cell shape and their underlying topographies can be used to browse the Connectivity Map database for genetic signatures that resemble the effect of small molecule or RNAi-induced gene signatures[32]. This brings new opportunities to research the molecular pathways underlying shape-induced phenotypes. To accelerate the field of biomaterial engineering, this database should link information not only of topography design but also other material properties such as stiffness, ligand density, and any other relevant biomaterial properties. Based on systematic banking and mining of biomaterials, cell phenotype and transcriptomics data, we foresee a universal tool that will help understand and engineer the interface between cells and biomaterials. The work presented in this study outlines a method to search the large design space that links material properties to cell phenotype.

Materials and methods

Cell culture. In this study we used human mesenchymal stem cells (hMSCs) from donor (d016), undergoing a total hip replacement. hMSCs were isolated from bone marrow after obtaining written informed consent from the patient. Ethical approval for using the bone marrow sample was obtained from the ethical advisory board of the Medisch Spectrum Twente, Enschede. All methods were carried out in accordance with local and relevant guidelines and regulations[52]. Cells were cultured in basic medium (αMEM medium (Gibco, 22-571-038) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich batch number 013M3396) at 37 °C in a humid atmosphere with 5% CO_2, unless stated differently. In all experiments, we used hMSCs at passage number four. Before cells seeding, surfaces were sterilized with 70% ethanol (Boom, 84010059.5000) and wetted in the basic medium during at least 2 h. We have previously systematically characterized in vitro lineage differentiation capacity, gene expression signature and in vivo capacity for ectopic bone formation of this donor (16) cells[52].

Table 1 summarize duration of the functional experiments, which are explained below.

Adipogenesis. To induce adipogenesis, hMSCs (d016) were cultured for 3 weeks in adipogenic medium (DMEM (Gibco, 41-965-062), 100 U/ml penicillin plus 100 mg/ml streptomycin (Gibco, 15140-122), 10% fetal bovine serum (FBS) (Sigma-Aldrich batch number 013M3396), 0.2 mM indomethacin (Sigma-Aldrich, 57413), 0.5 mM IBMX (Sigma-Aldrich, I5879), 10^{-6} M dexamethasone (Sigma-Aldrich, D8893), 10 ug/ml insulin (human, Sigma-Aldrich, I9278). Cells were seeded in three replicas at density 15,000 cells/cm^2 in 24 well plates, in basic medium. When the cells reached confluency on a flat surface, the culture medium was changed to adipogenic medium for 21 days, the medium was refreshed twice a week. To visualize lipids formation, cells were stained with oil red o (Sigma-Aldrich, O0625) as described before[63]. Briefly, cells were fixed with 10% formalin (Sigma-Aldrich, 501128) for 30 min at room temperature, rinsed with distilled water and washed with 60% isopropanol (VWR, 20922.364). The sample was stained for 5 min in freshly filtered oil red o solution (stock: 500 mg oil red o (Sigma-Aldrich, O0625), 99 ml isopropanol, 1 ml distilled water; stain: 15 ml stock plus 10 ml distilled water).

ALP induction.   hMSCs (d016) were seeded on topographies at density 15,000 cells/cm^2 in 96 well plates and cultured for 7 days in mineralization medium (basic medium supplemented with 2 mM L-glutamine (Gibco, 25030), 0.2 mM ascorbic acid (Sigma-Aldrich, A8960), 100 U/ml penicillin plus 100 mg/ml streptomycin (Gibco, 15140-122), 10^{-8} M dexamethasone (Sigma-Aldrich, D8893) plus 0.01 M β-glycerol phosphate (Sigma-Aldrich, 50020). Every surface had three replicas. The medium was changed every two days. Afterwards, cells were fixed with 4% Paraformaldehyde (PFA) (VWR, 30525), ALP was stained with immunofluorescent dye as described below.

Cell proliferation.   hMSCSc (d016) were starved for 24 h in αMEM (Gibco, 22-571-038) medium without FBS to synchronize their cell cycles. Then cells were trypsinized with trypsin–EDTA (0.05%) (Fisher Scientific, 25300054), trypsin was neutralized with the basic medium. Cells were seeded at a density of 10,000 cells/cm^2 in triplicates and cultured in the presence of 1% FBS (Sigma-Aldrich batch number 013M3396). Cells were allowed to adhere overnight and afterwards, cells were cultured for 48 h with 10 µM EdU component (Click-iT Plus EdU Alexa Fluor 594 Imaging Kit, Thermo Fischer Scientific) in the media. Next, incorporated EdU was imaged as described in the kit manual (Click-iT Plus EdU Alexa Fluo 594 Imaging Kit, Thermo Fischer Scientific).

Protein synthesis.   To assess global protein synthesis on the topographies we looked at global protein expression by virtue of L-methionine analogue incorporation. We seeded hMSCs (d016) in triplicates at density 20,000 cells/cm^2 on topographies in 96 well plates in basic medium and let them adhere for 24 h. Later, the medium was replaced with DMEM high glucose, no L-glutamine, no L-methionine, no L-cystine, (Thermo Fischer Scientific, 21013024) supplemented with 580 mg/l L-glutamine (Thermo Fisher Scientific, 25,030,081), and 63 mg/l L-Cystine dihydrochloride (Sigma-Aldrich, C6727-25G). Prior cells fixation we added L-homopropargylglycine, L-methionine analogue (Thermo Fischer Scientific) for 1.5 h. We further stained incorporated L-homopropargylglycine as described in the kit manual (Click-iT HPG Alexa Fluor Protein Synthesis Assay Kits, Thermo Fischer Scientific).

Apoptosis.   hMSCs (d016) were stained with dye for live cells, CellTracker Green CMFDA (Thermo Fischer Scientific, 11570166). During the next step, cells were seeded on topographies for 24 h in a basic medium at density 10,000 cells/cm^2 in 96 well plates. Next, we added a staurosporine (Abcam, ab120056), a protein kinase inhibitor, in the concentration of 0.5 µM and 250 ng/ml of annexin V (kind gift of Dr Leon J. Schurgers, Department of Biochemistry, University Maastricht) with a red fluorescent tag to detect apoptosis. To facilitate annexin V binding, the medium was supplemented with additional 0.7 mM of CaCl$_2$ (VWR, 10043-52-4). Cells were monitored in an environmental chamber, where the temperature was maintained at 37 °C in a humid atmosphere with 5% CO$_2$, during 66 h, 16 images per well were captured every 25 min. Imaging was acquired on Nikon Ti Eclipse epifluorescent inverted microscope.

Live imaging.   hMSCs (d016) were stained with dye for live cells, CellTracker Green CMFDA (Thermo Fischer Scientific, 11570166). Afterwards, cells were seeded on topographies for 24 h in a basic medium at density 10,000 cells/cm^2 in 96 well plates. Cells were monitored in an environmental chamber, where the temperature was maintained at 37 °C in a humid atmosphere with 5% CO$_2$, during 66 h. 16 images per well were captured every 25 min. Imaging was acquired on Nikon Ti Eclipse epifluorescent inverted microscope.

Staining and imaging.   Cells were fixed with 4% PFA (VWR, 30525) for ten minutes; next, samples were permeabilized with 0.01% Triton X-100 (VWR, 437002A) for 10 min with following blocking step in 0.5% bovine serum albumin (VWR, 421501J) at room temperature for 30 min. Samples were incubated with primary antibodies against ALP (RnD, mab1-148) overnight at 4 °C in 1:100 dilution. Labeling with secondary antibodies conjugated to fluorochrome Alexa Fluor 647 goat anti-mouse IgG (Thermo Fisher Scientific, A21236) in dilution 1:300 and Alexa Fluor 568 Phalloidin (Thermo Fisher Scientific, A12380) in dilution 1:200 was made during 1 h at room temperature, Next, after washing with PBS nuclear staining with 1:1,000 Hoechst 33342, trihydrochloride, trihydrate (Thermo Fischer, H1399) was done. Finally, the topographies were mounted on glass-bottom 24 well plates (Mobitec, 5231) or fluorocarbon-bottom 96 well plates (Mobitec, 5241) with ProLong Diamond antifade mountant (Thermo Fischer Scientific, P36965) after 2 PBS (VWR, E404-200TABS) washes. Imaging was acquired on Nikon Ti Eclipse epifluorescent inverted microscope.

Assessing cellular morphology with the cell painting assay.   hMSCs (d016) were seeded on topographies for 24 h in basic medium at density 10,000 cells/cm^2 in 24 well plates. Following the protocol of Bray et al., 30 min prior cells fixation 500 nM of MitoTracker deep red FM (Invitrogen, M22426) was added to cells. Afterwards, cells were washed with HBSS (Fisher Scientific, 11540476) and fixed with 4% PFA (VWR, 30525) for 20 min at room temperature. Samples were permeabilized with 0.1% Triton 100 (VWR, 437002A) for 10 min with the following washing with HBSS (Fisher Scientific, 11540476). We further added wheat germ agglutinin Alexa Fluor 594 conjugate (Invitrogen, W11262) at concentration 5 µg/ml, concanavalin A, Alexa Fluor 488 conjugate (Invitrogen, C11252) at concentration 50 µg/ml, SYTO 14 green fluorescent nucleic acid stain (Invitrogen, S7576) at concentration 10 µM, 1:40 Alexa Fluor 568 Phalloidin (Thermo Fischer Scientific, A12380) and 1:2,000 Hoechst 33342, trihydrochloride, trihydrate (Thermo Fischer, H1399) for 30 min. Finally, the topographies were mounted on glass-bottom 24 well plates (Mobitec) with ProLong Diamond Antifade Mountant (Thermo Fischer Scientific, P36965) after 2 PBS washes.

Transcriptional profiling.   hMSCs (d016) were seeded on selected topographies for 24 h in a basic medium at a density of 15,000 cells/cm² in 24 well plates in triplicate. Total RNA was isolated using the Nucleospin RNA isolation kit (Macherey–Nagel). Next, 100 ng of RNA was used to synthesize cRNA, following the instructions of the Illumina TotalPrep RNA amplification kit, and both RNA and cRNA quality was verified on a Bioanalyzer 2100 (Agilent Technologies).

Illumina HT-12 v4 expression Beadchips were used for microarray transcriptomics analysis. Briefly, 750 ng of cRNA was hybridized on the array overnight followed by washing and blocking. Next, by addition of streptavidin Cy-3, a fluorescent signal was developed and arrays were scanned on an Illumina Beadarray reader. Raw intensity values were corrected for background signal in BeadStudio (Illumina). Further data processing and statistical testing were performed using the online portal ArrayAnalysis (https://arrayanalysis.org/)[64]. Probe-level raw intensity values were quantile normalized and transformed using variance stabilization (VSN). To reduce the number of false positives, a detection threshold of 0.01 was used. A linear modelling approach with empirical Bayesian methods, as implemented in the R Limma package[65], was applied to look for differential expression of the resulting probe-level expression values. A false discovery rate (FDR) correction of the p-values was applied using the Benjamini and Hochberg method. Genes were considered differentially expressed at a corrected p-value < 0.05.

Pathway over-representation analysis was performed using the ConsensusPathDB (CPDB) tool (https://cpdb. molgen.mpg.de/), which provides a comprehensive pathway analysis covering most public resources of gene and protein interactions[32]. As input for the analysis, the set of differentially expressed genes mentioned above was used. To improve the statistical evaluation of the pathways, a background list containing all measured genes was also used in the calculations. Pathways with an FDR-corrected p-value < 0.05 were considered significant.

A protein network analysis was carried out in two steps. CPDB contains an induced network module, which uses the protein–protein interactions described in a large number of public resources, to build a network based on a list of input genes or proteins. As a first step, a network was generated using the same list of differentially expressed genes as used above in the pathway analysis, using a z-score threshold of 20. Only binary protein interactions of low, medium and high confidence were selected. Furthermore, the addition of intermediate genes to the network was allowed in order to improve inter-gene connectivity. In the second step, in order to better understand how this CPDB network may be regulated, the network was imported into CytoScape v.3.5.1[66] and the plugin CyTargetLinker was used to extend the network by adding transcription factors from the transcription factor target database TFe (Transcription Factor encyclopedia). TFe is a small-scale manual literature curation project containing 1531 human transcription factor-target interactions.

Selection of surfaces and clustering.   To be able to find cell shapes that were repeatedly induced by topography and remove any imaging artefacts or biological variation we performed the following filtering steps. Among other outlier detection tools we used, so called, 1.5 interquartile range (IQR) rule. It works as following: calculate 1st and 3rd quantile of the data (value of the first 25% and 75% of ordered data, correspondingly), difference between 3rd and 1st quantiles is IQR, all the data points that lie outside of the following range: 1st quantile − 1.5 × IQR, 3rd quantile + 1.5 × IQR, considered being outlines. First, we removed replicas based on cell density. We performed this on a per-surface basis, to avoid removal of the surfaces with a unique response. Replicas with too low or to high cell number were discarded using the 1.5 IQR rule. Afterwards, we removed cells that were outliers in terms of area and perimeter. Outliers were removed sequentially, first based on area and then based on perimeter values using the 1.5 IQR rule. We further employed Moutlier package in R[67] to remove outliers based on cell shapes. This approach allows finding outliers by using information from multiple features simultaneously. We did this by looking at the distribution of cell shapes in a 5-dimensional shape space and then removing outliers using the Mahalanobis-based outlier detection. These 5 shape features were selected from 11 features that have correlation index less than 0.7 between each other. Finally, for each surface, we retained only those replicates that were of good quality, those that were reproducible. The reproducibility was measured using the correlation of cell shapes between replicas. We excluded replicas that have correlation index less than 0.5, in comparison to all the replicas. Next, we summarized cell shape features per surface by taking median features of all the cells. Afterwards, we removed redundant surfaces, those that were highly correlated (Pearson correlation index above 0.95) to each other based on cell shape features. To reduce the dimensionality of the data further, we performed principal components analysis. We have chosen first seven principal components that captured the most variations of the data. We further used hierarchical clustering analysis with Ward linkage to find clusters of cell shapes. A distance matrix was calculated with Euclidean distances.

Image and data analysis.   Open-source software Cell Profiler 2.1 (CP) was used for image analysis[27]. In order to perform automated image analysis in CP, a robust pipeline able to recognize different cell features was built.

Training the model.   We employed Lasso regression model from Glmnet package in R[68] to find cell shape features that can predict the response of the phenotypic assays. The model was trained with Leave One Out Cross validation approach. We assessed the performance of the model by reporting the Pearson correlation index (r) between predicted and observed values on the training data. To show the goodness of fit we also reported R^2. The hyperparameter lambda was optimized separately for every model. Cells on flat control surfaces were excluded because of their distinct response, in comparison to the rest of topographies, which would bias the model.

Connectivity map analysis.   List of genes was uploaded into Connectivity Map (Subramanian, Narayan, et al. 2017), a library of gene expression signatures induced by chemical compounds or genetic interference and a connectivity score between the topography-induced DEGs and Connectivity Map DEGs were retrieved. We

focused on perturbant classes (PCL), which are groups of signatures in which members belong to the same gene family or are targeted by the same compound. To find cell shape signatures we have uploaded genes that have absolute correlation value above 0.5. Where genes with positive correlation values were loaded as "upregulated genes" and genes with negative correlation values were loaded as "downregulated genes. The results were further exported to csv file and visualized in R.

TopoChip and enlarged surfaces fabrication. The complete design of the TopoChip is described in detail elsewhere[24]. The TopoChip and enlarged surfaces were fabricated as described previously[69]. Briefly, a chromium mask for photolithography was constructed that contains the design for either the TopoChip or the topographies on expanded surfaces. Subsequently, through standard lithography and deep reactive etching, inverse structures of the topographies were generated on a silicon wafer. This mould was then utilized to make a positive mould in poly(dimethylsiloxane) (PDMS). Subsequently, a second negative mould in OrmoStamp hybrid polymer (micro resist technology Gmbh) was generated from the PDMS mould. This mould serves as the template for hot embossing the topographies in 190 μm thick PS films by applying 140 °C and 10 bar for 5 min. Before cell culture, PS films containing the topographical imprints were O2-plasma treated to enhance cell attachment.

Statistical analyses and data visualization. Statistical analysis was performed in R ver. 3.2.5[70], graphs were generated in R package ggplot2[71] and cowplot[72]. To determine a fold difference in the phenotypical assays, we have used original units as was measured by image analysis. To quantify the rank of the surfaces were sorted them based on the measured value.

References

1. Tosh, D. & Slack, J. M. W. How cells change their phenotype. *Nat. Rev. Mol. Cell Biol.* **3**, 187–194 (2002).
2. Potter, J. D. Morphogens, morphostats, microarchitecture and malignancy. *Nat. Rev. Cancer* **7**, 464–474 (2007).
3. Hu, X. *et al.* Microglial and macrophage polarization—New prospects for brain repair. *Nat. Rev. Neurol.* **11**, 56–64 (2015).
4. Liddiard, K. & Taylor, P. R. Understanding local macrophage phenotypes in disease: Shape-shifting macrophages. *Nat. Med.* **21**, 119 (2015).
5. Lin, J. Y. & Fisher, D. E. Melanocyte biology and skin pigmentation. *Nature* **445**, 843–850 (2007).
6. Semenza, G. L. Hypoxia-inducible factor 1: Control of oxygen homeostasis in health and disease. *Pediatr. Res.* **49**, 614–617 (2001).
7. Hahn, C. & Schwartz, M. A. Mechanotransduction in vascular physiology and atherogenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 53–62 (2009).
8. Nelson, W. J. Adaptation of core mechanisms to generate cell polarity. *Nature* **422**, 766–774 (2003).
9. Cui, Y. *et al.* Cyclic stretching of soft substrates induces spreading and growth. *Nat. Commun.* **6**, 6333 (2015).
10. Chen, Q. *et al.* Fate decision of mesenchymal stem cells: Adipocytes or osteoblasts?. *Cell Death Differ.* **23**, 1128–1139 (2016).
11. Cook, B., Hardy, R. W., McConnaughey, W. B. & Zuker, C. S. Preserving cell shape under environmental stress. *Nature* **452**, 361–364 (2008).
12. Gehwolf, R. *et al.* Pleiotropic roles of the matricellular protein Sparc in tendon maturation and ageing. *Sci. Rep.* **6**, 32635 (2016).
13. Lee, J., Abdeen, A. A., Wycislo, K. L., Fan, T. M. & Kilian, K. A. Interfacial geometry dictates cancer cell tumorigenicity. *Nat. Mater.* **15**, 856–862 (2016).
14. McBeath, R., Pirone, D. M., Nelson, C. M., Bhadriraju, K. & Chen, C. S. Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. *Dev. Cell* **6**, 483–495 (2004).
15. Vogel, V. & Sheetz, M. Local force and geometry sensing regulate cell functions. *Nat. Rev. Mol. Cell Biol.* **7**, 265–275 (2006).
16. Sun, Z., Guo, S. S. & Fässler, R. Integrin-mediated mechanotransduction. *J. Cell Biol.* **215**, 445–456 (2016).
17. Chiquet, M., Gelman, L., Lutz, R. & Maier, S. From mechanotransduction to extracellular matrix gene expression in fibroblasts. *Biochim. Biophys. Acta (BBA) Mol. Cell Res.* **1793**, 911–920 (2009).
18. Papachroni, K. K., Karatzas, D. N., Papavassiliou, K. A., Basdra, E. K. & Papavassiliou, A. G. Mechanotransduction in osteoblast regulation and bone disease. *Trends Mol. Med.* **15**, 208–216 (2009).
19. Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science (80-).* **330**, 55–60 (2010).
20. Zhang, J. *et al.* Regulation of cell proliferation of human induced pluripotent stem cell-derived mesenchymal stem cells via ether-à-go-go 1 (hEAG1) potassium channel. *Am. J. Physiol. Physiol.* **303**, C115–C125 (2012).
21. Roca-Cusachs, P., Gauthier, N. C., Del Rio, A. & Sheetz, M. P. Clustering of α5β1 integrins determines adhesion strength whereas αvβ3 and talin enable mechanotransduction. *Proc. Natl. Acad. Sci.* **106**, 16245–16250 (2009).
22. Wiche, G., Osmanagic-Myers, S. & Castanon, M. J. Networking and anchoring through plectin: A key to IF functionality and mechanotransduction. *Curr. Opin. Cell Biol.* **32**, 21–29 (2015).
23. Iskratsch, T., Wolfenson, H. & Sheetz, M. P. Appreciating force and shape—The rise of mechanotransduction in cell biology. *Nat. Rev. Mol. Cell Biol.* **15**, 825–833 (2014).
24. Unadkat, H. V. *et al.* An algorithm-based topographical biomaterials library to instruct cell fate. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16565–16570 (2011).
25. Masaki, C., Schneider, G. B., Zaharias, R., Seabold, D. & Stanford, C. Effects of implant surface microtopography on osteoblast gene expression. *Clin. Oral Implants Res.* **16**, 650–656 (2005).
26. Hulshof, F. F. B. *et al.* Mining for osteogenic surface topographies: In silico design to in vivo osseo-integration. *Biomaterials* **137**, 49–60 (2017).
27. Carpenter, A. E. *et al.* Cell Profiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
28. Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629 (2002).
29. Fletcher, D. A. & Mullins, R. D. Cell mechanics and the cytoskeleton. *Nature* **463**, 485 (2010).
30. Ron, A. *et al.* Cell shape information is transduced through tension-independent mechanisms. *Nat. Commun.* **8**, 2145 (2017).
31. Anselme, K., Wakhloo, N. T., Rougerie, P. & Pieuchot, L. Role of the nucleus as a sensor of cell environment topography. *Adv. Healthc. Mater.* **7**, 1701154 (2018).

32. Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 (2017).

33. Kamburov, A. *et al.* ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Res.* **39**, D712–D717 (2010).

34. Beijer, N. R. M. *et al.* Dynamic adaptation of mesenchymal stem cell physiology upon exposure to surface micropatterns. *Sci. Rep.* **9**, 1–14 (2019).

35. Castro-Raucci, L. *et al.* Titanium with nanotopography induces osteoblast differentiation by regulating endogenous bone morphogenetic protein expression and signaling pathway. *J. Cell. Biochem.* **117**, 1718–1726 (2016).

36. Nassiri, I. & McCall, M. N. Systematic exploration of cell morphological phenotypes associated with a transcriptomic query. *Nucleic Acids Res.* **46**, e116–e116 (2018).

37. Kilian, K. A., Bugarija, B., Lahn, B. T. & Mrksich, M. Geometric cues for directing the differentiation of mesenchymal stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4872–4877 (2010).

38. Muramatsu, T. *et al.* YAP is a candidate oncogene for esophageal squamous cell carcinoma. *Carcinogenesis* **32**, 389–398 (2010).

39. Luxenburg, C., Amalia Pasolli, H., Williams, S. E. & Fuchs, E. Developmental roles for Srf, cortical cytoskeleton and cell shape in epidermal spindle orientation. *Nat. Cell Biol.* **13**, 203–214 (2011).

40. Lamb, R. F. *et al.* Essential functions of ezrin in maintenance of cell shape and lamellipodial extension in normal and transformed fibroblasts. *Curr. Biol.* **7**, 682–688 (1997).

41. Moroishi, T., Hansen, C. G. & Guan, K.-L. The emerging roles of YAP and TAZ in cancer. *Nat. Rev. Cancer* **15**, 73–79 (2015).

42. Du, S. *et al.* Fructose-bisphosphate aldolase a is a potential metastasis-associated marker of lung squamous cell carcinoma and promotes lung cell tumorigenesis and migration. *PLoS ONE* **9**, e85804 (2014).

43. Kadandale, P., Stender, J. D., Glass, C. K. & Kiger, A. A. Conserved role for autophagy in Rho1-mediated cortical remodeling and blood cell recruitment. *Proc. Natl. Acad. Sci.* **107**, 10502–10507 (2010).

44. Santamaría, D. *et al.* Cdk1 is sufficient to drive the mammalian cell cycle. *Nature* **448**, 811 (2007).

45. Vairapandi, M., Balliet, A. G., Hoffman, B. & Liebermann, D. A. GADD45b and GADD45g are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. *J. Cell. Physiol.* **192**, 327–338 (2002).

46. Fischer, M., Quaas, M., Steiner, L. & Engeland, K. The p53–p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes. *Nucleic Acids Res.* **44**, 164–174 (2015).

47. David, M., Petit, D. & Bertoglio, J. Cell cycle regulation of Rho signaling pathways. *Cell Cycle* **11**, 3003–3010 (2012).

48. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* **9**, 153 (2009).

49. Siddappa, R., Licht, R., van Blitterswijk, C. & de Boer, J. Donor variation and loss of multipotency during in vitro expansion of human mesenchymal stem cells for bone tissue engineering. *J. Orthop. Res.* **25**, 1029–1041 (2007).

50. Alves, H. *et al.* A mesenchymal stromal cell gene signature for donor age. *PLoS ONE* **7**, e42908 (2012).

51. Vermeulen, S. *et al.* Mechanotransduction is a context-dependent activator of TGF-β signaling in mesenchymal stem cells. *Biomaterials* **259**, 120331 (2020).

52. Mentink, A. *et al.* Predicting the therapeutic efficacy of MSC in bone tissue engineering using the molecular marker CADM1. *Biomaterials* **34**, 4592–4601 (2013).

53. Jain, N., Iyer, K. V., Kumar, A. & Shivashankar, G. V. Cell geometric constraints induce modular gene-expression patterns via redistribution of HDAC3 regulated by actomyosin contractility. *Proc. Natl. Acad. Sci.* **110**, 11349–11354 (2013).

54. Mascharak, S. *et al.* YAP-dependent mechanotransduction is required for proliferation and migration on native-like substrate topography. *Biomaterials* **115**, 155–166 (2017).

55. Brusatin, G., Panciera, T., Gandin, A., Citron, A. & Piccolo, S. Biomaterials and engineered microenvironments to control YAP/TAZ-dependent cell behaviour. *Nat. Mater.* **17**, 1063–1075 (2018).

56. Pocaterra, A., Romani, P. & Dupont, S. YAP/TAZ functions and their regulation at a glance. *J. Cell Sci.* **133**, 230425. https://doi.org/10.1242/jcs.230425 (2020).

57. Klinker, M. W., Marklein, R. A., Lo Surdo, J. L., Wei, C.-H. & Bauer, S. R. Morphological features of IFN-γ-stimulated mesenchymal stromal cells predict overall immunosuppressive capacity. *Proc. Natl. Acad. Sci.* **114**, E2598–E2607 (2017).

58. Zhang, Y. *et al.* iPSC-MSCs with high intrinsic MIRO1 and sensitivity to TNF-α yield efficacious mitochondrial transfer to rescue anthracycline-induced cardiomyopathy. *Stem Cell Rep.* **7**, 749–763 (2016).

59. Ding, Y. *et al.* Rap1 deficiency-provoked paracrine dysfunction impairs immunosuppressive potency of mesenchymal stem cells in allograft rejection of heart transplantation. *Cell Death Dis.* **9**, 386 (2018).

60. Jiang, D. *et al.* Mitochondrial transfer of mesenchymal stem cells effectively protects corneal epithelial cells from mitochondrial damage. *Cell Death Dis.* **7**, e2467–e2467 (2016).

61. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* **26**, 694–701 (2007).

62. Hebels, D. G. A. J., Carlier, A., Coonen, M. L. J., Theunissen, D. H. & de Boer, J. cBiT: A transcriptomics database for innovative biomaterial engineering. *Biomaterials* **149**, 88–97 (2017).

63. De Boer, J., Wang, H. J. & Van Blitterswijk, C. Effects of Wnt signaling on proliferation and differentiation of human mesenchymal stem cells. *Tissue Eng.* **10**, 393–401 (2004).

64. Eijssen, L. M. T. *et al.* A user-friendly workflow for analysis of Illumina gene expression bead array data available at the arrayanalysis. org portal. *BMC Genom.* **16**, 482 (2015).

65. Smyth, G. K. Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer, New York, 2005).

66. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

67. Gupta, M., Gao, J., Aggarwal, C. & Han, J. Outlier detection for temporal data. *Synth. Lect. Data Min. Knowl. Discov.* **5**, 1–129 (2014).

68. Hastie, T. & Qian, J. Glmnet vignette. http//www.web.stanford.edu/Hast.pdf. Accessed 20 Sep 2016 (2014).

69. Vermeulen, S. *et al.* Identification of topographical architectures supporting the phenotype of rat tenocytes. *Acta Biomater.* **83**, 277–290 (2019).

70. Ihaka, R. & Gentleman, R. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).

71. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2016).

72. Wilke, C. O. cowplot: Streamlined plot theme and plot annotations for 'ggplot2'. *CRAN Repos* (2016).

Acknowledgements

Author contributions

A.V. and J.B. wrote the main manuscript text, A.V., S.V., M.K., N.R., R.R., N.B., J.P. performed experiments. A.V., A.C, D.H., S.S., S.E. performed data analysis and prepared figures. A.C. and J.B. provided resources and facilities. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-76019-z.

Correspondence and requests for materials should be addressed to J.B.

Reprints and permissions information is available at www.nature.com/reprints.

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