BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 21 2010, pages 2760–2766
doi:10.1093/bioinformatics/btq525

Systems biology

Advance Access publication September 16, 2010

Spatial quantification and classification of skin response following perturbation using organotypic skin cultures

Thora Pommerencke[1,2], Kathi Westphal[1,2], Claudia Ernst[1,2], Kai Safferling[1,2], Hartmut Dickhaus[1,2], Thorsten Steinberg[3], Pascal Tomakidi[3] and Niels Grabe[1,2,*]

[1]Hamamatsu Tissue Imaging and Analysis (TIGA) Center, BIOQUANT, University Heidelberg, [2]Institute for Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg and [3]Department of Oral Biotechnology, University of Freiburg, Freiburg, Germany

Associate Editor: Martin Bishop

ABSTRACT

Motivation: For a mechanistic understanding of skin and its response to an induced perturbation, systems biology is gaining increasing attention. Unfortunately, quantitative and spatial expression data for skin, like for most other tissues, are almost not available.

Results: Integrating organotypic skin cultures, whole-slide scanning and subsequent image processing provides bioinformatics with a novel source of spatial expression data. We here used this approach to quantitatively describe the effect of treating organotypic skin cultures with sodium dodecyl sulphate in a non-corrosive concentration. We first measured the differentiation-related spatial expression gradient of Heat-Shock-Protein 27 in a time series of up to 24 h. Secondly, a multi-dimensional tissue classifier for predicting skin irritation was developed based on abstract features of these profiles. We obtained a high specificity of 0.94 and a sensitivity of 0.92 compared with manual classification. Our results demonstrate that the integration of tissue cultures, whole-slide scanning and image processing is well suited for both the standardized data acquisition for systems biological tissue models and a highly robust classification of tissue responses.

Contact: niels.grabe@bioquant.uni-heidelberg.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 13, 2010; revised on August 25, 2010; accepted on September 9, 2010

1 INTRODUCTION

The human body is shielded against outside influences by the skin, exhibiting a carefully regulated equilibrium of cell proliferation, differentiation and cell death. Even though this homeostasis guarantees a monthly self renewal, the exposition to harmful substances can lead to severe perturbations of the skin (Perry and Trafeli, 2009; Schnuch, 2007; Slodownik *et al*., 2008). The skin therefore is a main target in toxicity testing and drug development. Systems biology is gaining increasing attention in these areas (Andersen and Krewski, 2009; Ho and Lieu, 2008; Krewski *et al*., 2009), offering a novel approach to cope with the complexity of the skin as a biological system (Grabe and Neuber, 2005). Besides mimicking the physiologic system's processes, the potential of respective computational models lies in the prediction of skin response following e.g. a chemical exposure.

For identifying those system elements and pathways mainly contributing to the observed system response, a standardized experimental setup is necessary. On the single cell level high content screening (HCS) based on cell lines or strains allows the quantification of phenotypic changes after perturbation by automated fluorescence microscopy and subsequent image analysis (Giuliano *et al*., 1997; Thomas, 2010). However, approaches addressing individual cells do not reflect the spatial complexity of the human skin with its stratified organization. For example, substances applied to the surface of normal skin have to pass the barrier built by the stratum corneum and the perturbation induced is buffered by robust homeostatic molecular networks leading to differentiation-dependent expression changes. Therefore, predictions based on these simple single cell model assumptions are not necessarily transferable to human skin. According to our best knowledge, up to now there is no suitable data acquisition method for the standardized and automated study of spatial skin response after perturbation.

Recently, automated whole slide scanners (Virtual Microscopy) became available, offering the opportunity to analyze whole tissue sections in high throughput. Thus, the combination of whole slide scanners with industrially standardized tissue cultures could open up a novel way for the systematic characterization of perturbations induced to the human skin.

Industrially manufactured organotypic skin cultures offer a well organized, reproducible spatial morphology and could therefore serve as a standardized source for spatial organized data. Such full thickness models are composed of epidermal keratinocytes growing air exposed on a matrix of native collagen type I containing viable fibroblasts (Bell *et al*., 1981). A perturbation inflicted upon such a skin tissue culture will, depending on its strength, alter tissue homeostasis hallmarked by a specific spatial gene- and protein-expression. We have shown previously how such differentiation-dependent protein expression can be measured by quantitative spatial profiles (QSPs) based on automated image analysis of native skin sections, stained by indirect immunofluorescence (Grabe *et al*., 2007; Pommerencke *et al*., 2007).

We therefore set out to develop a novel data generation pipeline for systems biology experiments comprised of perturbing a standardized skin tissue culture, histological sectioning, immunofluorescence staining, whole slide scanning and automated

*To whom correspondence should be addressed.

image analysis. We consider this approach to open up the possibility for a standardized data acquisition of expression pattern changes reflecting skin response after perturbation, thus being well qualified for the toxicological study and characterization of chemicals. We assessed the feasibility of the proposed system by disturbing tissue homeostasis of skin cultures through the treatment with the detergent and classical irritant sodium dodecyl sulphate (SDS) and measuring the induced change as an altered Heat-Shock-Protein HSP27 expression (Boxman *et al*., 2002) deploying quantitative spatial profiling. To evaluate the reproducibility and suitability of the quantified tissue response for testing skin irritation, we built a classifier distinguishing between irritated and non-irritated skin samples.

2 METHODS

2.1 Experimental setup

MatTek EpiDermFT^{TM} skin cultures were either treated with phosphate buffered saline (PBS) or 0.4% SDS for varying durations of 1, 6, 16 and 24 h or received no treatment. In case of PBS and SDS, 25 µl of the respective solution were applied on a mesh on top of the cultures. Thereafter, the cultures were cryo conserved and cut into sections of 6 µm. Each mode of treatment (e.g. 16 h SDS) was applied to four cultures and for each culture about five sections underwent an immunhistological fluorescence triple staining. For the subsequent automatic segmentation of the epithelium, a reference staining is required delineating the epithelial area while omitting the stratum corneum and which is stable against SDS. A staining that proved to be appropriate (at least for a good approximation of the epithelium) comprised a blue nuclear stain using the dye 4′,6-Diamidin-2-phenylindol (DAPI). Labeling the basal lamina was performed with an antibody against the membrane protein Laminin 5 labeled with Alexa 488. Additionally, the biomarker for the quantification of skin irritation (i.e. Hsp 27) was marked in red (Alexa 594).

2.2 Image acquisition

The fluorescence sections were imaged with the Nanozoomer HT from Hamamatsu Photonics capable of scanning whole slides. The slides were scanned at 20× magnification (with a resolution of 0.46 µm/pixel) in three z-layers with a spacing of 2 µm each. In the resulting 'virtual slide' those tissue regions to be analyzed were selected in a computer-assisted way. For each region automatically the best focused z-layer was determined and saved as an image to be available for the following segmentation and profiling algorithms.

2.3 Segmentation of the epithelium

Without a staining marking the whole epithelium, the epithelial area had to be approximated using the staining of the basal lamina and the cell nuclei. Therefore, the first step was the segmentation of the basal lamina, scanned in the green channel. After contrast enhancement and noise reduction mainly a thresholding based on the Otsu's method (Otsu, 1979) was applied. To exclude unspecific and diffuse staining observed in the stratum granulosum and corneum from the determined mask, the thresholding is combined with an edge detection based on the Canny method (Canny, 1986). The operator was parameterized to detect long stretched edges like the basal lamina.

In a second step, cell nuclei were detected in the blue image channel based on a stepwise thresholding method to account for strongly as well as weakly stained nuclei. Only image regions close to strong intensity gradients were considered. Nuclei clusters were identified by size and split using a watershed method (Meyer, 1994) with local maxima as seeds. Finally, false positive detected cell nuclei were filtered out based on the object features size, roundness, inner intensity variation and contrast (Pommerencke *et al*., 2009).

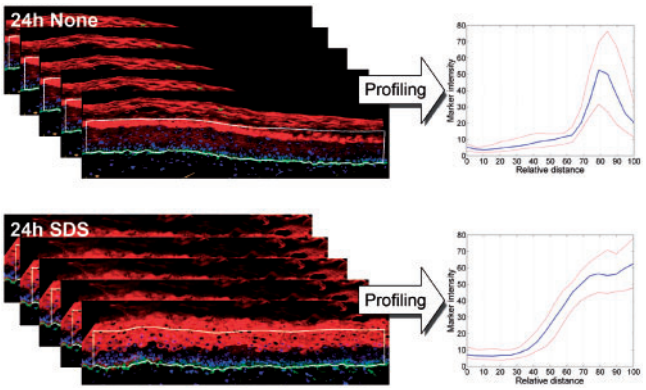

Fig. 1. Protein expression profiling on fluorescent images. Fluorescent images from different tissue samples or locations are automatically segmented based on nuclei and basal lamina information. The determined epithelial region is delineated by a white line. For all images of one treatment mode the median protein expression in the epithelium (intensity in the red channel) is profiled against the relative distance from basal lamina to surface.

Ring-like stained nuclei, identified by a high ratio of border intensity to inner intensity, were excluded from the watershed procedure.

For artificial skin cultures, we can assume that the epithelium exhibits almost the same thickness (distance from basal lamina to surface) over the whole section neglecting border regions. Therefore, the surface (or more precisely the transition between stratum granulosum and stratum corneum) can be approximated by smoothing the detected basal lamina and a succeeding shift of the resulting line towards the epithelium. A shift of 1.4 times the maximum (95% quantile) nuclei distance from basal lamina in the respective epithelial region proofed to be well fitting. The epithelial side exhibits a higher nuclei density than the side of the culture's collagen matrix. Therefore, the shifting direction could be determined by the nuclei density in a band around the basal lamina.

For a closed epithelial area with sides vertical to the surface, the surface pixels were fitted by a line. The endpoints of the surface line were connected with the detected basal lamina by orthogonals to the surface.

2.4 Relative distance and biomarker profiling

The spatial biomarker profiles were generated by measuring the mean biomarker intensity (i.e. Hsp 27 recorded in the red channel) in multiple bands inside the epithelium. The number and width of these bands is defined by intervals of relative distance (Pommerencke *et al*., 2008), where the basal lamina is set to a distance of 0% and the surface to 100%. Finally, the averaged marker intensity is plotted against the corresponding relative distance for all intervals. For further analysis, the median of all profiles from either each treatment mode (e.g. 6 h SDS) or each individual culture were calculated as well as for the estimation of the confidence interval the 25% and 75% quantiles (red lines in Fig. 1). Beside a marker profile based on the whole epithelial area, also profiles quantifying the marker intensity inside the nuclei against relative distance were generated.

2.5 Feature extraction from expression profiles

Changes of the tissue's state (either due to a tissue response following treatment or a disease) lead to changes of the histological protein expression patterns and hence to a shift or modified height of the biomarker profile. For quantification and later classification of the tissue responses, the information in the biomarker profiles was consolidated into intelligible features characteristic of the tissue's state. In case of the protein Hsp27, SDS treatment leads to a premature and cytoplasmic expression pattern with planar appearance (see Fig. 1). In the profile, a premature expression is reflected

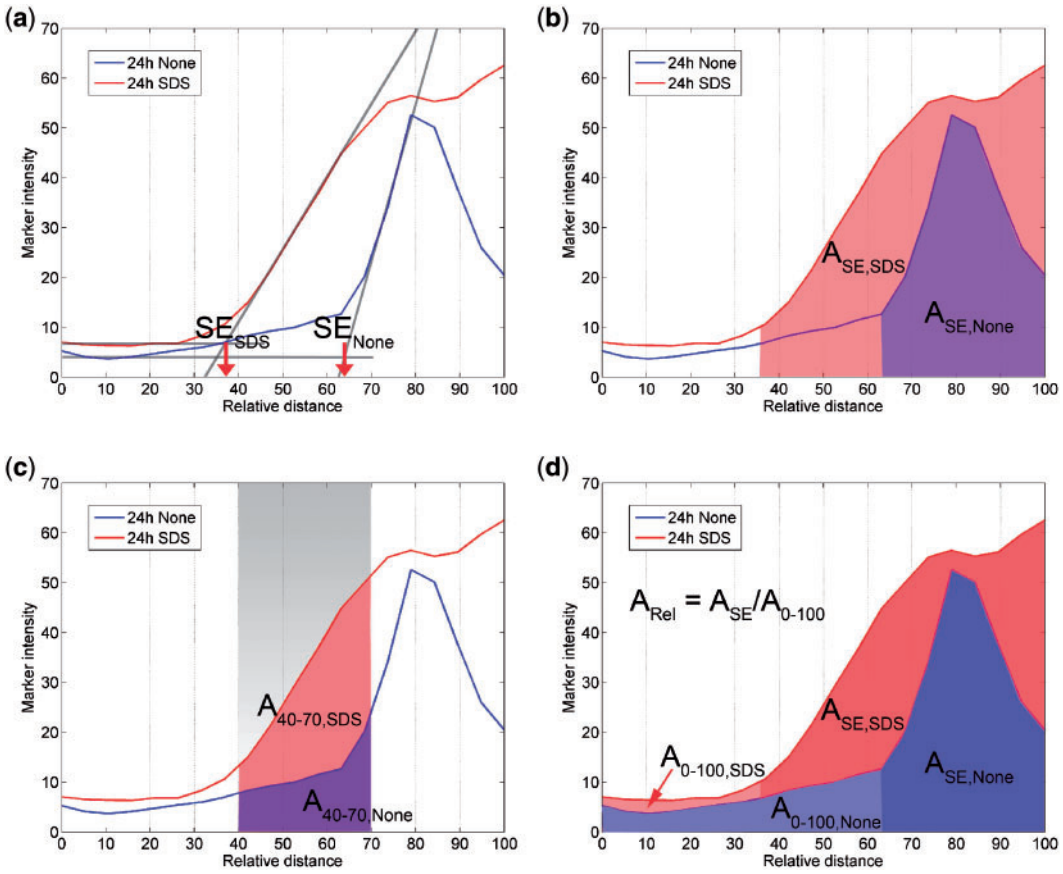

Fig. 2. Feature extraction from the QSPs. (**a**) Shows the determination of the starting point of expression (*SE*). In (**b**) the area of expression is coloured (A_{SE}). (**c**) Highlights the area between 40% and 70% relative distance (A_{40-70}). (**d**) Illustrates the calculation of the relative expression area A_{Rel} by dividing the A_{SE} by the total area below the QSP.

by an early increase of the marker intensity while the planar histological expression pattern is reflected by a high integral of the profile from the starting point of expression to profile's end.

In the following, we present five possible features which are based on the described observations:

- *StartExpression (SE)*: to determine the starting point of expression the profile curve with the highest slope is fitted with a line. The intersection of that line with the horizontal through the minimum marker intensity was set as the starting point of expression (Fig. 2a). To approximate the region of highest slope, the intensity difference (*difInt*) between maximum and minimum intensity left to the maximum is calculated. For line fitting those discrete profile points left to the maximum are considered being in the range of [*difInt*/2 − *difInt*/5, *difInt*/2 + *difInt*/5]. If there are no profile points equalling the interval limits, those points closest to the interval limits, but outside the range, are chosen for end points.
- *ProfileExpressionArea* (A_{SE}): the whole extent of expression was calculated by summing up all marker intensities beginning at *SE* (Fig. 2b).

$$A_{SE} = \int_{SE}^{100} f_{QSP}(x)\mathrm{d}x \tag{1}$$

- *ProfileExpressionArea40–70* (A_{40-70}): As the main change of expression can be observed between 40% and 70% relative distance, also the amount of marker intensities confined to this distance range was calculated (Fig. 2c).

$$A_{40-70} = \int_{40}^{70} f_{QSP}(x)\mathrm{d}x \tag{2}$$

- *RelativeExpressionArea* (A_{Rel}): As a measure independent of staining intensity we also calculated the proportion defined by the A_{SE} compared with the whole profile area. A_{Rel} is calculated as the ratio of A_{SE} to the total profile integral (Fig. 2D).

$$A_{Rel} = \frac{A_{SE}}{\int_{0}^{100} f_{QSP}(x)\mathrm{d}x} \tag{3}$$

- *NucleiIntensity40–70*: As literature reports a translocation of Hsp27 into nuclei following SDS treatment (Boxman *et al.*, 2002), also the average marker intensity located in the nuclei was measured in the range of 40% and 70% relative distance. For this feature, a QSP of marker intensity inside the nuclei against relative distance was calculated.

2.6 Tissue classification

In any kind of toxicological application, single tissue samples of unknown state have to be classified according to quantitative parameters describing the tissue's state. We here ask, to what extent the presented profile features allow the assignment of a degree of irritation to a single culture. As a reference for an automatic classification, all cultures were manually classified based

$$F_{opt} = \{F \subseteq \{f_1, \ldots, f_5\} \mid \sigma_F \text{ is minimal}\}$$

$$\sigma_F = \frac{1}{n}\sum_{i=1}^{n} |k_i - p(x_i, \theta_i)|$$

$$p(x_i, \theta_i) = \begin{cases} 1 & \text{if } d_F(x_i) < \theta, \\ 0 & \text{else.} \end{cases}$$

$$\theta_i = \{\theta \in \mathbb{R}_0^+ \mid d_{ROC}(\theta, i) \text{ is minimal}\}$$

$$d_{ROC}(\theta, i) = \left| \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} \text{sensitivity } (\{k_j\}, \{p(x_j, \theta)\}) \forall_{j \neq i} \\ 1 - \text{specificity } (\{k_j\}, \{p(x_j, \theta)\}) \forall_{j \neq i} \end{pmatrix} \right|$$

Fig. 3. Formalized classification process. The first line describes the feature selection, in which the optimal subset F_{opt} is defined as that subset F of all features $f_{1..5}$ yielding the minimum classification error σ_F. The classification error is calculated as the average deviation of the control classes k and the predicted classes p by the classification model in a leave-one-out cross-validation. A culture x_i is predicted to belong to the 'none class' ($p_i = 1$) if its Mahalanobis distance $d_F(\boldsymbol{x}_i)$ falls below a certain threshold θ_i determined on the remaining $n-1$ cultures. This threshold is defined as that Mahalanobis distance for which the distance $d_{ROC}(\theta, i)$ of the ROC curve to the upper left corner is minimal. The ROC curve is determined by calculating the sensitivity and specificity on the $n-1$ cultures for varying thresholds θ of the Mahalanobis distance.

on the visual inspection of one exemplary stained section for each culture. According to the observed HSP27 expression pattern, two classes were discriminated (no or strong change of expression).

For the automatic classification, the Mahalanobis distance (Duda *et al.*, 2000) was used, which independently quantifies the distance of one sample to a control group in a multi-dimensional space while considering correlation between the different dimensions. Each dimension reflects one of the basic features (*SE*, A_{SE}, etc.). In this way, the Mahalanobis distance maps several features into one single measure, which then can be used for classification by thresholding. As a control group, we selected the untreated cultures ('none cultures'). The Mahalanobis distance d is calculated as:

$$\mathrm{d}(x) = \sqrt{(x - \mu_{\mathrm{None}})^T \Sigma^{-1} (x - \mu_{\mathrm{None}})} \quad (4)$$

with $\boldsymbol{x}$ denoting the n-dimensional feature vector of one culture and μ the n-dimensional mean vector of the None group. The superfix$T$is the matrix transpose and Σ^{-1} the inverse of the n-by-n covariance matrix.

For the class assignment (none versus irritated) a separating threshold for the Mahalanobis distance is required. For this, a ROC curve (Zweig and Campbell, 1993) was generated by plotting the sensitivity against 1-specificity for each possible threshold and resulting class assignments. An optimal threshold for the separation of the 'none class' and the 'irritant class' refers to the point on the receiver operating characteristic (ROC) curve with the minimum Euclidean distance to the upper left corner (100% sensitivity, 100% specificity). For a formalization of the classification process see Fig. 3.

To assess the accuracy and the predictive power of the presented classification method, a leave-one-out-cross-validation (Kohavi, 1995) was performed. The validation results for all test samples were averaged, giving an estimate for the performance of the classifier on an independent dataset. This validation was carried out for each possible feature subset. Finally, that subset yielding the smallest classification error was selected. The algorithms described above were implemented in Matlab R2008b.

3 RESULTS

For the quantification of the treatment effect, the median of all biomarker profiles for each treatment mode (e.g. 24 h SDS) was calculated. Based on these profiles, the five presented profile features were determined (except for the *NucleiIntensity40-70*, which is not extracted from the profiles, but directly from the section images). Figure 4 shows the change of the starting point of expression depending on time as well as treatment mode. *SE* ranges between 38% (SDS) and 62% (None) relative distance. The observation that most of the expression changes occurred during 40–70% relative distance motivated the definition of the features A_{40-70} and *NucleiIntensity40-70*. In Figure 4, the time-dependent alterations of the individual expression profile features observed in SDS-treated tissue cultures are compared to 'none cultures'.

The strongest change is noticeable in the feature A_{40-70}. At a treatment duration of 24 h, the area increases to 2.6-fold of the area measured in 'none cultures'. But also *SE* drops to 60% of the starting point in 'none cultures'. However, we also measured a change in PBS-treated cultures. Comparing SDS-treated cultures to PBS-treated cultures discovers a 1.9-fold change in ProfileExpressionArea40-70 at 16 h treatment duration.

For the classification of single cultures, the profiles of all sub-images belonging to one culture were averaged. Based on the resulting median profile, the culture features were calculated. The feature subset yielding the minimal classification error after translation into the Mahalanobis distance consists of the four features *SE*, A_{SE}, A_{40-70} and A_{Rel}. The calculated Mahalanobis distance for each culture is shown in Figure 4. The average threshold chosen for classification is illustrated as a horizontal in that diagram. The bars of automatically irritated classified cultures are striped and those manually classified by visual inspection are coloured in red. The validation of the classification procedure with the Leave-one-out-method yielded a classification error of 0.064, an average specificity of 0.94 and a sensitivity of 0.92 compared with manual classification (see exemplary ROC-curve in Fig. 4d).

4 DISCUSSION

We here presented an approach for the quantitative capturing and subsequent classification of tissue response following substance-induced perturbations. The approach comprises standardized skin tissue cultures, whole slide-scanning, automated image processing and classification. We consider such a system ideally suited for the quantitative toxicological study of chemicals. Since chemicals of similar mechanisms of toxicity are expected to show similar phenotypic effects, the endpoints of this method also qualify for classification.

In contrast to HCS, allowing the spatial quantification of molecules in single cell strains or cultures, our approach addresses the question of how a full tissue with its complex structure quantitatively reacts to perturbations. Besides the spatial layering, a skin culture also exhibits a barrier formed by the stratum corneum that is not present in cell strains. Substances applied to the surface of normal skin have to pass this barrier before perturbations can be induced at all.

Common *in vitro* methods for skin irritation testing like the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) test measure inside a homogenate thus loosing all spatial information. Furthermore, the irritative effect is correlated to the degree of cellular destruction (Coquette *et al.*, 2003; Mosmann, 1983). This makes these methods blind for early and subtle modifications of the tissue before corrosion sets in.

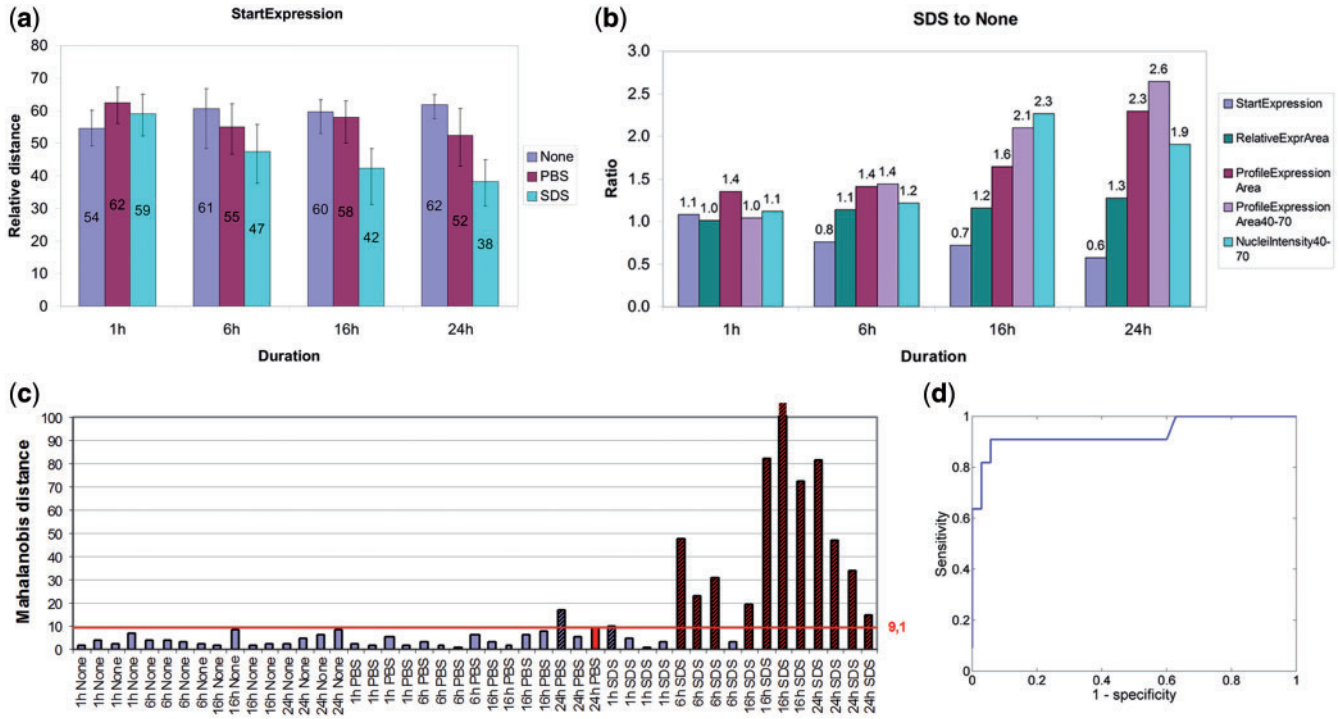

Fig. 4. Classification of skin cultures based on quantified expression changes following irritation. (**a**) Shows the median starting point of expression with 25% and 75% quantile for each time point and treatment mode. (**b**) For each of the five profile feature the ratio of the average SDS culture value and that value determined in the None cultures is plotted. (**c**) For each culture its Mahalanobis distance based on the four features SE, A_{SE}, A_{40-70} and A_{Rel} is shown. The average threshold used for classification is indicated by a red horizontal line. Those cultures automatically set as irritated are striped and those cultures being identified during visual inspection to have a changed expression pattern are marked in red. (**d**) Exemplary ROC curve during leave-one-out cross-validation.

Therefore, we here presented an approach based on measuring the spatial and quantitative change of protein expression in intact, immunofluorescence-stained tissue sections. Recording modifications at the protein level guarantees a certain expression stability, as not being as transient as mRNA expression changes. Based on the determined QSPs a multi-parametrical and automatic classifier was built, assigning to each tissue culture the status irritated or non-irritated.

In our experimental setting, tissue cultures were treated for different time durations with the standard irritant SDS. The visual inspection of the cultures identified a considerable change in Hsp27 expression (premature, planar and increased) after ~6 h of treatment. This change is clearly reflected in the QSPs when measuring marker intensity against the normalized distance to basal lamina. To quantify such changes in the QSPs, we presented five different features extracted from the profiles (SE, A_{SE}, A_{40-70}, A_{Rel} and *NucleiIntensity40-70*). The premature expression is reflected in an early increase of the QSP. With SE we were able to determine the onset of expression. The increase of total expression is manifested in an increase of the QSP's integral i.e. the area below the profile. We therefore developed features quantifying the profile's area (A_{SE}, A_{40-70}, A_{Rel}). The determination of SE for all modes of treatment revealed that most changes occur between 40% and 70% relative distance. Hence, the features A_{40-70} and *NucleiIntensity40-70* focus on this part of the QSP. The one-by-one comparison of all features identified A_{40-70} as the feature with the most prominent changes after treatment. However, the feature subset selection for classification by first translating the multi-dimensional problem into the Mahalanobis distance and then minimizing the classification error in a leave-one-out-validation revealed an optimal subset consisting of SE, A_{SE}, A_{40-70} and A_{Rel}.

Interestingly, the feature *NucleiIntensity40-70* reflecting the translocation of Hsp27 into the nuclei as described in literature did not improve the classification result. However, also the visual inspection of the stained sections did not reveal a noticeable translocation. The increase during treatment mainly arises from the general expression increase in the area between 40% and 70% relative distance.

Based on the selected feature subset we were able to reproduce the manual classification with our automatic classification by the Mahalanobis distance to 'none cultures'. We report a specificity of 0.94 and a sensitivity of 0.92 compared with the manual classification results based on the visually observed phenotype. We therefore conclude that the presented features are well suited for the characterization of the tissue's response. The classification result in conjunction with the good confidence intervals of the calculated profiles argues for a well reproducible tissue response. We showed that the observed spatial changes could be described by profile features allowing a precise classification. This indicates a highly suitable method for a standardized data acquisition in the context of skin irritation.

To further increase the accuracy of our profiling method a staining unambiguously delineating the epithelium would be preferable to improve tissue segmentation. In the presented method, the

epithelium is approximated based on a basal lamina and nuclear staining assuming a constant epithelial thickness. However, our literature search as well as own experiments did not reveal a suitable protein marker stable under topical treatment with SDS.

As a marker for skin irritation we chose Hsp27 based on literature (Boxman *et al*., 2002). Nevertheless, many other potential biomarkers, as reported in association with skin irritation, could be deployed with the presented method. Usually, such biomarkers are identified by complementary techniques like RNA and protein microarrays, or mass spectrometry (Fletcher and Basketter, 2006). Our method could easily be extended to a multi-marker analysis by staining of serial sections each with another biomarker or the use of quantum dots thus giving a more comprehensive view on the mechanisms of the tissue response.

In our approach, whole slide scanners are the key to a spatial analysis as they allow the imaging of complete tissue sections in a highly automated way. The resulting images have a resolution of up to 0.23 μm/pixel and thus also facilitate subcellular analyses. We here deployed the nuclei staining to detect a translocation of Hsp27 to the nucleus. However, many other stainings e.g. for the membrane are possible thus allowing the expression analysis in individual cells.

An important question in our study was how standardized organotypic skin cultures actually are. Even though the good confidence intervals of the QSPs and the classification result suggest a highly reproducible tissue response we observed variations in staining intensity as well as epithelial thickness. A future way to compensate staining variations would be to develop further intensity-independent measures like texture and to increase the impact of intensity-independent measures during classification like the start of expression. Another possibility could be normalization of the measured marker intensity against a reference marker. We tried normalizing against the DAPI intensity inside a section, but the correlation of variations in DAPI intensity and marker intensity was low (data not shown). Thus, further effort will be to put on the establishment of a suitable reference marker. Variations in epithelial thickness are mostly compensated by our profiling based on relative distance. However, the penetration of a chemical strongly depends on the thickness of the stratum corneum, which could be captured by an additional staining.

Lastly, we would like to stress the enormous potential of data acquisition deploying tissue cultures. Such an approach not only complements data acquired from cell strains or lines but also provides a replacement for animal tests. Besides the important ethical issue, skin cultures also exhibit a higher degree of standardization than animal experiments due to absent inter-animal variations. Since about 1944, the irritative potential was predicted using rabbits (Draize *et al*., 1944) that finally showed to be over-predictive of human skin irritation (Robinson *et al*., 2002; Welss *et al*., 2004).

Concluding, we here presented an innovative approach for the standardized and automated quantification of skin response after perturbation mainly deploying tissue cultures and whole slide-scanning. At the example of irritation testing we showed that the tissue response is highly reproducible, thus allowing the classification based on spatial protein expression changes. Due to its high degree of standardization and automation this method now opens up new possibilities for the data acquisition for systems biological models of the skin. By combining molecular expression data with spatial information at the cell and further more organ level the presented way of data generation is highly suited for multi-scale systems biological projects aiming at the tissue level at multiple levels of biological abstraction (e.g. the USA 'virtual liver' project). Such multi-scale models are highly promising for the understanding of tissue perturbations thus facilitating the risk assessment of substances or drug development.

The approach presented here allows for the first time the controlled acquisition of quantitative and spatial data after perturbation of organotypic skin cultures and thereby forms the basis for quantitative computational models of higher complexity systems like the human skin.

ACKNOWLEDGEMENTS

We thank Annette Kohl for experimental assistance and Dana Sittner for providing reference MTT data. Furthermore, we would like to thank Helena Kandarova and Patrick Hayden from Mattek for advice regarding the experimental protocols.

Funding: Federal Institute of Risk Assessment (FK-1328-192); BMBF (FORSYS: 0315263 and MEDSYS: 0315401B).

Conflict of Interest: none declared.

REFERENCES

Andersen,M.E. and Krewski,D. (2009) Toxicity testing in the 21st century: bringing the vision to life. *Toxicol. Sci.*, **107**, 324–330.

Bell,E. *et al*. (1981) Living tissue formed in vitro and accepted as skin-equivalent tissue of full thickness. *Science*, **211**, 1052–1054.

Boxman,I. *et al*. (2002) Proteomic analysis of skin irritation reveals the induction of HSP27 by sodium lauryl sulphate in human skin. *Br. J. Dermatol*., **146**, 777–785.

Canny,J. (1986) A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **8**, 679–698.

Coquette,A. *et al*. (2003) Analysis of interleukin-1 (IL-1) and interleukin-8 (IL-8) expression and release in vitro reconstructed human epidermis for the prediction of in vivo skin irritation and/or sensitization. *Toxicol. In Vitro*, **17**, 311–321.

Draize,J. *et al*. (1944) Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J. Pharmacol. Exp. Ther*., **82**, 377–390.

Duda,R. *et al*. (2000) *Pattern Classification*, John Wiley & Sons, New York.

Fletcher,S.T. and Basketter,D.A. (2006) Proteomic analysis of the response of EpiDerm^{TM} cultures to sodium lauryl sulphate. *Toxicol. In Vitro,* **20,** 975–985.

Giuliano,K.A. *et al*. (1997) High-content screening: a new approach to easing key bottlenecks in the drug discovery process. *J. Biomol. Screen.*, **2**, 249–259.

Grabe,N. and Neuber,K. (2005) A multi-cellular systems biology model predicts epidermal morphology, kinetics, and Ca++ flow. *Bioinformatics*, **21**, 3541–3547.

Grabe,N. *et al*. (2007) Reconstructing protein networks of epithelial differentiation from histological sections. *Bioinformatics*, **23**, 3200–3208.

Ho,R.L. and Lieu,C.A. (2008) Systems biology—an evolving approach in drug discovery and development. *Drugs R D*, **9**, 203–216.

Kohavi,R. (1995) A Study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence (IJCAI)*. Morgan Kaufmann, San Francisco, CA, USA, pp. 1137–1143.

Krewski,D. *et al*. (2009) Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal.*, **29**, 474–479.

Meyer,F. (1994) Topographic distance and watershed lines. *Signal Processing*, **38**, 113–125.

Mosmann,T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, **65**, 55–63.

Otsu,N. (1979) A threshold selection method from grey-level histograms. *IEEE Trans. Syst. Man Cybern.*, **9**, 62–66.

Perry,A.D. and Trafeli,J.P. (2009) Hand dermatitis: review of etiology, diagnosis, and treatment. *J. Am. Board Fam. Med.*, **22,** 325–330.

Pommerencke,T. *et al*. (2007) *Ermittlung von räumlichen Proteinexpressionsmustern mittels Bildverarbeitung*. *Bildverarbeitung für die Medizin*, Springer, Berlin, Heidelberg.

Pommerencke,T. *et al.* (2008) Nuclear staining and relative distance for quantifying epidermal differentiation in biomarker expression profiling. *BMC Bioinformatics*, **9**, 473.

Pommerencke,T. *et al*. (2009) *Vollautomatische Einzelzellerkennung auf fluoreszenten Gewebeschnitten humaner Epidermis*. *Bildverarbeitung für die Medizin*, Springer, Berlin, Heidelberg.

Robinson,M.K. *et al*. (2002) Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. *Food Chem. Toxicol.*, **40**, 573–592.

Schnuch,A. (2007) Unwanted effects due to cosmetics—a review. *Allergologie*, **30**, 411–430.

Slodownik,D. *et al*. (2008) Irritant contact dermatitis: a review. *Australas. J. Dermatol.*, **49**, 1–11.

Thomas,N. (2010) High-content screening: a decade of evolution. *J. Biomol. Screen.*, **15**, 1–9.

Welss,T. *et al*. (2004) In vitro skin irritation: facts and future. State of the art review of mechanisms and models. *Toxicol. In Vitro*, **18**, 231–243.

Zweig,M. and Campbell,G. (1993) Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–77.