REGULAR ARTICLE

A perceptually optimised bivariate visualisation scheme for high-dimensional fold-change data

André Müller[1] · Ludwig Lausser[1] · Adalbert Wilhelm[2] · Timo Ropinski[3] ·
Matthias Platzer[4] · Heiko Neumann[5] · Hans A. Kestler[1,4]

Abstract

Visualising data as diagrams using visual attributes such as colour, shape, size, and orientation is challenging. In particular, large data sets demand graphical display as an essential step in the analysis. In order to achieve comprehension often different attributes need to be displayed simultaneously. In this work a comprehensible bivariate, perceptually optimised visualisation scheme for high-dimensional data is proposed and evaluated. It can be used to show fold changes together with confidence values within a single diagram. The visualisation scheme consists of two parts: a uniform, symmetric, two-sided colour scale and a patch grid representation. Evaluation of uniformity and symmetry of the two-sided colour scale was performed in comparison to a standard RGB scale by twenty-five observers. Furthermore, the readability of the generated map was validated and compared to a bivariate heat map scheme.

A. Müller, L. Lausser, H. Neumann, H.A. Kestler contributed equally.

H. Neumann, H.A. Kestler are joint senior authors.

✉ Hans A. Kestler
hans.kestler@uni-ulm.de

Heiko Neumann
heiko.neumann@uni-ulm.de

[1] Institute of Medical Systems Biology, Ulm University, 89069 Ulm, Germany

[2] Psychology and Methods, Jacobs University, 28759 Bremen, Germany

[3] Visual Computing Group, Ulm University, 89069 Ulm, Germany

[4] Leibniz Institute on Ageing, Fritz Lipmann Institut, Beutenbergstr. 11, 07745 Jena, Germany

[5] Institute of Neural Information Processing, Ulm University, 89069 Ulm, Germany

1 Introduction

Bioinformatics is a field with a broad range of applications and challenges (Augen 2005; Baldi and Brunak 2001). Based on the analysis of high-dimensional data (e.g. gene expression profiles etc.) it can lead to a deeper understanding of fundamental mechanisms of oncogenesis, tumour progression, and metastasis with the potential to generate new hypotheses for diagnosis and therapeutics (Kim et al. 2012). Besides a plethora of different statistical and heuristic approaches, the inspection of visual data maps is often conducted as an important and typically first step in the analysis of data of this kind.

In these investigations a common scenario is the comparison of gene expression profiles from two different populations such as carcinoma versus inflammation (Buchholz et al. 2005). In such an instance the two populations are characterised by the difference in their gene expression levels. A basic method of describing the change of expression is using their logarithmized ratio, commonly called fold change in this context. Bilban et al. (2002) indicated that gene expression ratios are no reliable markers alone. Fold change scores can be enriched with a measure of confidence (Fensterer et al. 2004). Hence, the ultimate goal of this paper is to present a bivariate visualisation tool that encompasses both the fold changes and their standard errors.

Traditionally gene expression fold changes are visualised as red and green patches for positive and negative fold changes (see Fig. 1a). The commonly used red-green colour palette is usually generated by uniformly sampling two channels of the standard RGB space (Eisen et al. 1998). The logarithmized fold changes are then linearly mapped onto the chosen colour palette.

The standard RGB space was designed as a colour mapping to control signals for technical devices such as computer monitors (Foley et al. 1997). It is rooted in the trichromatic colour theory (by Young and von Helmholtz (Young 1802)) which was originally based on the human ability to perceive any colour due to our photoreceptor sensitivity. The horseshoe like CIE xy chromaticity diagram, representing perceivable colours, does not show isotropic colour difference sensitivities, but depends on colour and location is this space (Lee 2005). Uniformly sampling such a colour space does therefore not result in perceptually equidistant stimuli in terms of just noticeable differences (JNDs) (Newman 1933), but rather a nonlinear mapping is necessary to achieve this. Certain values may become over- or underemphasised alone by choosing the wrong colour scale. As an example Fig. 2 shows two equally sampled red-green palettes with different perceptual properties: a standard RGB and a perceptually optimised scale (OPTIM). The commonly used bright tones of the standard RGB palette show a lower contrast to adjacent colours when compared to dark tones. This suggests the non-uniformness of the scale. Furthermore, the green half-scale appears much brighter than the red half-scale of the standard RGB palette indicating the non-symmetry of the scale.

The colour scale in Fig. 2b was computed by the *OPT-SCALE* algorithm (Kestler et al. 2006). Here, the corresponding colours of the red and the green half-scale reflect equal quantitative stimuli and neither of them shows a dominant effect. In each half-scale, the different colour tones are perceived as almost perfectly equidistant.

It was shown in Ware (Ware 2004, p. 136) that bi-variate colour scales for encoding two dimensions are difficult to read. Based on these observations we propose the use
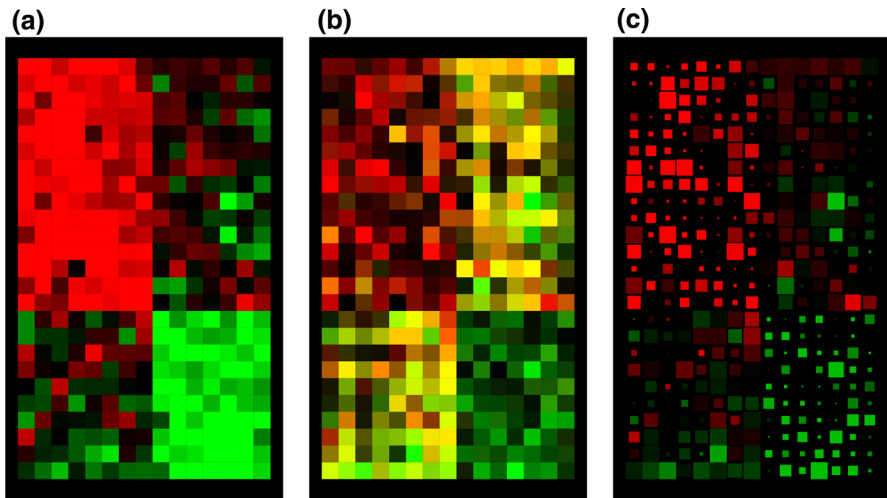
Fig. 1 The figure shows a random data set being rendered by three different maps: **a** the standard RGB representation showing only fold changes without confidences, **b** a bivariate heat map using the HSV colour space for encoding ratios and confidences into a single colour, and **c** the colour patch approach using an optimised red-green colour scale (encoding the fold changes) together with the modulation of patch sizes (confidence values). All three maps represent the same input data set. Rows and columns were ordered according to the ratios using a complete linkage hierarchical clustering algorithm applied to rows and columns independently. Distance metrics was the L_2 norm applied to the log ratios of columns and rows. Please also see the online supplement for colour reproductions of the figures (color figure online)
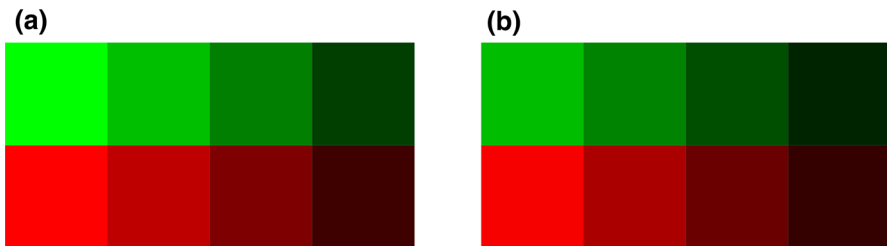
Fig. 2 Red-green scales showing 8 levels: **a** standard RGB: equidistant levels within the sRGB space, and **b** OPTIM: perceptually equidistant levels (color figure online)

of perceptually separable visual dimensions (Carswell and Wickens 1990): colour and size for encoding both values to achieve improved readability and the reduction of mapping artefacts in accordance with (Carswell and Wickens 1990).

2 Methods

2.1 Fold changes and their confidence measures

The definition of fold changes and confidence values depend on the specific application area. A fold change r is usually defined as a factor of change of a measurement value

between two conditions (e.g., values at different times of a process or measurements among different environmental conditions). It could be defined by gains or losses in stock exchange relative to a basic level A representing such as

$$r = \frac{B - A}{A}, \qquad A > 0. \tag{1}$$

An associated confidence measure a may originate from a statistical model or be the total magnitude of the measurements, e.g., the basic level $a = A$ or the mean measurements $a = (A + B)/2$. In the analysis of gene expression data fold changes are commonly defined as log ratios $M = \log_2 R/G$ and confidences as average log intensities $A = \frac{1}{2}log_2(R \cdot G)$ often displayed as MA plot (Dudoit et al. 2002) to shown intensity dependencies of singular fold changes. Here R and G can be intensity measurements within a microarray of a sample (e.g., tumour material) versus a control condition (healthy tissue) among multiple genes. Before rendering the dimensionality of data sets must be reduced. Feature selection methods try to omit attributes carrying no interesting information content such as those having no variation among all samples (Lausser et al. 2017; Müssel et al. 2016; Schirra et al. 2016; Völkel et al. 2015; Kraus et al. 2015; Lausser et al. 2013) or having low confidences over all samples (noise would dominate the signal in this case).

In the following we assume the existence of fold changes $\hat{r}_{ij} \in \mathbb{R}$ and associated confidences $\hat{a}_{ij} \in \mathbb{R}_0^+$ with features (such as, e.g., different genes) in the rows $i = 1 \ldots m$ and samples (e.g., tissue probes of different subjects or probes at different times) in the columns $j = 1 \ldots n$.

2.1.1 Dynamics reduction

In practice the dynamics of the raw fold changes $\hat{r}_{ij}$ and confidences $\hat{a}_{ij}$ needs to be limited and scaled to ranges $[-1, 1]$ and $[0, 1]$, respectively, before further rendering. Otherwise outliers could dominate the map. Hence the following thresholded scaling functions are proposed:

$$a_{ij} = \Psi_{0,1}\left(\frac{\hat{a}_{ij}}{\theta_a}\right) \tag{2}$$

$$r_{ij} = \Psi_{-1,1}\left(\frac{\hat{r}_{ij}}{\theta_r}\right) \tag{3}$$

with cutoff values $\theta_a > 0$ and $\theta_r > 0$ and hard thresholding function

$$\Psi_{a,b}(x) = \begin{cases} a & \text{for } x < a \\ x & \text{for } a \leq x \leq b \\ b & \text{for } x > b \end{cases} . \tag{4}$$

Soft thresholding functions (e.g., sigmoids) can be used instead of Eq (4) as well.

2.2 Optimal colour scales

We previously described the *OPT-SCALE* algorithm (Kestler et al. 2006) (an extension of the linear optimal scale algorithm of Levkowitz (Levkowitz 1997, pp.141)) which we briefly outline here. The procedure (see Alg 1) creates a bi-coloured palette

$$\mathbf{c} = \langle c_n^- \dots c_1^-, c_0, c_1^+ \dots c_n^+ \rangle \tag{5}$$

comprising n colours $\mathbf{c}^-$ representing negative log ratios (green tones), n colours $\mathbf{c}^+$ representing positive log ratios (red tones), and one central colour c_0 (usually neutral dark grey or black) representing a log ratio of 0 in order to create a symmetric palette.

Algorithm 1

procedure OPT- SCALE($\hat{\mathbf{c}}^+, \hat{\mathbf{c}}^-, n$)

    Calculate colour distance $\Delta = \min_{k \in \{+,-\}} d\left(\hat{\mathbf{c}}_1^k, \hat{\mathbf{c}}_m^k\right) / (n-1)$

    for $h \in \{+, -\}$ **do**

        Select a set $\mathbf{c}^h$ of n different colours from $\hat{\mathbf{c}}^h$ that

        minimizes $\left| d\left(c_{i-1}, \hat{c}_i\right) - \Delta \right| \, \forall \, i \in \{1, \dots, n-1\}$

        and contains c_1^h

    end for

    return $\langle \mathbf{c}^-, \mathbf{c}^+ \rangle$

end procedure

Hence a much larger perceptually ordered input palette (i.e. a rank can be assigned by the values)

$$\hat{\mathbf{c}} = \langle \hat{c}_m^- \dots \hat{c}_1^-, c_0, \hat{c}_1^+ \dots \hat{c}_m^+ \rangle, \qquad m \gg n \tag{6}$$

is sub-sampled, thereby conserving the ordering. The input palette $\hat{\mathbf{c}}$ is usually perceptually non-uniform. In the sample coding scheme shown here we used an 128-times oversampling i.e. $m = 128 \cdot n$ (we used $n = 64$). The colours for the sub-scale are chosen according to a perceptual distance measure d) such that all adjacent colours have approximately constant distance Δ:

$$d(c_{(i-1)}^h, c_i^h) \approx \Delta \qquad i = 2 \dots n \quad h \in \{+, -\} . \tag{7}$$

2.2.1 Perceptual distance measures

Distance measures (e.g. ℓ_2) between RGB coordinates do not correspond to perceptual differences between two stimuli. The Commission Internationale de l'Eclairage (CIE) proposed psychophysical derived colour spaces such as the CIE LUV and CIE LAB space in order to approximate a perceptually uniform colour space (CIE 2004). The benefit of these colour spaces is, that the ℓ_2 distance between two colour coordinates is nearly proportional to the perceptual stimulus difference i.e. for a difference within the (L^*, a^*, b^*) coordinate space: $\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$. Here L^*

is the lightness value of a colour. The pairs of values (a^*, b^*) $((u^*, v^*)$ for the LUV space) are the colour's coordinates of the corresponding colour space. If a colour scale $\hat{\mathbf{c}}$ is taken from one of these colour spaces, a subsampled scale $\mathbf{c}$ built by *OPT-SCALE* has the following properties:

- *U*niformity: Two adjacent colours c_{i-1}^h, c_i^h within same half-scale $h \in \{+, -\}$ should always have similar perceptual distances: i.e. $d(c_{i-1}^h, c_i^h) \approx \Delta \; \forall i \in \{2, \ldots, n\}$ with a constant $\Delta > 0$. Uniformity is necessary in order to map fold changes as linearly as possible to visual stimuli.
- *S*ymmetry: The absolute intensity of a stimulus originating from the negative half-scale c_i^- shall be similar to that of the corresponding positive stimulus c_i^+. According to the central colour c_0 in palette $\mathbf{c}$ the stimulus difference $d\left(c_0, c_i^+\right)$ and $d\left(c_0, c_i^-\right)$ should correspond in their magnitude, so that $\left|d\left(c_0, c_i^+\right)\right| \approx \left|d\left(c_0, c_i^-\right)\right| \; \forall i \in 1 \ldots n$.
  Symmetry is necessary in order to let negative log ratios appear with similar intensity as positive log ratios of the same magnitude.

2.3 Bivariate heat maps

One approach for visualising fold changes together with confidence values is the use of a bivariate colour scale mapping of both attributes onto a single colour tone (cf. Figs. 1b and 5). As baseline and as a means of comparison for the proposed visualisation scheme a bivariate heat map based on the hue-saturation-value (HSV) colour space (Gonzalez and Woods 2002) is used. Here the hue value H is modulated by the fold-changes (ranging from green to red - a traditional colour scale for gene expressions). The confidence is encoded as brightness V. Saturation S was kept constant at its maximum. More reliable entities are represented by brighter colours than those of lower confidence.

2.4 The patch grid visualisation approach

A further possibility for showing log ratios together with confidences is the patch grid visualisation approach which we describe here and which is experimentally compared to the bivariate heat map described in the last section (see Figs. 1c and 4). The main idea behind this approach is the use of two (nearly) perceptually independent visual channels (size and colour) in order to optimise the perceived information content.

In the patch grid visualisation approach each entity (fold change r_{ij} together with confidence a_{ij}) is represented by a square patch of size $s_{ij} \geq 0$ (edge length) representing the confidence filled by a colour c_{ij} encoding for the fold change. The squares (patches) are arranged as regular grid thereby showing the features (e.g., genes) in the rows and the samples in the columns. The background colour was chosen to be neutral dark grey. Spatial effects - i.e. the influence of adjacent stimuli - are not considered in this work as then the comparability to a common legend would have been lost (Robertson and O'Callaghan 1986, p. 30).

The patch colour c_{ij} is taken from the previously described OPT-SCALE colour palette by a lookup of r_{ij} and a linear interpolation of the colour coordinates. The patch sizes s_{ij} were chosen according to Weber's law (Weber 1905; Stuart et al. 1993) describing the threshold of perception of two physical stimuli I and $I + \Delta I$

$$k = \Delta I / I = const. \tag{8}$$

For a range of $[s_{min}, s_{max}] \subset \mathbb{R}^+$ a sequence of edge lengths s_l, $l \in \{1, \dots, n\}$ is constructed that fulfils $(s_l - s_{l-1})/s_l = k$ for all $l > 1$

$$s_l = s_{min} \cdot (1 + k)^{l-1} \quad \text{with} \quad k = \sqrt[n-1]{\frac{s_{max}}{s_{min}}}, \tag{9}$$

where $s_1 = s_{min}$ and $s_n = s_{max}$.

2.4.1 Ordering

For enhancing the perceptual continuity of the generated visual maps the order of rows and columns can be rearranged. For this a plethora of different methods is available such as the leaf ordering of hierarchical clusterings (e.g., single, average, complete linkage clusterings) (Jain and Dubes 1988) as well as the application of traveling salesman problem (TSP) rearrangements (Climer and Zhang 2006). In contrast to clustering algorithms the TSP optimises the sum of all adjacent distances thereby creating a slightly different arrangement.

2.5 Perceptual threshold detection

Techniques for detecting a perceptual threshold reach back to Fechner (1860), (Palmer 1999, Appendix A) and include methods of adjustment, limits, and constant stimuli. Here, the *best PEST* method - a constant stimulus method with adaptive step size - was used (Pentland 1980; Lieberman and Pentland 1982; Treutwein 1995). The algorithm tries to find an optimum perceptual threshold by maximum likelihood estimation of a sensitivity function Ψ giving a relationship between stimulus $X \in \mathbb{R}$ and a subject response $Z \in \{0, 1\}$:

$$\Pr(Z = 1 | X) = \Psi(X). \tag{10}$$

The logistic function

$$\Psi_\sigma(x) = \left(1 + e^{-x/\sigma}\right)^{-1} \tag{11}$$

was chosen with a fixed σ. The 50% decision point θ^* is sought where the probability of a positive or negative response is equal. The subject's response defines the random variable $Z \in \{0, 1\}$ to find the perceptual threshold denoted by

$$\Pr(Z | \theta^*) = 0.5. \tag{12}$$

The likelihood function is defined as

$$\mathcal{L}(\theta|x_1 \ldots x_n) = \Pi_{i=1}^{n} \mathcal{L}(\theta|x_i). \tag{13}$$

with stimulus level $x_i \in \mathbb{R}$ and subject response $z_i \in \{0, 1\}$ at trial $i = 1 \ldots n$. The likelihood of a single stimulus is $\mathcal{L}(\theta|x_i) = \Psi(x_i - \theta|z_i)$ with

$$\Psi(x|z) = \begin{cases} \Psi(x) & \text{if } z = 1 \\ 1 - \Psi(x) & \text{otherwise.} \end{cases} \tag{14}$$

The best PEST estimator shows at step n the stimulus being most likely the sought threshold value within range $[a, b]$:

$$x_{n+1} = \max_{\theta \in [a,b]} \Pi_{i=1}^{n} \Psi(x_i - \theta|z_i). \tag{15}$$

The first presented stimulus is $x_1 = (a+b)/2$ and should be far enough from the sought threshold θ. For numeric stabilisation Eq. (13) is logarithmized. As this transformation is strictly monotone the location of the maximum is not changed. For an implementation the interval $[a, b]$ is equally sampled at N points $a = u_1 < \cdots < u_N = b$ and a vector containing the current log likelihood function is generated (Lieberman and Pentland 1982).

3 Results

The following properties of the visualisation method were evaluated:

E1 *Uniformity* of the colour scale: the perceptual error is independent of the position within the scale.
E2 *Symmetry* of the colour scale: both parts of the bi-coloured scale are balanced.
E3 *Readability* of the visualisation scheme: visual entities can be matched with a legend at low error.

The experiments for properties E1 and E2 were performed for the standard RGB scale and the optimised red-green scale (OPTIM scale). The optimised scale was calculated by *OPT-SCALE* using CIE LAB coordinates. Property E3 is compared for the described bivariate HSV scale and the patch grid method.

3.1 Evaluation and viewing conditions

The visualisation scheme was evaluated by 25 knowledgeable potential users from Ulm University (students, (8), staff personnel (2), research assistants (15), female/male = 6/19, age: 22–39 years (female), 24–37 years (male)) with an educational background of mathematics/computer science (19) and medicine/biology (6). All had normal or corrected to normal visual abilities. None of the participants had a dyschromatopsia. A 20 inch monitor was used. The viewing distance was about 70 cm. Colours
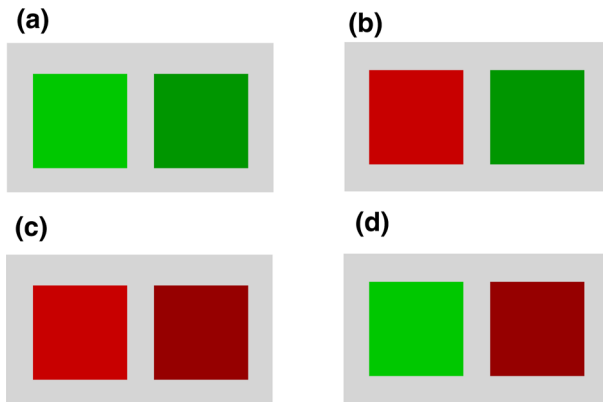
Fig. 3 Setup of the evaluations E1 (left pair of diagrams panels **a** and **c**) and E2 (right pair of diagrams, panels **b** and **d**). The target stimulus is always on the left side of each pair whereas the right stimulus is adaptable (+/−) via the keyboard (color figure online)

were equalised using a calibration device (Bits++, Cambridge Research Systems) in combination with a colorimeter (ColorCAL, Cambridge Research Systems).

3.2 Uniformity (E1) and symmetry (E2)

The described best PEST scheme was used. In these evaluations two square shaped stimuli are shown (see Fig. 3): one fixed reference (target) stimulus on the left side and one adaptable stimulus on the right side. The observers' task is to adjust the brightness of the right stimulus until it matches the brightness of the left stimulus. For E1 both stimuli were selected from the same side of the scale (green or red) whereas for E2 the stimuli were selected from opposite sides in order to match the brightness of a green colour tone with those of a red colour tone.

Six colour stimuli were equidistantly chosen from the interval [0.1, 0.7]. All stimulus values were scaled to [0, 1]. Each stimulus was shown twice, once for the RGB scale and once for the OPTIM scale (LAB space). At most 40 PEST adjustment steps had to be performed for each experiment. The 56 experiments (i.e. 14 colour tones × 2 colour scale × 2 experimental types - uniformity/symmetry) were randomly permuted for each evaluator. The target stimuli were chosen from the range [0.1, 0.7] for 6 identically spaced steps (it is assumed that the maximum stimulus range is defined by the interval [0, 1]). In each experiment a stimulus was shown twice, once for the RGB scale and once for the OPTIM scale. An observer had to do 28 adjustments for each experiment. All 56 single evaluations were shown in sequence, which was randomly permuted for each observer.

3.2.1 Initialization

The best PEST procedure is initialised such that the first presented stimulus x_1 is far enough from the target stimulus x^* to be always detected as different. As the initial

Table 1 Best PEST parameters for experiments 1 and 2

Parameter	Value	Description
N	1000	Resolution of the probability vector
n_{trials}	40	Number of adaptations
Δ	2	Range of scale
σ	1/5	Reciprocal steepness of the psychophysical function

Table 2 Results of the uniformity test E1 (RGB)

$Q_{\Delta x}(z-x)$	Stimulus x														
	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	Σ
≥ 2	0	0	0	0	0	1	1	0	0	0	0	1	2	0	5
1	4	2	4	1	0	0	0	0	2	6	3	2	7	1	32
0	**14**	**17**	**17**	**19**	**22**	**23**	**24**	**23**	**21**	**18**	**19**	**21**	**15**	**19**	**272**
−1	5	6	4	5	3	1	0	0	0	0	2	1	1	5	33
≤ -2	2	0	0	0	0	0	0	2	2	1	1	0	0	0	8
Accuracy	0.56	0.68	0.68	0.76	0.88	0.92	0.96	0.92	0.84	0.72	0.76	0.84	0.60	0.76	**0.76**

The count of observations falling in the given quantisation group are shown. On the bottom right the median accuracy is shown. Correct responses are shown in bold

stimulus for the best PEST procedure is in the center $m = a + \Delta/2$ of the defined range $[a, a+\Delta]$, a was adapted such that $m+2\sigma \leq x^* \leq m+4\sigma$ or $m-4\sigma \leq x^* \leq m-2\sigma$ uniformly distributed with equal probability. The used parameters for E1 and E2 are summarised in Table 1.

For analysing the results it is of interest if a subject categorised a colour square x erroneously to a response $z \gg x$ or $z \ll x$. To be more rigorous we define a quantisation function

$$Q_{\Delta x}(u) = \left\lfloor \frac{u}{\Delta x} + \frac{1}{2} \right\rfloor \tag{16}$$

mapping response failures $Q_{\Delta x}(z-x)$ to an integer k. The quantisation resolution matches the stimuli distance $\Delta x = 0.1$. So $k = 0$ means that the colour patch was correctly classified, $k = +1$ or $k = -1$ means that the adjacent bin was erroneously found.

Tables 2 (RGB) and 3 (OPTIM) categorise the response errors into the five classes $\leq$-2,-1, 0, 1, ≥ 2 for each stimuli s. There are the correct responses $Q = 0$ and the responses of the two adjacent stimuli $Q = -1$ or $Q = 1$, and those which are more than two stimuli steps away: $Q \leq -2$ or $Q \geq 2$. For both scales the median number of correct matches is very similar 76% (RGB) versus 84% (OPTIM). In stimuli with higher magnitude $|s|$, more confusions between s and its neighbours can be seen for the RGB scale than for the OPTIM scale. The improvement was not significant (p-value 0.1722) using a one-sided paired Wilcoxon rank sum test.

Table 3 Results of the uniformity test E1 (OPTIM)

$Q_{\Delta x}(z-x)$	Stimulus x														
	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	Σ
≥ 2	1	1	1	0	0	3	0	0	2	0	0	0	0	1	9
1	2	4	0	2	5	2	3	4	1	1	1	2	2	1	30
0	**21**	**20**	**23**	**23**	**18**	**17**	**21**	**16**	**22**	**20**	**24**	**21**	**22**	**23**	**291**
−1	1	0	1	0	2	3	1	2	0	2	0	2	1	0	15
≤ −2	0	0	0	0	0	0	0	3	0	2	0	0	0	0	5
Accuracy	0.84	0.80	0.92	0.92	0.72	0.68	0.84	0.64	0.88	0.80	0.96	0.84	0.88	0.92	**0.84**

The count of observations falling in the given quantisation group are shown. On the bottom right the median accuracy is shown. Correct responses are shown in bold

Table 4 Results of the symmetry test E2 (RGB)

$Q_{\Delta x}(z-x)$	Stimulus x														
	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	Σ
≥ 2	1	2	1	0	1	1	0	1	1	0	0	0	1	1	10
1	0	0	0	0	1	0	0	4	1	0	0	0	1	0	7
0	**0**	**0**	**0**	**4**	**2**	**1**	**1**	**11**	**2**	**3**	**0**	**0**	**2**	**0**	**26**
−1	0	0	2	1	5	2	6	9	17	5	2	3	0	4	56
≤ −2	24	23	22	20	16	21	18	0	4	17	23	22	21	20	251
Accuracy	0.00	0.00	0.00	0.16	0.08	0.04	0.04	0.44	0.08	0.12	0.00	0.00	0.08	0.00	**0.04**

The count of observations falling in the given quantisation group are shown. On the bottom right the median accuracy is shown. Correct responses are shown in bold

Tables 4 and 5 show the number of correct matches and the adjacent matches for the RGB and OPTIM scales. For the RGB scale the responses tend to be higher than the original stimulus. 26 examples (of 350) have been classified correctly. Only few responses are classified as neighbouring stimuli (63). For the optimised scale the responses are more centred. More stimuli have been categorised correctly (82). Of the misclassified examples 126 are categorised as adjacent stimuli. Comparing the accuracies with a one-sided paired Wilcoxon rank sum test showed a significant better accuracy of the OPTIM scale with p-value 0.001652.

3.3 Readability (E3)

The readability of the patch grid visualisation scheme was investigated by presenting a random patch grid with a marked patch. Colour and size of the patch had to be matched independently to a legend at the side (see Fig. 4).

In the HSV colouring scheme the user had to match the selected value with a bivariate colour scale (see Fig. 5).

A set of 12 colour tones and 8 different patch sizes was used. The colours were chosen from a red-green palette by using the *OPT-SCALE* algorithm within the LAB

Table 5 Results of the symmetry test E2 (OPTIM)

| $Q_{\Delta x}(z-x)$	Stimulus x														
	−0.7	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	Σ
≥ 2	8	8	3	5	3	2	3	0	4	7	5	9	4	4	65
1	4	5	2	2	0	2	1	1	3	2	7	1	3	4	37
0	**8**	**7**	**10**	**4**	**6**	**5**	**3**	**10**	**8**	**2**	**4**	**3**	**6**	**6**	**82**
−1	4	4	7	8	8	2	8	10	8	10	5	3	4	8	89
≤ −2	1	1	3	6	8	14	10	4	2	4	4	9	8	3	77
Accuracy	0.32	0.28	0.40	0.16	0.24	0.20	0.12	0.40	0.32	0.08	0.16	0.12	0.24	0.24	**0.24**

The count of observations falling in the given quantisation group are shown. On the bottom right the median accuracy is shown. Correct responses are shown in bold
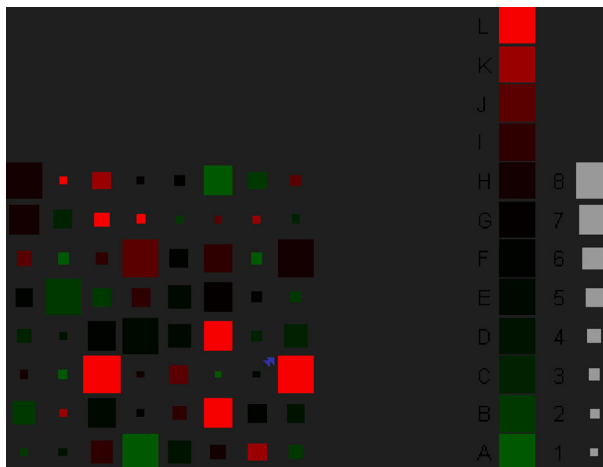
Fig. 4 Setup of experiments E3 for validating the readability of a patch grid. The evaluator had to match colour and size of a selected patch on the left side with the legends on the right side (color figure online)

space. The resulting colour scale contained 6 red and 6 green colours. The colour tones were equidistantly chosen within the range [0.05, 1.0] on both sides of the bi-coloured scale. For each visualisation schemes 30 combinations were randomly drawn without replacement from all possible 96 ratio/confidence combinations (independently for each subject). The edge lengths (always quadratic patches) were subsampled from the interval [0.2, 1.0]. A sequence of $n = 8$ edge lengths was constructed (Eq. 9).

The objective of the readability experiment was the matching of colour and patch size on two legends (one for the colour tone and one for the patch size) for the patch grid visualisation scheme. The bivariate colour map (HSV) had a two dimensional legend showing ordered colour tones in the rows (ranging from green to red) and varying brightness values among the columns.

Tables 6, 7, 8, and 9 summarise the results from the readability experiment as cooccurrence matrices. The median accuracy over all shown fold changes was 0.91 for the patch grid and 0.45 for the bivariate HSV colour map. A one-sided paired Wilcoxon
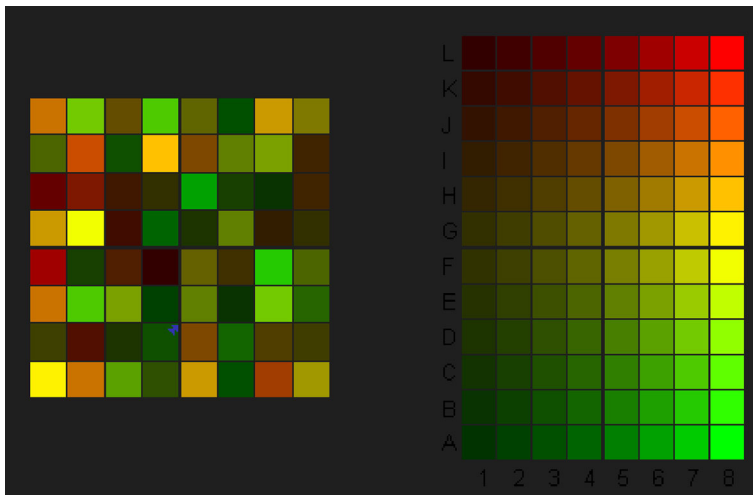
Fig. 5 Setup of experiment E3 for validating the readability of the HSV bivariate colouring scheme. Each user had to match a marked colour square on the left side of the diagram with an entry on the right side of the palette (color figure online)

rank sum test reveals a significant better accuracy for the patch grid system (p-value 0.002669). The confidence values could be read with median accuracy 0.905 in the patch grid approach and 0.37 in the bivariate HSV colour map (p-value 0.007813).

4 Discussion

Heatmap-representations are a standard visualisation concept for a large variety of high dimensional molecular high-throughput profiles. Their correct interpretation can improve the analysis of mRNA or miRNA profiles as well as the analysis of methylation or mutation patterns (Gress et al. 2017; Ogechukwu et al. 2017; Taudien et al. 2016; Gress et al. 2011). However, heatmaps also provide an overwhelming amount of information. Ambiguous representations can blur the perception of existing information and can therefore be misleading (Fig. 1). In this work we evaluate a novel bivariate visualisation scheme which is mainly applied to gene expression data, but not limited to this application, showing fold changes and confidence values within a single diagram.

In a first step the used OPTIM colour scale (Kestler et al. 2006) was investigated by comparing its uniformity (E1) and symmetry (E2) property to a standard RGB scale commonly used in the rendering of gene expression data. Here *uniformity* is the property of perceptual equidistance within a colour scale and *symmetry* is the property of balancedness between two half-scales. As a consequence reddish and green colour tones are matched by their brightness such that the encoded value is equally well perceived independent of its sign. These are necessary prerequisites for a bicolour visualisation that conveys meaning in the display of gene expression fold changes. The uniformity experiments (E1) showed a slightly higher accuracy among all stimuli for

Table 6 E3 using the OPTIM patch grid visualisation scheme: Stimuli and responses of the fold changes encoded as colour hue from the OPTIM colour space are shown as a cooccurrence matrix

Response z	Stimulus x												
	−6	−5	−4	−3	−2	−1	1	2	3	4	5	6	
−6	**59**	2	0	0	0	0	0	0	0	0	0	0	
−5	3	**66**	6	0	0	0	0	0	0	0	0	0	
−4	0	1	**55**	6	0	0	0	0	0	0	0	0	
−3	0	0	9	**61**	2	0	0	0	0	0	0	0	
−2	0	0	0	7	**56**	5	0	0	0	0	0	0	
−1	0	1	0	0	5	**59**	2	0	0	0	0	0	
1	0	0	0	0	0	2	**61**	6	0	0	0	0	
2	0	0	0	0	0	0	0	**60**	4	0	0	0	
3	0	0	0	0	0	0	0	1	**60**	0	0	0	
4	0	0	0	0	0	0	0	0	1	**69**	9	0	
5	0	0	0	0	0	0	0	0	0	2	**64**	1	
6	0	0	0	0	0	0	1	0	0	0	4	**66**	**Median**
Accuracy	0.95	0.94	0.79	0.82	0.89	0.89	0.95	0.90	0.92	0.97	0.83	0.99	**0.91**

The columns represents the colour stimulus and the rows the corresponding answers of the evaluators. The median total accuracy over all stimuli and its standard deviation are shown in the bottom right field. Correct responses are shown in bold

Table 7 E3 using a HSV bicolour scale: Stimuli and responses of the fold changes encoded as colour hue from the HSV colour space are shown as a cooccurrence matrix

Response z	Stimulus x												
	−6	−5	−4	−3	−2	−1	1	2	3	4	5	6	
−6	**45**	21	4	0	0	0	0	0	0	0	0	0	
−5	14	**29**	9	0	0	0	0	0	0	0	0	0	
−4	2	20	**25**	16	1	0	0	0	0	0	0	0	
−3	0	0	28	**25**	4	1	0	0	0	0	0	0	
−2	0	0	3	28	**22**	1	0	0	0	0	0	0	
−1	0	0	0	4	26	**18**	12	1	0	0	0	0	
1	0	0	0	1	7	25	**27**	3	0	0	0	0	
2	0	0	0	0	2	14	19	**32**	1	0	0	0	
3	0	0	1	0	1	5	5	29	**42**	0	0	0	
4	0	0	0	0	0	1	0	1	21	**45**	2	0	
5	0	0	0	0	0	0	1	1	1	26	**73**	0	
6	1	0	0	0	0	1	0	0	0	0	2	**67**	**Median**
Accuracy	0.73	0.41	0.36	0.34	0.35	0.27	0.42	0.48	0.65	0.63	0.95	1.00	**0.45**

The columns represents the colour hue stimulus and the rows the corresponding answers of the evaluators. The median total accuracy over all stimuli and its standard deviation are shown in the bottom right field. Correct responses are shown in bold

Table 8 E3 using the patch grid visualisation scheme: The cooccurrence matrix of confidence stimuli versus responses coded as rectangle sizes is shown

Response z	Stimulus x								
	1	2	3	4	5	6	7	8	
1	**98**	5	0	0	0	0	0	0	
2	15	**95**	2	0	0	0	0	0	
3	0	9	**88**	8	1	1	0	0	
4	0	1	4	**87**	8	0	0	0	
5	0	0	0	1	**83**	8	0	0	
6	0	0	0	1	7	**105**	7	0	
7	0	0	0	0	0	1	**87**	1	
8	0	1	0	0	1	0	0	**91**	**Median**
Accuracy	0.87	0.86	0.94	0.90	0.83	0.91	0.93	0.99	**0.905**

The median total accuracy over all stimuli and its standard deviation are shown in the bottom right field. Correct responses are shown in bold

Table 9 E3 using the HSV visualisation scheme: The cooccurrence matrix of confidence stimuli versus responses coded as HSV brightness value is shown

Response z	Stimulus x								
	1	2	3	4	5	6	7	8	
1	**103**	84	30	9	0	0	1	1	
2	7	**24**	47	20	3	1	0	0	
3	2	3	**15**	40	19	0	1	0	
4	0	0	2	**26**	35	6	0	0	
5	0	0	0	2	**32**	30	1	0	
6	0	0	0	0	7	**48**	10	1	
7	1	0	0	0	4	27	**42**	7	
8	0	0	0	0	0	3	39	**83**	**Median**
Accuracy	0.91	0.22	0.16	0.27	0.32	0.42	0.45	0.90	**0.37**

The median total accuracy over all stimuli and its standard deviation are shown in the bottom right field. Correct responses are shown in bold

the OPTIM scale (median accuracy: 0.84) compared to the standard RGB scale (mean accuracy: 0.76). The symmetry experiments (E2) - the matching of the brightness of reddish and green colour tones - revealed a strong imbalance within the RGB scale; a red stimulus was matched with a considerable darker green (mean accuracy: 0.04). The OPTIM scale showed a better balance and much more stimuli were categorised correctly (mean accuracy: 0.24) and is therefore less likely to overemphasise positive or negative fold changes of expression levels.

For an increasing magnitude the confusion of neighbouring stimuli increased for the standard RGB scale. Here, the perceptual differences between single colours decrease, which makes it harder to distinguish single stimuli. This effect was not observable in the OPTIM scale experiments where colours were chosen perceptually equidistant. For

small magnitudes the perceptual differences in the RGB scale are disproportionately high. Here, more adjustments were registered for the RGB scale than for the OPTIM scale. Considering the complete scale the confusion with adjacent stimuli is more uniform for the OPTIM scale compared to the standard RGB colour scale. Overall the OPTIM scale seems to be more appropriate for visualising quantitative information than the standard RGB scale. Especially in a symmetry task the OPTIM scale is much more useful than the RGB scale. The OPTIM scale therefore allows an improved discrimination of different fold changes.

In a second step the readability (E3) of the proposed patch visualisation scheme compared to a bivariate colour coding (using the HSV colour space) was evaluated. Here, the ability of each observer for directly matching both encoded attributes as a single visual entity with a legend was evaluated. In the patch grid approach the readability showed a much higher accuracies on fold changes (median accuracy: 0.91) and confidences (median accuracy 0.905) compared to the HSV visualisation: here the fold changes (coded as colour hue) were categorised with a median accuracy of 0.45 and the confidences (coded as brightness value) of only 0.37.

The improved interpretability of the proposed patch grid visualisation approach compared to a simple bivariate colour scale was demonstrated. The focus was here to add a confidence value to a standard visualisation approach without disturbing the readability of the so generated map by introducing a second attribute. Showing confidences together with fold changes is important in rendering gene expressions but may be of interest in other application areas as well, such application domains include analysing economic growth rates or assessment of investment strategies; here, the observed dimensions may include time and geographical region, business area, or business strategy. Another example could be the analysis of the agricultural output of different crops among different environmental conditions (soil, fertiliser, temperature, humidity etc.).

Compliance with ethical standards

References

Augen J (2005) Bioinformatics in the post-genomic era. Addison-Wesley, Boston

Baldi P, Brunak S (2001) Bioinformatics: The machine learning approach. MIT Press, Cambridge

Bilban M, Buehler L, Head S, Desoye G, Quaranta V (2002) Defining signal thresholds in DNA microarrays: exemplary application for invasive cancer. BMC Genom 3(1):1

Buchholz M, Kestler HA, Bauer A, Böck W, Rau B, Leder G, Kratzer W, Bommer M, Scarpa A, Schilling MK et al (2005) Specialized DNA arrays for the differentiation of pancreatic tumors. Clin Cancer Res 11(22):8048–8054

Carswell CM, Wickens CD (1990) The perceptual interaction of graphical attributes: configurality, stimulus homogeneity, and object integration. Percept Psychophys 47(2):157–168. https://doi.org/10.3758/BF03205980

CIE (2004) Colorimetry, 3rd edition. Tech. Rep. 15, Commission Internationale de l'Eclairage, http://www.cie.co.at

Climer S, Zhang W (2006) Rearrangement clustering: pitfalls, remedies, and applications. J Mach Learn Res 7:919–943

Dudoit S, Yang YH, Callow MJ, Speed TP (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Statistica Sinica 12(1):111–139 http://www.jstor.org/stable/24307038

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. PNAS 95(25):14863–14868

Fechner GT (1860) Elemente der Psychophysik, 2nd edn. Breitkopf and Härtel, Leipzig, Germany

Fensterer H, Giehl K, Buchholz M, Ellenrieder V, Buck A, Kestler HA, Adler G, Gierschik P, Gress TM (2004) Expression profiling of the influence of RAS mutants on the TGFB1-induced phenotype of the pancreatic cancer cell line PANC-1. Genes Chromosom Cancer 39(3):224–235. https://doi.org/10.1002/gcc.20000

Foley J, van Dam A, Feiner S, Hughes J (1997) Computer graphics-principles and practice in C, 2nd edn. Addison-Wesley, Boston

Gonzalez RC, Woods RE (2002) Digital image processing, 2nd edn. Prentice Hall, New Jersey

Gress TM, Kestler HA, Lausser L, Fiedler L, Sipos B, Michalski CW, Werner J, Giese N, Scarpa A, Buchholz M (2011) Differentiation of multiple types of pancreatico-biliary tumors by molecular analysis of clinical specimens. J Mol Med 90(4):457–464

Gress TM, Lausser L, Schirra LR, Ortmller L, Diels R, Kong B, Michalski CW, Hackert T, Strobel O, Giese NA, Schenk M, Lawlor RT, Scarpa A, Kestler HA, Buchholz M (2017) Combined microRNA and mRNA microfluidic TaqMan array cards for the diagnosis of malignancy of multiple types of pancreatico-biliary tumors in fine-needle aspiration material. Oncotarget 8(64):108223–108237

Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice-Hall, Inc

Kestler HA, Müller A, Buchholz M, Gress TM, Palm G (2006) A perceptually optimized scheme for visualizing gene expression ratios with confidence values. In: André E, Dybkjær L, Minker W, Neumann H, Weber M (eds) Perception and interactive technologies, LNAI, vol 4021. Springer, Berlin, pp 73–84

Kim KI, Tk Kim, Kim IW, Ahn KS, Yoon SS, Shin WG, Oh JM (2012) Copy number variations in normal karyotype acute myeloid leukaemia and their association with treatment response. Basic Clin Pharmacol Toxicol 111(5):317–324

Kraus JM, Lausser L, Kestler HA (2015) Exhaustive k-nearest-neighbour subspace clustering. J Stat Comput Simul 85(1):30–46

Lausser L, Müssel C, Maucher M, Kestler HA (2013) Measuring and visualizing the stability of biomarker selection techniques. Comput Statistics 28(1):51–65

Lausser L, Szekely R, Schirra LR, Kestler HA (2017) The Influence of Multi-class Feature Selection on the Prediction of Diagnostic Phenotypes. Neural Processing Letters pp 1–18

Lee HC (2005) Introduction to Color Imaging Science. Cambridge University Press, Cambridge

Levkowitz H (1997) Color theory and modeling for computer graphics, visualization, and multimedia applications. Kluwer

Lieberman H, Pentland A (1982) Microcomputer-based estimation of psychophysical thresholds: The best pest. Behav Res Methods Instrum 14(1):21–25

Müssel C, Schmid F, Blätte TJ, Hopfensitz M, Lausser L, Kestler HA (2016) BiTrinA—multiscale binarization and trinarization with quality analysis. Bioinformatics 32(3):465–468

Newman EB (1933) The validity of the just noticeable difference as a unit of psychological magnitude. Trans Kansas Acad Sci 36:172–175

Ogechukwu B, Lausser L, Huse K, Bergmeier C, Jahn N, Groth M, Bens M, Sahm A, Gall M, Witte OW, Kestler HA, Schwab M, Platzer M (2017) Tissue-, sex-, and age-specific DNA methylation of rat glucocorticoid receptor gene promoter and insulin-like growth factor 2 imprinting control region. Physiol Genomics 49(11):690–702

Palmer SE (1999) Vision science: photons to phenomenology. MIT Press, Cambridge

Pentland A (1980) Maximum likelihood estimation: the best pest. Percept Psychophys 28(4):377–379

Core Team R (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Robertson PK, O'Callaghan JF (1986) The generation of color sequences for univariate and bivariate mapping. IEEE Comput Graphics Appl 6(2):24–32

Schirra LR, Lausser L, Kestler HA (2016) Selection stability as a means of biomarker discovery in classification. Analysis of large and complex data. Springer, Berlin, pp 79–89

Stuart GW, Bossomaier TRJ, Johnson S (1993) Preattentive processing of object size: implications for theories of size perception. Perception 22:1175–1193

Taudien S, Lausser L, Giamarellos-Bourboulis EJ, Sponholz C, Felder FS, Schirra LR, Schmid F, Gogos CSG, Petersen BS, Franke A, Lieb W, Huse K, Zipfel PF, Kurzai O, Moepps B, Gierschik P, Bauer M, Scherag A, Kestler HA, Platzer M (2016) Genetic factors of the disease course after sepsis: rare deleterious variants are predictive. EBioMedicine 12:227–238

Treutwein B (1995) Adaptive psychophysical procedures. Vision Res 35(17):2503–2522

Völkel G, Lausser L, Schmid F, Kraus JM, Kestler HA (2015) Sputnik: ad hoc distributed computation. Bioinformatics 31(8):1298

Ware C (2004) Information Visualization, 2nd edn. Morgan Kaufmann, Burlington

Weber EH (1905) Der Tastsinn und das Gemeingefuehl [1846] hg, vol Ewald. Hering, Leipzig

Young T (1802) The bakerian lecture: on the theory of light and colours. Philos Trans R Soc Lond 92:12–48. https://doi.org/10.1098/rstl.1802.0004