nature biotechnology

Analyzing proteome topology and function by automated multidimensional fluorescence microscopy

Walter Schubert[1,5], Bernd Bonnekoh[2], Ansgar J Pommer[5], Lars Philipsen[5], Raik Böckelmann[2], Yanina Malykh[5], Harald Gollnick[2], Manuela Friedenberger[1,6], Marcus Bode[1,5,6] & Andreas W M Dress[3,4,6]

Temporal and spatial regulation of proteins contributes to function. We describe a multidimensional microscopic robot technology for high-throughput protein colocalization studies that runs cycles of fluorescence tagging, imaging and bleaching *in situ*. This technology combines three advances: a fluorescence technique capable of mapping hundreds of different proteins in one tissue section or cell sample; a method selecting the most prominent combinatorial molecular patterns by representing the data as binary vectors; and a system for imaging the distribution of these protein clusters in a so-called toponome map. By analyzing many cell and tissue types, we show that this approach reveals rules of hierarchical protein network organization, in which the frequency distribution of different protein clusters obeys Zipf's law, and state-specific lead proteins appear to control protein network topology and function. The technology may facilitate the development of diagnostics and targeted therapies.

Establishing how networks of proteins associate in time and space to generate function is an important goal of post-genomic research[1]. Formation of interaction networks within a cell requires proteins to be at the right place, at the right time and concentration. Molecular networks, which enable specific cellular functions, obey a unique protein colocation and anti-colocation code[2].

Fluorescent proteins have been used to examine close molecular colocations/interactions using fluorescence resonance energy transfer (FRET) microscopy[3–5], or the location of the cellular proteome in intact yeast[6]. Direct fluorescence detection of up to 17 different proteins is now possible[7]. Nevertheless, none of these methods can image sufficient numbers of proteins in a single sample to map protein network features. Scanning mass spectrometry has been explored to establish protein networks[8,9], but its spatial resolution remains far inferior to that of fluorescence microscopy.

We have previously described a technique termed multi-epitope-ligand kartographie (MELK), henceforth referred to as multi-epitope-ligand cartography (MELC), that can map the location of several proteins in the one sample of cells or tissue using sequential rounds of fluorescent detection *in situ*[2]. Here we demonstrate that MELC can colocalize at least a hundred molecules in a single sample and we describe an approach in which proteins detected in each data point are represented as a multidimensional vector. The summed vectors from an individual cell or tissue can be represented by what we term a toponome map (from the Greek τόπος, location, and νόμος, a rule or law), revealing hierarchical features of protein organization. After demonstrating the method's validity and robustness, we show that the frequencies of protein clusters obey Zipf's law, a power law for highly organized systems, and that MELC can readily be used to identify *in situ* the lead[2] or hub proteins[10,11] responsible for organizing protein networks. We also show that MELC can identify new diagnostic features and therapeutic targets in diseases.

RESULTS

Multidimensional analysis of protein locations

MELC builds on fluorescence imaging by bleaching each dye completely after labeling and imaging, and then applying another set of labeled antibodies (or other affinity reagents) to image the distribution of additional molecules. By repeating these steps, large sets of molecular distribution maps can be acquired with light-microscopic resolution, thereby overcoming the limitations of traditional fluorescence microscopy (**Supplementary Fig. 1a** online).

To test the MELC technique, we mapped 18 cell-surface receptors in peripheral blood mononuclear cells (PBMCs) using monoclonal antibody (mAb) library 1 (**Supplementary Table 1** online) conjugated to fluorescein isothiocyanate (FITC) or phycoerythrin (PE). PBMCs were isolated and fixed on cover slips, incubated with FITC- or PE-conjugated mAbs (**Fig. 1a**; CD4; CD8) and washed. After storing the resulting images, we bleached the remaining fluorescent signals by soft multiwavelength excitation centered on 485 nm for FITC and 546 nm for PE (to avoid damaging proteins by energy transfer), and stored a post-bleaching image (data not shown). We repeated this cycle nine times with one mAb or two mAbs per cycle. Protein distributions varied strikingly from cell to cell (**Fig. 1b**), reflecting the heterogeneity of protein localization in PBMCs from healthy donors.

[1]Molecular Pattern Recognition Research (MPRR) Group, Institute of Medical Neurobiology, Otto-von-Guericke-University Magdeburg, D-39120 Magdeburg, Germany. [2]Clinic for Dermatology and Venereology, Otto-von-Guericke-University Magdeburg, D-39120 Magdeburg, Germany. [3]CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences (SIBS), CN-200031 Shanghai, China. [4]MPI for Mathematics in the Sciences, D-04103 Leipzig, Germany. [5]ZENIT Technology Park, D-39120 Magdeburg, Germany. [6]These authors contributed equally to this work. Correspondence and requests for additional data should be addressed to W.S. (walter.schubert@medizin.uni-magdeburg.de).

Received 1 February; accepted 16 August; published online 1 October 2006; doi:10.1038/nbt1250

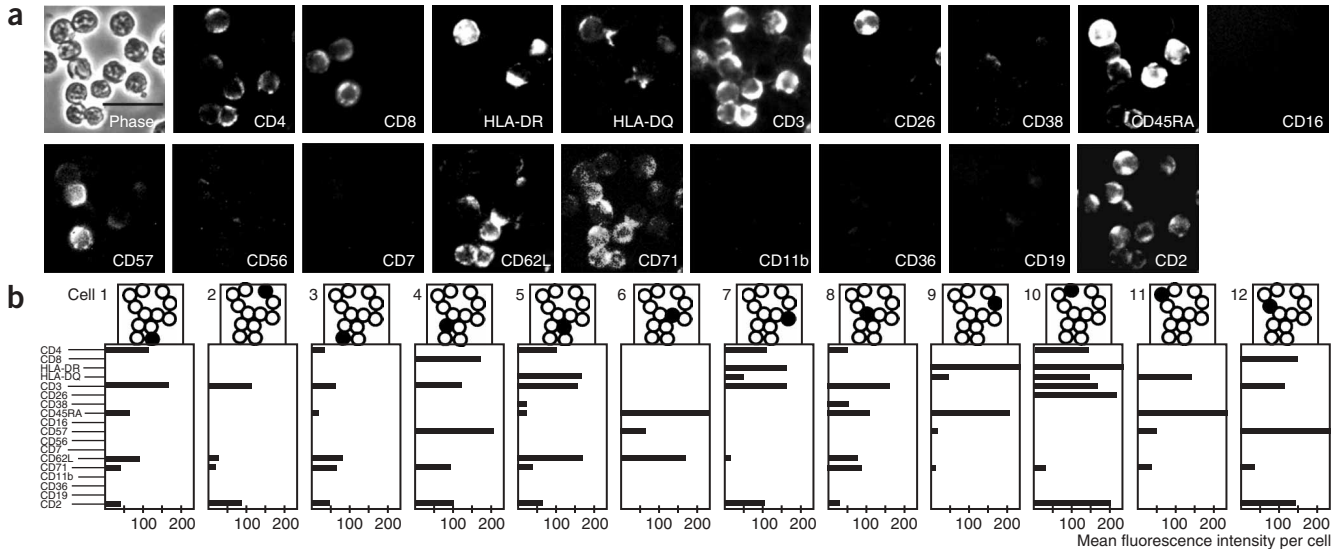

Figure 1 Protein mapping by MELC. (**a**) Phase contrast view of PBMCs from a healthy donor, followed by MELC maps of 18 cell surface proteins (subset of library 1) in the same cells. Although some proteins are undetectable, they were readily detected in other samples (data not shown). (**b**) Mean protein levels in each cell, showing striking heterogeneity among the 12 cells analyzed. Scale bar, 20 μm.

To obtain a protein colocalization map, we visualized MELC data by superimposing images (**Supplementary Video 1** online). However, superimposition provides no direct estimate of which protein combinations are formed *in vivo*. To address this issue, we represented MELC data as vectors. Because each MELC map represents the distribution of a single protein epitope, superimposition of these maps provides a list of the epitopes present at each pixel. This list can be expressed either as a vector of fluorescence intensities or, after introducing appropriate threshold values for each epitope, as a 0.1-vector denoting the presence or absence of every protein at each pixel. Thus each primary data set (the fluorescence intensity images) produces a collection of 0.1-vectors (binarized data sets) termed combinatorial molecular phenotypes (CMPs; **Fig. 2a–e**). The rationale for binarization is fourfold: (i) although a continuum of concentrations may exist for each protein in a sample, MELC data sets based on related intensity levels are extremely complex and difficult to handle when large data sets have to be compared, (ii) binarized protein patterns reflect the most prominent features of protein localization (**Fig. 2f,g**), (iii) binarized data sets can rapidly be translated into binary CMPs for immediate quantification and (iv) binary CMPs can be easily visualized and readily explored as a toponome map of any cell or tissue (**Fig. 2d,e**), and tested for robustness (see below).

Our analysis indicated that some CMPs (or slight variants, as defined in Methods, Construction of 2D toponome maps) denote functional regions of the cell or tissue in question. These features were defined as CMP motifs: sets of CMPs which (i) invariably contain one or more proteins (lead protein(s)), (ii) never contain one or more proteins (absent proteins) and (iii) may or may not contain additional proteins (wild-card proteins). Each CMP motif can thus be denoted by a sequence of 1s, 0s and *s, indicating the lead, absent and wild-card proteins, respectively. By assigning one color to each CMP motif, we can subsequently produce cell and tissue maps, summarizing protein organization in what we term a two-dimensional (2D) or three-dimensional (3D) toponome map (**Fig. 2d,e**, respectively, and **Supplementary Video 1** online).

To test MELC for signal separation, we extracted nuclear and cytoplasmic markers of a human hepatocyte (**Fig. 2d**) as positive and negative controls in the imaging procedure. Golgi/ER structures, cytokeratin and the nucleus are selectively stained by six different molecular markers at the expected subcellular locations, demonstrating that MELC reliably produces established results (**Fig. 2f,g**; CMP list in **Supplementary Table 2** online). A similar control experiment using rhabdomyosarcoma cells also showed compartment-specific staining/signal separation (**Supplementary Fig. 2** online). Moreover, molecules known to be associated with each other, such as alpha- and beta-tubulin, CD3 and CD4, and nine postsynaptic proteins, were frequently found to be colocated by MELC (**Fig. 2c**; **Fig. 3b–d**; **Supplementary Fig. 2** online; and corresponding CMP lists in **Supplementary Table 2** online).

The position of a given protein in the labeling sequence might cause differences in the protein's apparent level, pushing its level in a pixel above or below its threshold, and thereby affecting the CMPs. We therefore repeated our original experiment (**Fig. 1a**) using (i) the original sequence of antibodies, (ii) the same sequence inverted and (iii) the same sequence randomly permuted, resulting in a total of 45 informative labeling reactions (**Supplementary Fig. 3a–d** online). CMPs are not significantly affected by the order in which proteins are labeled (**Fig. 3a**).

In this study, several experts manually selected the thresholds from an automatically generated range. In addition, a fully automated system based on maximizing 'mutual information'[12] also gives closely comparable values (A. Borissenko, A.V.M.D. & W.S., unpublished data). Both approaches produce very similar results (**Fig. 3a**, insert; **Fig. 3b–d**), confirming that our findings are not influenced by changes in the algorithm used to set thresholds. Performance tests revealed that MELC was specific and robust (**Supplementary Fig. 3** online).

To increase the rate of data acquisition and map proteins in a wide variety of cells and tissues, we built 15 robotic work stations (for details see **Supplementary Fig. 1b,c** online and **Supplementary Table 3** online). CMPs and CMP motifs were detected in all samples, suggesting that proteomic organization into CMPs is universal (examples in **Figs. 2d,e** and **3b,c,e,f** and **Supplementary Videos 1–3** online).

Protein organization in model diseases

To explore how disease affects protein organization, we analyzed skin biopsies from a patient with psoriasis[13]. More than 90 different molecules were mapped in each tissue section within 100 MELC cycles, comparing affected and unaffected skin (**Supplementary Video 3**, part a, online). To confirm that the maps were specific and reproducible, we remapped selected proteins on the same tissue section at different stages of the procedure. The resulting CMPs reveal substantial differences between diseased and normal skin (**Fig. 4a**,**b**).

To corroborate these findings, we analyzed the location of 49 molecules *in situ* in patients with psoriasis, atopic dermatitis (a related but distinct inflammatory skin disease[14]) and healthy

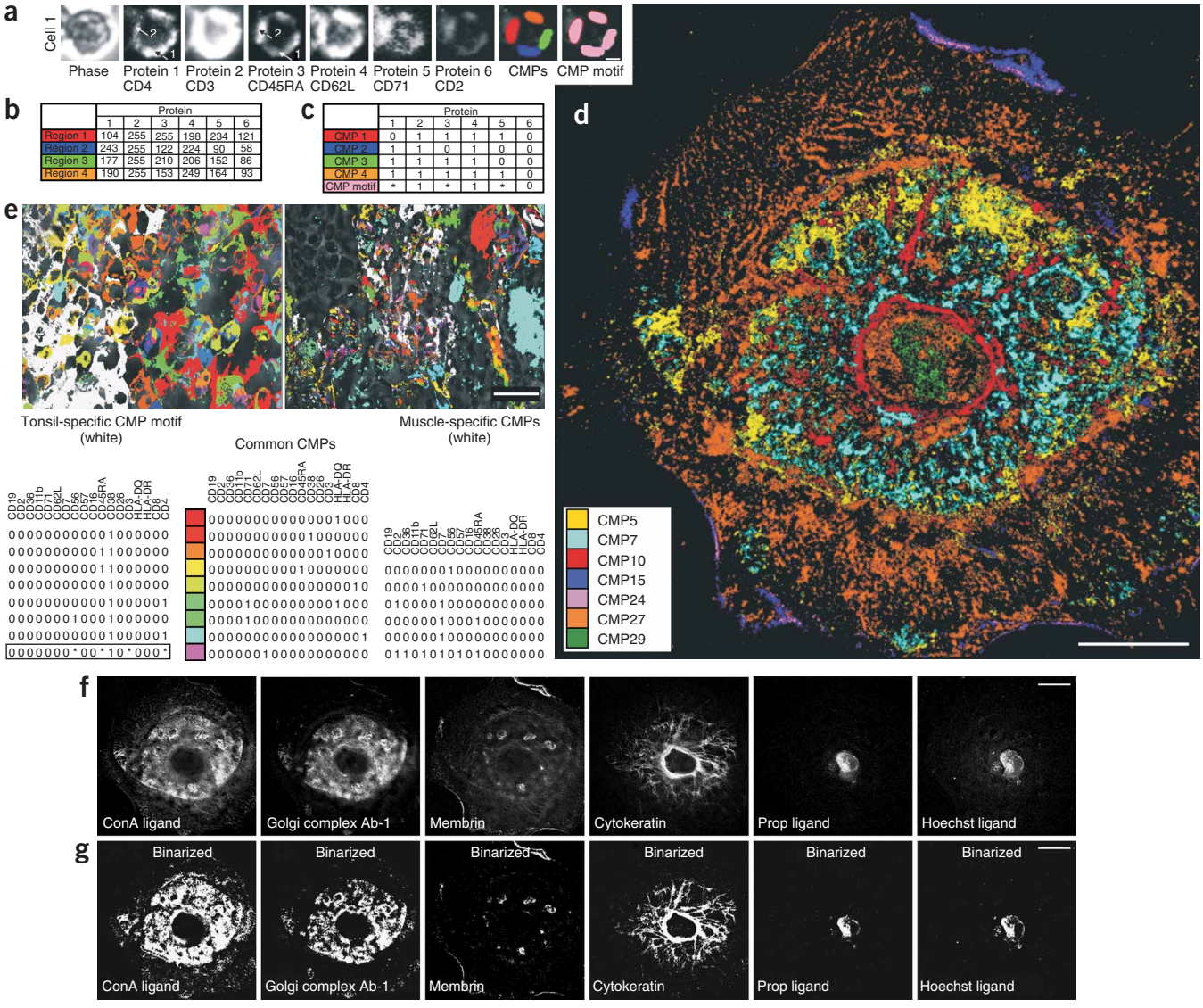

Figure 2 Construction of toponome maps. (**a**) Phase contrast image of cell 1 from **Figure 1b**, followed by MELC maps of six proteins in this cell. Arrows 1 and 2 designate regions where the protein 1 (CD4)/protein 3 (CD45RA) ratio is high and low, respectively. These regions are colored red and blue in the CMPs subpanel, and two additional regions with different protein abundances are colored orange and green. (**b**) Average fluorescence signal for each protein in each of the regions defined in **a**. (**c**) Binary representation of the data in **a**, obtained by setting a constant threshold and adjusting the signal for each protein in each region to 1 or 0 depending on whether it lies above or below this threshold resulting in different CMPs. A CMP motif characteristic of all four regions can then be defined as shown on the bottom line and visualized (purple) in the right image of **a**. Note that in this experiment, the antibody concentrations used for labeling were adjusted to give the threshold (for CMPs, see also **Supplementary Table 2** online). (**d**) Toponome map of a primary human hepatocyte produced by mapping 15 molecules (library 2), defining a set of subcellular CMPs and representing seven CMPs by different colors to display a set of mutually exclusive compartments. (The same CMPs are mapped in **Supplementary Video 1** online and are defined in **Supplementary Table 2** online). (**e**) Toponome maps of the tonsil and polymyositic muscle produced by mapping the 18 proteins mapped in **Fig. 1a**, defining prominent CMPs and representing each by a different color. CMPs unique to each tissue are shown in white; those unique to the tonsil collectively define a CMP motif. (**f**) Control experiment illustrating that six molecular markers (subset of library 2), specific either for cytoplasmic (left, four images) or nuclear structures (right, two images), give selective staining of the corresponding compartments. Note that these markers were extracted from the MELC data set of **d** and displayed in one single optical plane of the same cell obtained by deconvolution of z stack images. (**g**) Binarized data set based on threshold setting in the images shown in **f** (for image superimpositions and corresponding CMPs see **Supplementary Table 2** online). Note that the image data shown in **Figures 1a**,**b** and **2a**–**g** are representative examples of a multitude of independent MELC experiments in a broad variety of cells and tissues. Scale bars, 2 μm (**a**), 20 μm (**d**,**f**,**g**), 40 μm (**e**).

controls ($n = 6$, each). The levels of these 49 proteins in psoriasis and atopic dermatitis were generally similar when averaged across the entire tissue section (**Fig. 4c**). Toponome analysis, however, identified CMP motifs on putative T cells associated with each condition (**Supplementary Fig. 4** online), identified candidate keratinocyte stem cells *in situ* (**Supplementary Fig. 5** online) and singled out CD11a as a lead protein and potential therapeutic target in atopic dermatitis (**Supplementary Table 4** online). The latter finding was validated by CD11a-specific, FITC-labeled efalizumab (Raptiva); the binding pattern of this biological drug was nearly identical to CD11a expression (**Supplementary Fig. 6** online).

To find disease-specific CMP motifs, we worked with 48 molecules (library 7 except CD1a) in an extended cohort. We identified a $CD36^+$, $HLA\text{-}DR^+$, $CD29^-$ motif whose abundance discriminates among the three skin conditions ($n = 10$, each; $P < 0.00018$; **Fig. 4d**). Analysis of the remaining molecules revealed a larger motif expressing the lead proteins CD58/CD138/pan-CK. This larger motif predominantly denotes keratinocytes from the upper epidermis (CMP motif I, **Fig. 4e**), and is a possible hallmark of the hyperactivated suprabasal keratinocyte islands typical of psoriasis[15].

Frequency distribution of CMPs *in vivo*

Of the 2^{48} (2.81×10^{14}) possible CMPs in the previous experiment, we detected only 531,075, 599,345 and 281,743 CMPs in patients with psoriasis, atopic dermatitis and healthy controls, respectively ($n = 10$, each). We therefore attempted to correlate the number of

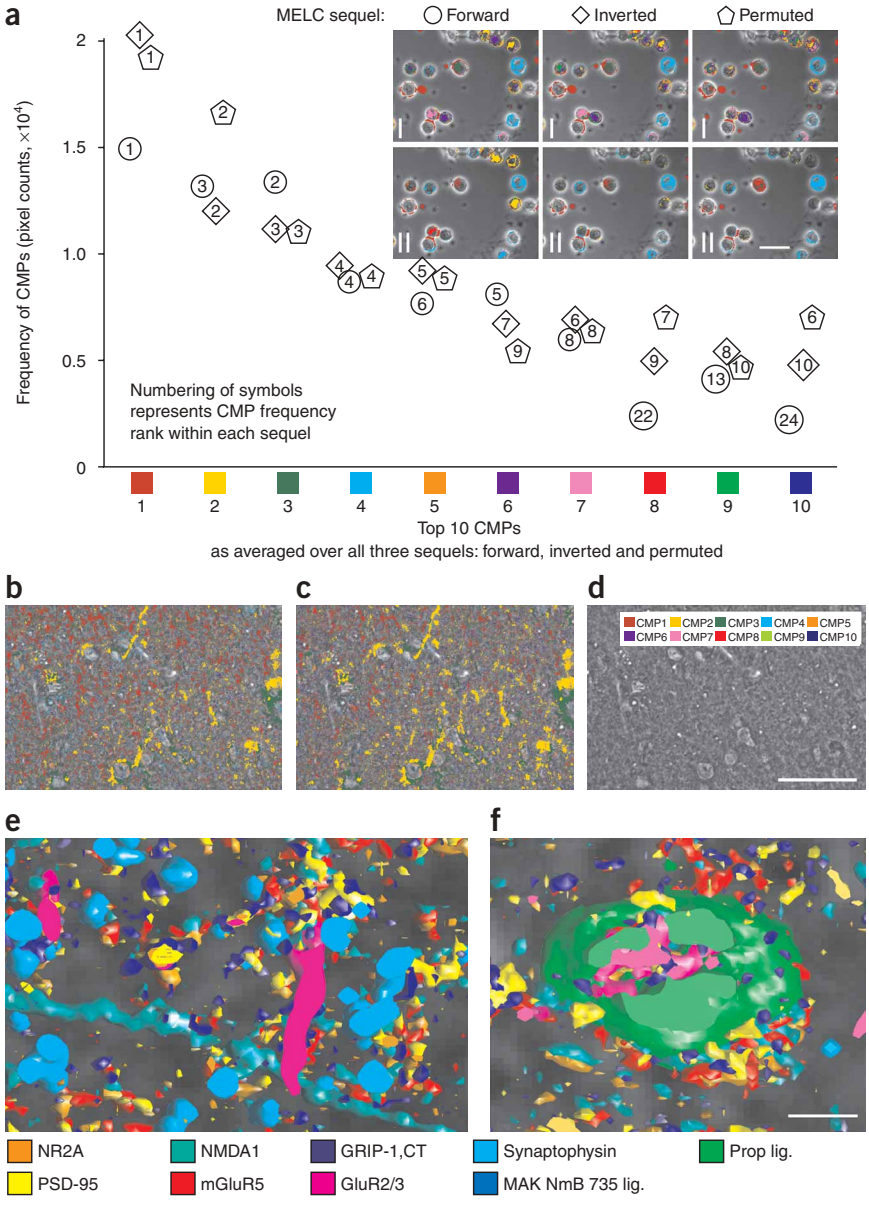

Figure 3 Robustness of CMP detection and 3D colocalization mapping. (**a**) Effect of repeated imaging and antibody order on the apparent abundance of the top ten prominent CMPs detected by MELC, using a sequence of 15 antibodies (subset of library 1), the same sequence inverted and the same sequence randomly permuted, all performed sequentially on the one sample of PBMCs using expert-adjusted protein thresholds for binarization (primary data shown in **Supplementary Fig. 3** online). Note that 90% of the top ten CMPs (as averaged over all three sequels) remain the top ten, and only some change their frequency ranks in the forward, inverted and permuted sequels. Inset shows a detail of a larger visual field illustrating the location of the ten most frequent CMPs in each of the three separate sequels of one run using (I) thresholds defined by human experts, as in the main graph or (II) thresholds defined by an automatic system based on maximizing mutual information[12], demonstrating that the apparent distribution of the CMPs is largely unaffected by the method used to set thresholds. CMPs are defined in **Supplementary Table 2** online. Numbering of symbols represents CMP frequency rank within each sequel. (**b**) Toponome map (library 8) of ten different synaptic proteins in rat prefrontal cerebral cortex produced using thresholds set by human experts. Note that the differentially colored CMPs highlight specific synaptic subregions of the brain tissue: (yellow) dendritic and perineuronal synaptic regions, (green) a subset of perineuronal synapses; (red and blue) subsets of synapses in the neuropil, together forming the major fraction of the synaptic neuronal network (representative data of three independent experiments). (**c**) Toponome map of the same sample produced using thresholds defined by the automatic system based on maximizing mutual information, again showing that the method used to set thresholds has little effect on the apparent distribution and frequency of CMPs detected (note that the CMP colors indicate the same CMPs in both experiments of **b** and **c**; CMPs are superimposed on the corresponding phase contrast image). (**d**) Phase contrast image corresponding to **b** and **c**, displayed alone; inset denotes the ten CMPs of **b** and **c** (for CMP color code see **Supplementary Table 2** online). (**e**,**f**) 3D MELC imaging experiment showing surface-rendered, colocalization map of seven postsynaptic proteins together with propidium iodide staining (library 4) in the neuropil (**e**) and a neuronal cell body (**f**) from rat spinal cord (lamina I/II). The distribution of each molecule is represented by colors as indicated, omitting overlaps between the distribution patterns. Scale bars, 1 μm (**e**,**f**), 10 μm (**a** insert), 50 μm (**b**–**d**).

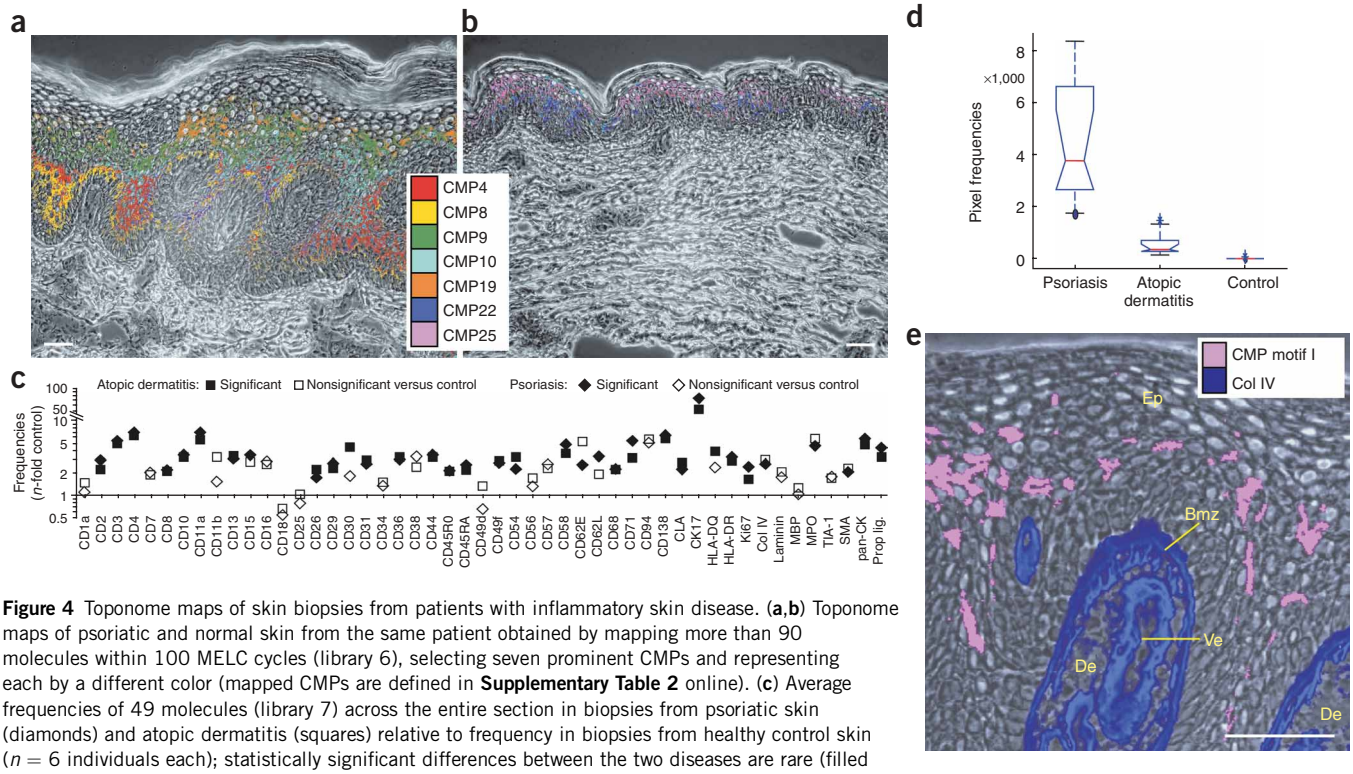

Figure 4 Toponome maps of skin biopsies from patients with inflammatory skin disease. (**a**,**b**) Toponome maps of psoriatic and normal skin from the same patient obtained by mapping more than 90 molecules within 100 MELC cycles (library 6), selecting seven prominent CMPs and representing each by a different color (mapped CMPs are defined in **Supplementary Table 2** online). (**c**) Average frequencies of 49 molecules (library 7) across the entire section in biopsies from psoriatic skin (diamonds) and atopic dermatitis (squares) relative to frequency in biopsies from healthy control skin ($n = 6$ individuals each); statistically significant differences between the two diseases are rare (filled symbols, significant difference versus controls; open symbols, insignificant difference). (**d**) Box plot showing the overall frequency of CMP motif I ($CD29^-$/$CD36^+$/$CD58^+$/$CD138^+$/HLA-DR^+/pan-CK^+) in psoriatic skin and atopic dermatitis compared with healthy control skin ($n = 10$ individuals each), showing that the frequency of this motif discriminates significantly between these diseases, in contrast to the frequency of single molecules (**c**). (**e**) Toponome map of psoriatic epidermis showing the distribution of this motif (purple) and collagen type IV (blue, marking the basal membrane zone (Bmz) and blood vessels (Ve) to distinguish the skin micro-compartments, epidermis (Ep) and dermis (De)) overlaid on the corresponding phase contrast image. Scale bars, 50 μm.

observed CMPs with their individual frequency. The relationship conformed to Zipf's law[16–19], in which the probability $P(e_n)$ of the n most frequent individual event e_n is proportional to a (negative) power of n (**Fig. 5**, blue line). Reducing the number of molecules examined weakened the relationship (**Fig. 5**, red and green lines), and it was totally absent in randomly generated data (**Fig. 5**, inset). The findings indicate that CMPs obtained by MELC unravel the highly organized toponome.

Protein organization in experimental pathology

To investigate alterations of protein organization in experimentally induced disease, we chose a model of chronic constriction injury (CCI), in which chronic neuropathic pain is induced by unilateral ligature of the sciatic nerve in rats[20]. This procedure is thought to modulate interneuronal synapses in lamina I/II of the spinal cord at lumbar level 3/4. We mapped ten synaptic proteins in this area using MELC in two treated rats (four measurements) and two healthy controls (three measurements) (library 8). These groups were distinguished by 167 out of 10^3 CMP motifs (t-test screening $P < 0.01$), of which 91% were more frequent in controls.

The most prominent CMP motif in controls (with lead proteins GRIP-1, CT and GluR2/3) showed a punctuate distribution associated with dendrites in lamina I/II (**Fig. 6a**), indicative of axo-dendritic synapses of the first afferent neuron. By contrast, a new motif in treated rats (with lead proteins PSD-95 and NR2A, which were never among lead proteins in controls) was totally restricted to single dendritic structures on the treated side (**Fig. 6b**). In CCI, the general variation in lamina I/II synapses thus appears to be replaced by a $PSD\text{-}95^+$/$NR2A^+$ motif on a subset of dendrites that probably plays a significant role in propagating chronic neuropathic pain. PSD-95 knockout mice are insensitive to CCI, supporting this conclusion[21]. As PSD-95 is an adaptor protein for NR2A in postsynaptic structures, inhibiting the interaction between NR2A and PSD-95 may offer an approach for treating chronic neuropathic pain. Moreover, the data show that proteins assembled in the average postsynaptic density, as revealed by conventional proteomics in recent studies, can be highly modified in individual synaptic regions, thus providing insight into the functional organization of synapses in the central nervous system.

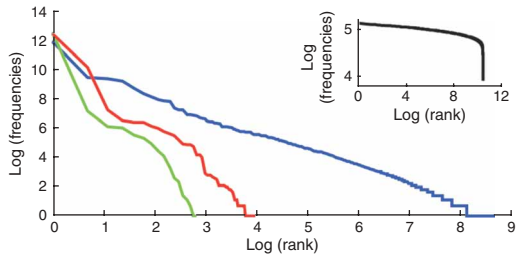

Figure 5 Zipf's plot of the relationship between the rank and the frequency of CMP motifs in human skin. Data were obtained for the 49 molecules recognized by library 7 (blue), or subsets of this library recognizing 10 molecules (red) or five molecules (green). The (log-log) linearity seen with data for 49 molecules is progressively less apparent as fewer molecules are examined. Inset, Zipf's plot of the relationship between the rank and the frequency of motifs obtained with 49 randomly distributed molecules, showing that the linearity has disappeared.

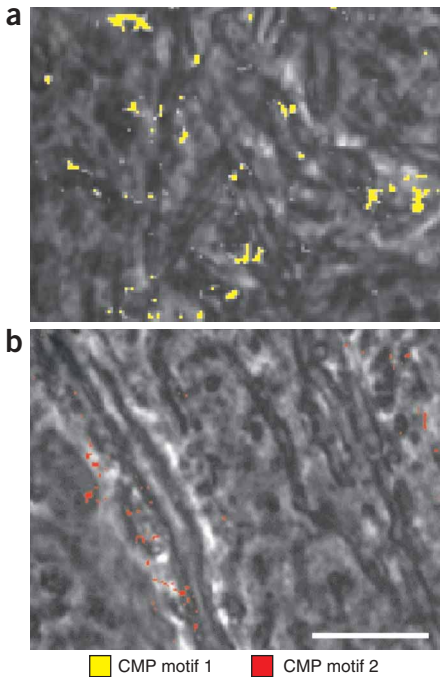

Figure 6 Toponome maps of lamina I/II in rat spinal cords with and without CCI as a model of chronic neuropathic pain. (**a**) Toponome map (library 8) showing the distribution of a CMP motif 1 apparently associated with dendrites (dark streaks) and characteristic of normal synaptic regions superimposed on a phase contrast image. (**b**) Toponome map showing the distribution of CMP motif 2, characteristic of chronic neuropathic pain (absent in **a**) and spatially associated with a small number of dendrites superimposed on a phase contrast image. Note that CMP motif 1 is specifically lost and CMP motif 2 is specifically acquired in chronic neuropathic pain (CMP motifs are defined in **Supplementary Table 2** online). Together two healthy control and two CCI rats were analyzed within an orienting study. Scale bar, 10 μm.

Identifying functional protein networks

To explore whether toponome maps can also be used to identify subcellular protein networks, we mapped 23 different cell surface proteins in the rhabdomyosarcoma cell line TE671 (**Fig. 7**). These muscle tumor cells spontaneously enter an exploratory state in which they form three cell extensions, one of which withdraws as the cell proceeds to migrate. Four cell surface CMPs were present throughout this process, forming an ordered sequence from the tip of two cell extensions towards the cell body in the exploratory state (**Fig. 7a,c**). The alanine-specific protease APN (CD13) was a lead protein in all these CMPs, as shown by its localization (**Fig. 7d**) and the breakdown of the fluorogenic substrate bis-ala-rhodamine 110 (**Fig. 7b**). Inhibition of aminopeptidases, especially APN, interferes with cell migration by an unknown mechanism[22,23]. Based on the lead protein hypothesis[2] we investigated the relationship between APN's activity and cell polarization. Treating spherical TE671 cells (**Fig. 7e**) with RB3014, a potent and specific small molecule inhibitor of APN[22], both disrupted the normal subcellular distribution of CMPs and blocked the cells' ability to enter the exploratory state for over 16 h (**Supplementary Video 4**, part a, online). When cells in the exploratory state were incubated with a mAb against membrane-bound APN (SJ1D1), the cell extensions both lost their ordered arrangement of CMPs (**Fig. 7g**) and were withdrawn as the cells returned to the spherical state (**Fig. 7f**) ($n = 4$). This proves several points. First, both the presence and the proteolytic activity of APN are essential for the topological integrity of CMPs; second, when this integrity is disturbed, the cell can no longer polarize or migrate, which may point to a route for selective inhibition of metastasis; and third, the cell surface proteins assembled as CMPs are organizationally dependent on APN, indicating that these proteins form a molecular network. We have thus detected a molecular network *in situ* and demonstrated that its cellular function is dependent on the continuing activity of its lead protein.

Notably, the average fluorescence intensities of the 23 cell surface proteins per cell, including APN, showed little if any reliable difference between the spherical and the exploratory cell states (**Fig. 7h**). At the CMP level, however, these states were readily distinguishable (**Fig. 7i**), underlining the advantages of the toponome approach to study cell functions.

We have collected several hundred images of this cell polarization process to construct, *in silico*, a video of normalized CMP movements in a synthetic TE671 cell with and without APN inhibition (**Supplementary Video 4**, part b, online). Intriguingly, APN appears not to be among the lead proteins in normal muscle cells and neither RB3014 nor SJ1D1 have a significant effect on such cells (results not shown).

DISCUSSION

MELC allows a large number of molecular components to be colocalized in fixed cells and tissue. It can detect any molecule for which a fluorescently labeled ligand is available. In addition, it can identify transient or rare protein associations in three dimensions in a subset of cells or in subcellular compartments with unprecedented detail, and can be used to reconstruct the dynamics of molecular networks by *in silico* modeling. MELC can thus be used to show when and where proteins colocalize in cellular or tissue compartments, which proteins are excluded, how these associations are altered by pathology and how proteins are functionally interrelated. This makes it possible to systematically decipher the rules governing protein co- and anti-colocation within individual cells. Recently established approaches of automated analysis of subcellular patterns of single proteins in cells[24,25] and analysis of similar patterns in tissues[26] are important for high-throughput analysis. MELC is an automated multidimensional fluorescence imaging technology with functional resolution.

Construction of tag libraries requires thorough calibration procedures to exclude sterical hindrance of tags. Sterical hindrance may occur when different monoclonal antibodies are directed against neighboring epitopes. Such interference may provide important biological information. The tag libraries used in this study were highly calibrated by determining the proper sequence of labeling cycles. Therefore, an important limitation of MELC is not sterical hindrance, but the discovery of new proteins. We could however address this limitation in the future by using large recombinant randomized antibody libraries. Because a fixed sample withstands the MELC procedure for at least two weeks, we may attempt experiments comprising more than 1,000 tags and screening of tissue microarrays.

The protein clusters or CMPs described in this study differ from previously described protein complexes by including weakly or transiently interacting proteins, rather than being restricted to strongly interacting proteins whose complexes withstand purification. Interestingly, very few protein localizations detected in this study, in particular in the rhabdomyosarcoma cell line, have previously been reported. CMPs represent the basic combinatorial molecular units of proteomes in all the cell and tissue types we have analyzed, and are organized into higher order units, so-called CMP motifs. Their frequencies follow the rules observed in many highly organized systems, as they fulfill Zipf's law. The frequencies of individual

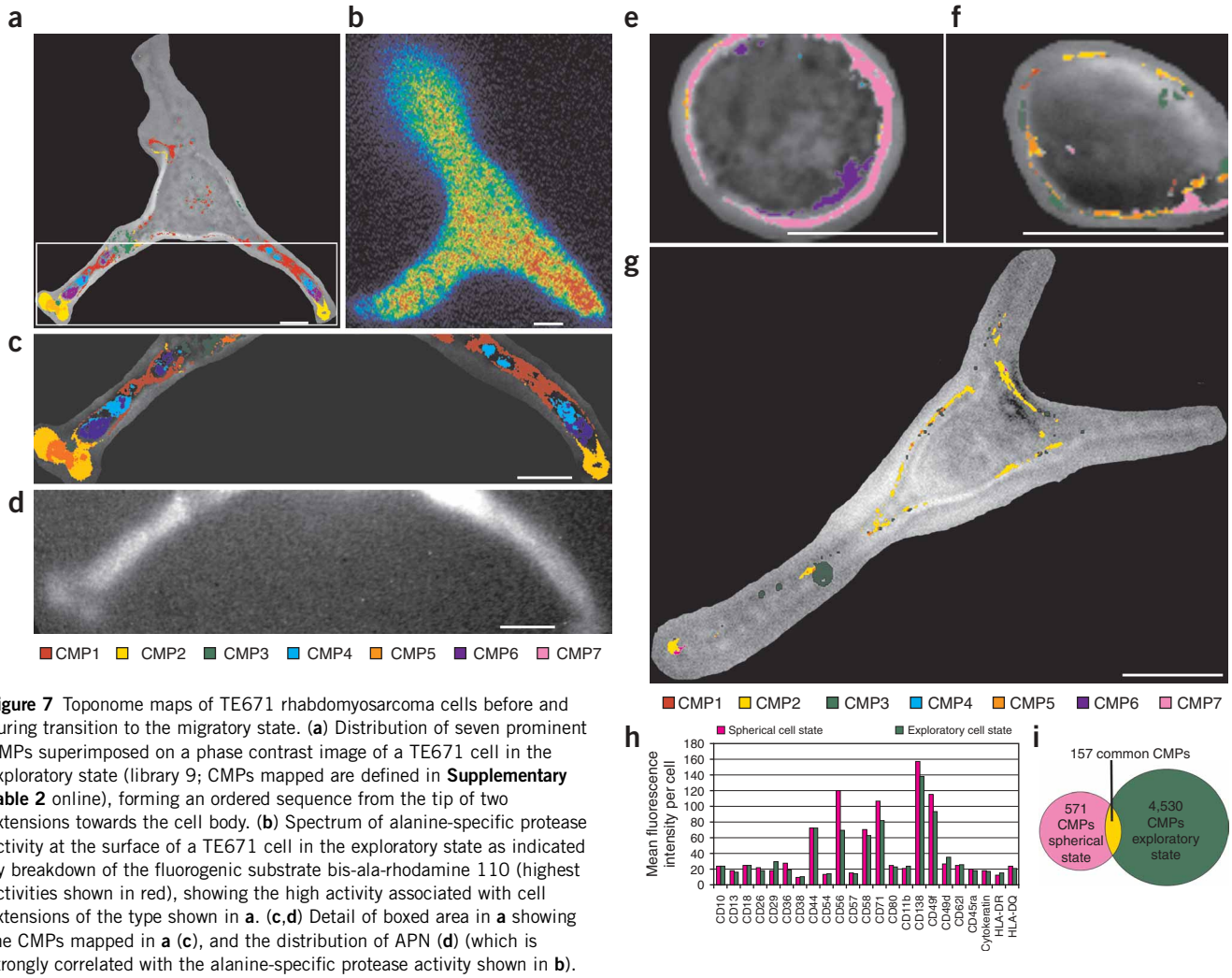

Figure 7 Toponome maps of TE671 rhabdomyosarcoma cells before and during transition to the migratory state. (**a**) Distribution of seven prominent CMPs superimposed on a phase contrast image of a TE671 cell in the exploratory state (library 9; CMPs mapped are defined in **Supplementary Table 2** online), forming an ordered sequence from the tip of two extensions towards the cell body. (**b**) Spectrum of alanine-specific protease activity at the surface of a TE671 cell in the exploratory state as indicated by breakdown of the fluorogenic substrate bis-ala-rhodamine 110 (highest activities shown in red), showing the high activity associated with cell extensions of the type shown in **a**. (**c**,**d**) Detail of boxed area in **a** showing the CMPs mapped in **a** (**c**), and the distribution of APN (**d**) (which is strongly correlated with the alanine-specific protease activity shown in **b**). (**e**,**f**) Distribution of the same CMPs superimposed on phase contrast images of TE671 cells in the spherical state before (**e**) and after (**f**) treatment with mAb SJ1D1. Note the disordered arrangement of the CMPs in **f** compared with **e**. (**g**) Distribution of the same CMPs superimposed on a phase contrast image of a TE671 cell in the exploratory state treated with SJ1D1, showing that the ordered sequence of CMPs seen in the untreated cell (**a**) has been lost as the cell reverts to the spherical state (see also **Supplementary Video 4**, part a, online). (**h**) Mean fluorescence intensities of the 23 molecules per cell investigated in the spherical cell of (**e**) and polarized cell of (**a**). (**i**) Total number of CMPs of these molecules, giving unique CMP clusters for the two cell states with very little overlap. Note, however only little difference in the degree of expression of single epitopes as shown in **h**. Together a total of 190 MELC experiments and 93 corresponding live cell imaging experiments were performed to reconstruct and explore experimentally the time-dependent steps of the cell polarization toponome dynamics. Scale bars, 10 μm.

molecules[17,27], and molecular interactions[28] all appear to obey similar power laws. Our study indicates that this law also governs the frequency of protein colocations *in vivo*. MELC makes protein systems *in situ* amenable to mathematical analysis using methods from combinatorial statistics and geometry.

The CMP motifs we have described in model skin diseases, chronic pain and tumor cell polarization are assembled around state-specific lead proteins, probably representing the interaction hubs typical of scale-free networks[10,11]. We show that lead proteins are highly predictive molecular elements controlling the formation of molecular networks and the associated cellular functionalities. MELC has been used to identify FcγRIII as a lead protein and potential target in amyotrophic lateral sclerosis[29], a finding that has now been confirmed by the phenotype of an Fcγ receptor knockout mouse[30].

In conclusion, MELC provides a tool to analyze protein organization in cells and to identify candidate target proteins in disease. Our findings also demonstrate how functional protein networks can be deciphered at the single cell level: once lead proteins have been identified in functionally relevant CMP motifs, they can be selectively inhibited or deleted, and the functional consequences observed. A systematic MELC-based approach using a proteome-wide set of affinity ligands might decipher the complete functional plan of a cell or a tissue using lead proteins as anchor points.

METHODS

Patients and antibodies. All participants gave informed consent and the experiments were approved by the local ethics committee. Demographic data of patients and controls are presented in **Supplementary Table 5** online. Samples were examined using a blind analysis. Punch biopsies 6 mm in diameter of representative skin conditions were taken at the time of patient admission and snap-frozen in liquid nitrogen.

Antibodies were conjugated to PE or FITC and are listed in **Supplementary Table 1** online.

Preparation of cells and tissue sections. We prepared 5-µm tissue sections at –25 °C using a cryotome, fixed in acetone at –20 °C for 10 min and stored at –20 °C for several days or –80 °C for longer time intervals until use. PBMCs were isolated from the blood of a healthy volunteer using Ficoll, placed on a slide and air-dried before snap freezing. Cultured human cells (TE671 rhabdomyosarcoma cell line, and primary human hepatocytes) were washed with 10% DMSO in PBS, fixed in fresh 4% paraformaldehyde, and washed in PBS, before air drying and snap freezing as before. Before use (see section on MELC below), all samples were rehydrated in PBS at 20 °C, incubated with normal goat serum for 30 min, and washed again in PBS.

Sciatic nerve ligation to model CCI in rats was performed as described[20] (provided by F. Hucho and coworkers, FU Berlin). After killing the animals 7 d later, the spinal cord was removed, and small tissue blocks of the relevant spinal cord layers L3/4 were prepared as above.

To localize alanine-specific proteases, we incubated TE671 cells in 10^{-5} M bis-ala-rhodamine 110, and the corresponding fluorogenic signal was imaged over 180 s using a confocal laser scanning microscope. Live cells were imaged using a cell observer (Zeiss; **Supplementary Video 4** online). Fresh human muscle biopsies and rat brain tissue were prepared as described above.

MELC. Aspects of MELC technology have been described previously[2]. In brief, a slide with a given specimen was placed on the stage of an inverted wide-field fluorescence microscope (Leica or Zeiss) equipped with fluorescence filters for FITC and PE. Fluorochrome-labeled antibodies and wash solutions were added and removed robotically under temperature control, avoiding any displacement of the sample and objective. In each cycle, a pair of antibodies was added; phase contrast and fluorescence images were acquired by a high-sensitivity cooled CCD camera; the sample was washed with PBS and bleached at the excitation wavelengths; and postbleaching phase contrast and fluorescence images were acquired. Data acquisition was fully automated using software developed by ZENIT Technology Park and the MPRR group (**Supplementary Fig. 1c** online). For analyzing protein abundance in skin samples, the frequency of pixels positive for each molecule in each tissue compartment (such as epidermis or dermis) was normalized to a horizontal width of 100 µm to compensate for the vertical stratification of epithelial tissue, and tissue compartments were distinguished using masks defined by a pan-cytokeratin antibody. Controls for MELC data were performed using conventional immunocytochemistry or in some cases flow cytometry analyses of PBMCs, to examine whether the frequencies and distribution patterns of given labeled proteins were comparable (data not shown).

Construction of 2D toponome maps. Fluorescence images produced by each antibody were aligned pixel-wise using the phase contrast images, with an in-register accuracy of +/– one pixel. Background and illumination faults were then removed by flat-field correction. Fluorescent pixels or voxels were then parsed by regarding the list of fluorescence intensities $I_1, I_2, I_3 \ldots I_n$ for proteins 1, 2, 3…n in any particular pixel or voxel as the values of an n-dimensional vector associated with that pixel or voxel. This vector can then be binarized by selecting thresholds $T_1, T_2, T_3 \ldots T_n$ for proteins 1, 2, 3…n, and setting the vector values for any protein m to zero if $I_m < T_m$ and to 1 if not, using thresholds manually set by human experts from within an automatically generated range. The binarized images were then combined to form a list of CMPs representing the proteins expressed in each pixel, or groups of CMPs representing regions of interest. This procedure is performed by using MATLAB programming language allowing us 52 epitopes to be represented in one 2D array of doubles, to be stored and analyzed. For higher numbers of epitopes several 2D arrays are combined. CMPs were defined as closely related when they had in common the same lead protein(s), but were distinct by the presence or absence of one or several other proteins. Different degrees of CMP variation were classified ranging from slight (one different protein) to extended (two and more different proteins).

Data analysis, statistics and visualization. To analyze CMPs we developed 'MotifFinder' and 'MotifAnalyzer' software packages to search for and visualize CMP motifs. Briefly, to cluster CMPs and identify CMP motifs of interest, MotifFinder looks for CMPs or CMP motifs whose overall frequency differs significantly in two different sample groups (e.g., healthy versus diseased tissue). To do this, it calculates the frequency of each n-member CMP motif with a bounded number of 'non-wild cards' for each run from either group, starting with all CMP motifs consisting entirely of wild cards except at one or two positions and progressively decreasing the number of wild cards in the motifs up to a limit imposed by computational resources. Following this, it identifies the CMPs or CMP motifs whose frequency distribution differs significantly in the two sample groups using the Wilcoxon rank-sum test or Student's t-test. MotifAnalyzer then allows the user to visualize where the pixels (or voxels) carrying these motifs are located in a given biological structure, permitting visual exploration of the data and suggesting possible biological explanations for the motif distributions that are observed. The motifs of greatest interest are then color-coded and superimposed on the corresponding biological structures to create toponome maps.

3D imaging. 3D imaging of MELC runs was performed by generating and visualizing z-stack raw images for each antibody signal from top to bottom of a sample; deconvoluting these images using a standard algorithm (XCOSM)[31] working with a specific point spread function; setting thresholds for each antibody signal from each optical plane as before; overlaying all binarized images to construct large scale protein colocalization maps using a MATLAB algorithm; and constructing 3D toponome maps in the same way as was done for two dimensions. The latter two visualization steps were performed by using algorithms provided by IMARIS software packages.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank K. Neubert for supplying bis-ala-rhodamine 110, B.P. Roques for RB3014, and A. Borissenko and P. Karcher for computational help. The study was supported by the *Bundesministerium für Bildung und Forschung* (NBL-3 FKZ: 01ZZ0107, 01ZZ0407; NGFN2 01 GR 0446, CELLECT 0312844, BioChance 0312452), the DFG (627/1-8; INK "Bildinformation") and Land Saxony-Anhalt. A.W.M. Dress thanks the Center for Combinatorics at Nankai University for hospitality during the preparation of this manuscript. We are grateful for animal tissue provided by F. Hucho/O. Bogen and fruitful discussions, and appreciate critical reading by A. Leech. We thank Anja Bastian, Katrin Brennecke and Franziska Böckelmann for technical assistance.

AUTHOR CONTRIBUTIONS

W.S. invented MELC and designed the study. M.F. and M.B. closely cooperated with L.P., A.J.P. and W.S. to develop robots and corresponding biological assays. A.W.M.D. performed mathematical image analyses. W.S., M.F. and M.B. designed, performed and interpreted data of multidimensional analyses of protein locations. CCI tissue material was analyzed and interpreted by W.S. and corresponding statistical motif analyses were done by L.P. Experiments with rhabdomyosarcoma cells were performed by M.F. and analyzed and interpreted in close cooperation with M.B., L.P. and W.S. 3D-images were generated by M.B. in close cooperation with W.S. W.S., M.B. and M.F. prepared the corresponding parts of the manuscript. B.B., R.B. and H.G. closely cooperated with A.J.P., Y.M. and L.P. in the dermatological and skin tissue–related investigations. In detail, B.B., A.J.P., R.B. and H.G. designed these clinico-experimental studies. B.B. and H.G. performed the clinical study part. R.B. did most of the accompanying conventional immunohistology. A.J.P., Y.M. and L.P. did the MELC experiments. R.B. binarized most of the MELC images and performed statistics. L.P. contributed all computer analyses and visualizations. B.B., A.J.P., L.P., R.B., Y.M. and H.G. interpreted these data and prepared the corresponding parts of the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that their competing financial interests in this work are too numerous to itemize.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Collins, F.S., Green, E.D., Guttmacher, A.E. & Guyer, M.S. US National Human Genome Research Institute. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
2. Schubert, W. Topological proteomics, toponomics, MELK technology. *Adv. Biochem. Engin. Biotechnol.* **83**, 189–209 (2003).

3. Herman, B., Krishnan, R.V. & Centonze, V.E. Microscopic analysis of fluorescence resonance energy transfer (FRET). *Methods Mol. Biol.* **261**, 351–370 (2004).
4. Jares-Erijman, E.A. & Jovin, T.M. FRET imaging. *Nat. Biotechnol.* **21**, 1387–1395 (2003).
5. Ganesan, S., Ameer-Berg, S.M., Ng, T.T., Vojnovic, B. & Wouters, F.S. A dark yellow fluorescent protein (YFP)-based resonance energy-accepting chromoprotein (REACh) for Forster resonance energy transfer with GFP. *Proc. Natl. Acad. Sci. USA* **103**, 4089–4094 (2006).
6. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
7. Perfetto, S.P., Chattopadhyay, P.K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat. Rev. Immunol.* **8**, 648–655 (2004).
8. Pacholski, M.L. & Winograd, N. Imaging with mass spectrometry. *Chem. Rev.* **99**, 2977–3006 (1999).
9. Stoeckli, M., Chaurand, P., Hallahan, D.E. & Caprioli, R.M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat. Med.* **7**, 493–496 (2001).
10. Bray, D. Molecular networks: the top-down view. *Science* **301**, 1864–1865 (2003).
11. Han, J.-D.J. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
12. Dress, A.W.M., Lokot, T., Pustylnikov, L.D. & Schubert, W. Poisson numbers and poisson distributions in subset surprisology. *Ann. Combinatorics* **8**, 475–485 (2004).
13. Gottlieb, A.B. *et al.* TNF inhibition rapidly down-regulates multiple proinflammatory pathways in psoriasis plaques. *J. Immunol.* **175**, 2721–2729 (2005).
14. Boguniewicz, M. & Leung, D.Y. Atopic dermatitis. *J. Allergy Clin. Immunol.* **117**, S475–480 (2006).
15. Böckelmann, R. *et al.* Suprabasal overexpression of the hsRPB7 gene in psoriatic epidermis as identified by a reverse transcriptase-polymerase chain reaction differential display model comparing psoriasis plaque tissue with peritonsillar mucosa. *Am. J. Pathol.* **158**, 367–372 (2001).
16. Zipf, G.K.. *Human Behaviour and the Principle of Least Effort* (Addison-Wesley, Cambridge MA, 1949).
17. Hoyle, D.C., Rattray, M., Jupp, R. & Brass, A. Making sense of microarray data distributions. *Bioinformatics* **18**, 576–584 (2002).
18. Ferrer i Cancho, R. & Sole, R.V. Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 788–791 (2003).
19. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
20. Bennett, G.J. & Xie, Y.K. A peripheral mononeuropathy in rat that produces disorders of pain sensation like those seen in man. *Pain* **33**, 87–107 (1988).
21. Garry, E.M. *et al.* Neuropathic sensitization of behavioral reflexes and spinal NMDA receptor/CaM kinase II interactions are disrupted in PSD-95 mutant mice. *Curr. Biol.* **13**, 321–328 (2003).
22. Chen, H., Roques, B.P. & Fournie-Zaluski, M.C. Design of the first highly potent and selective aminopeptidase N (EC 3.4.11.2) inhibitor. *Bioorg. Med. Chem. Lett.* **9**, 1511–1516 (1999).
23. Kehlen, A., Lendeckel, U., Dralle, H., Langner, J. & Hoang-Vu, C. Biological significance of aminopeptidase N/CD13 in thyroid carcinomas. *Cancer Res.* **63**, 8500–8506 (2003).
24. Conrad, C. *et al.* Automatic identification of subcellular phenotypes on human cell arrays. *Genome Res.* **14**, 1130–1136 (2004).
25. Chen, X. & Murphy, R.F. Objective clustering of proteins based on subcellular location patterns. *J. Biomed. Biotechol.* **2**, 87–95 (2005).
26. Uhlen, M. *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteomics* **4**, 1920–1932 (2005).
27. Koonin, E.V., Wolf, Y.I. & Karev, G.P. The structure of the protein universe and genome evolution. *Nature* **420**, 218–223 (2002).
28. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
29. Schubert, W. Method of blocking cytotoxic activity in patients with amyotrophic lateral sclerosis using antibodies to FcγRIII. US patent no. 6,638,506 (first published as international patent application WO 99/29731, 1999).
30. Mohamed, H.A. *et al.* Immunoglobulin Fc gamma receptor promotes immunoglobulin uptake, immunoglobulin-mediated calcium increase, and neurotransmitter release in motor neurons. *J. Neurosci. Res.* **69**, 110–116 (2002).
31. Conchello, J.A. & McNally, J.G. Fast regularization technique for expectation maximization algorithm for computational optical sectioning microscopy. in *Three-Dimensional Microscopy: Image Acquisition and Processing* Cogswell, C.J., Kino, G.S. & Wilson, T. (eds.) *Proc. SPIE* **2655**, 199–208 (1996).