High-throughput subcellular protein localization using cell arrays

Y.-H. Hu, D. Vanhecke, H. Lehrach and M. Janitz[1]

Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Fabeckstrasse 60–62, 14195 Berlin, Germany

Abstract

Accomplishment of the human and mouse genome projects resulted in accumulation of extensive gene sequence information. However, the information about the biological functions of the identified genes remains a bottleneck of the post-genomic era. Hence, assays providing simple functional information, such as localization of the protein within the cell, can be very helpful in the elucidation of its function. Transfected cell arrays offer a robust platform for protein localization studies. Open reading frames of unknown genes can be linked to a His_6-tag or GFP (green fluorescent protein) reporter in expression vectors and subsequently transfected using the cell array. Cellular localization of the transfected proteins is detected either by specific anti-His-tag antibodies or directly by fluorescence of the GFP fusion protein and by counterstaining with organelle-specific dyes. The high throughput of the method in terms of information provided for every single experiment makes this approach superior to classical immunohistological methods for protein localization.

The intracellular localization of gene products can provide important information about their function. Since the eukaryotic cell is characterized by a high degree of compartmentalization, most of the protein activities can be assigned to a particular cellular compartment. Moreover, interactions of proteins within regulatory networks greatly rely on proper localization of their components. Aberrant translocation of proteins is in many cases highly correlated with pathological changes in cell physiology. Intracellular localization together with other data, such as gene sequence, expression levels and interaction with other proteins, significantly contribute to the integrative functional characterization of novel genes.

Until recently protein localization studies have been limited to a particular gene of interest. More recently, following the completion of the genome sequencing programmes from various organisms, in particular those of human and mouse, great efforts have been made towards the establishment of genome-wide protein localization studies. The first large-scale studies were performed on fission yeast, where 250 proteins have been analysed [1]. Subsequent studies in budding yeast, like the one performed by Kumar et al. [2], comprised even greater number of proteins. A first high-throughput localization study in mammalian cells has been performed by Simpson et al. [3] and comprised 107 human genes transfected into the Vero cell line. Recently, automated transfection and immunostaining of mammalian cells on a 96-well plate format has been established [4]. However, a microwell plate-based approach is still characterized by a significant reagent consumption and a need for automation of liquid dispensing.

A recently developed TCA (transfected cell array) technique [5] represents a cost-effective alternative for high-throughput functional genomics. The principle of the TCA technique is based on transfection of DNA or RNA molecules immobilized on a solid surface into mammalian cells with subsequent detection of the physiological effects caused by the introduction of the foreign nucleic acid on these cells (for a review, see [6]). In our laboratory, we optimized this platform for high-throughput protein localization studies, which we used for the characterization of a variety of human proteins.

Control proteins localizing to known intracellular compartments, as well as new genes of interest, are cloned into a Gateway pENTRY vector. Using the Gateway technology, these genes can be rapidly recloned into various expression vectors for different organisms and they can contain different protein tags. For our intracellular localization studies, the genes were recloned into the pDEST26 vector in order to construct mammalian expression plasmids containing a His_6-tag at the N-terminus of every open reading frame. These plasmid constructs were then spotted in an array format on to a glass slide, treated with transfection reagent and subsequently covered by a monolayer of HEK-293T cells (human embryonic kidney 293T cells). Only the cells growing on top of each spot on the glass slide become transfected (reverse transfection) with plasmid DNA. After 2 or 3 days, the cell monolayer is fixed and the expressed proteins are detected using anti-His antibodies. Compartmentalization of the proteins is determined by counterstaining each cellular compartment and subsequent analysis using confocal microscopy (Figure 1). Because of the high-throughput cloning strategy used, approx. 60% of the transfected plasmid constructs could be detected and the corresponding proteins classified according to the intracellular localization (nucleus, endoplasmic

Figure 1 | Confocal microscopy examination of protein compartmentalization

Gene expression vectors are spotted on to a glass surface and reverse transfected for overexpression in cells growing on top of the spots (**A**). Using microscopy and different counterstaining dyes, localization of the different proteins in particular cellular compartments can be determined in a single experiment. Images (**B**–**E**) show localization of EGFP (enhanced GFP), lamin A, KDEL (Lys-Asp-Glu-Leu) receptor 1 and cadherin 8 proteins respectively. Scale bars on (**B**–**E**) correspond to 10 μm.
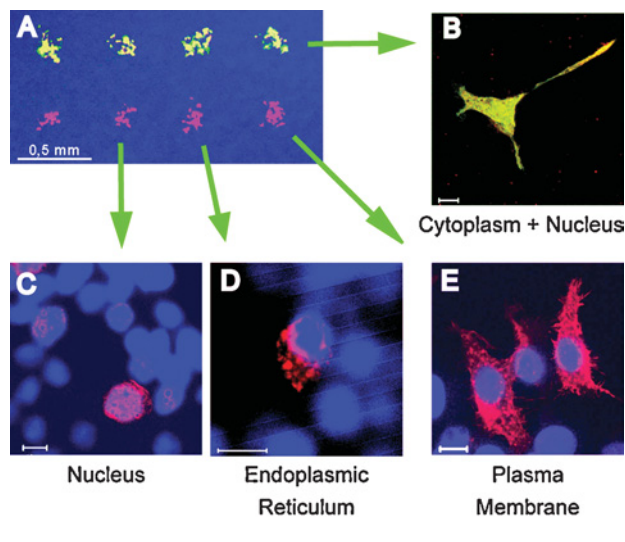
reticulum, cytoplasm, Golgi and plasma membrane) (Y.-H. Hu, unpublished work).

A His tag instead of a GFP (green fluorescent protein) fusion protein was used to track the expressed proteins, since GFP has been described to contribute to the aberrant localization of fusion proteins by changing protein folding and/or by masking the targeting signals of proteins [7–9]. Moreover, the His_6-tag can subsequently be used for high-throughput purification of recombinant proteins for downstream studies such as generation of protein chips or protein structure analysis.

By combining the cell array with the detection of proteins using the small His_6-tag epitope and a number of counterstaining methods for organelle identification, comprehensive information on the subcellular phenotype of novel proteins can be obtained in a very cost-effective way. Furthermore, changes in cell morphology and proliferation as a result of the overexpression of some of the exogenous proteins can be observed, further contributing to the identification of the putative functions of these proteins. The cell array-based experimental platform described in this report can be further up-scaled to genome-wide protein expression analysis. This is because up to 8000 DNA spots can be easily immobilized on a glass slide and subsequently transfected. For such a large scale, however, automatic image acquisition has to be optimized for the identification and classification of subcellular phenotypes. Automatic microscope-based image acquisition systems on glass slides and microwell plates have recently been established [4,10]. For example, Conrad et al. [10] implemented machine-learning methods to train and modulate the automatic classification of subcellular protein localization patterns. In these experiments, an average accuracy of 82% was achieved for determination of 11 phenotypes. In contrast, only 50% accuracy was achieved for detection of endoplasmic reticulum-, microtubule- and mitochondria-specific proteins due to the structural similarities of these organelles under the microscope. There are still several challenges concerning localization classification methods for large-scale screenings. Biological variations of the same cellular compartment in different cells, as well as variations as a result of expression of exogenous proteins, still hamper accurate automated phenotype analysis.

Nevertheless, automated high-throughput approaches based on the cell array format are expected to contribute to the acceleration of functional genomic studies. Integration of existing genome-wide gene expression data with simple functional protein properties, such as intracellular localization patterns, will lead to further elucidation of physiological and pathological cellular processes on the global level.

References

1 Ding, D.Q., Tomita, Y., Yamamoto, A., Chikashige, Y., Haraguchi, T. and Hiraoka, Y. (2000) Genes Cells **5**, 169–190
2 Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y. et al. (2002) Genes Dev. **16**, 707–719
3 Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R. and Wiemann, S. (2000) EMBO Rep. **1**, 287–292
4 Liebel, U., Starkuviene, V., Erfle, H., Simpson, J.C., Poustka, A., Wiemann, S. and Pepperkok, R. (2003) FEBS Lett. **554**, 394–398
5 Ziauddin, J. and Sabatini, D.M. (2001) Nature (London) **411**, 107–110
6 Vanhecke, D. and Janitz, M. (2004) Oncogene **23**, 8353–8358
7 Pouli, A.E., Kennedy, H.J., Schofield, J.G. and Rutter, G.A. (1998) Biochem. J. **331**, 669–675
8 Brosius, U., Dahmel, T. and Gärtner, J. (2002) J. Biol. Chem. **277**, 774–784
9 Palmer, E. and Freeman, T. (2004) Comp. Funct. Genom. **5**, 342–353
10 Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lörch, T., Ellenberg, J., Pepperkok, R. and Eils, R. (2004) Genome Res. **14**, 1130–1136