

Principles of Bioimage Informatics: Focus on Machine Learning of Cell Patterns

Luis Pedro Coelho^{1,2,3}, Estelle Glory-Afshar^{3,6}, Joshua Kangas^{1,2,3},
Shannon Quinn^{2,3,4}, Aabid Shariff^{1,2,3}, and Robert F. Murphy^{1,2,3,4,5,6}

¹ Joint Carnegie Mellon University–University of Pittsburgh Ph.D.
Program in Computational Biology

² Lane Center for Computational Biology, Carnegie Mellon University

³ Center for Bioimage Informatics, Carnegie Mellon University

⁴ Department of Biological Sciences, Carnegie Mellon University

⁵ Machine Learning Department, Carnegie Mellon University

⁶ Department of Biomedical Engineering, Carnegie Mellon University

Abstract. The field of bioimage informatics concerns the development and use of methods for computational analysis of biological images. Traditionally, analysis of such images has been done manually. Manual annotation is, however, slow, expensive, and often highly variable from one expert to another. Furthermore, with modern automated microscopes, hundreds to thousands of images can be collected per hour, making manual analysis infeasible.

This field borrows from the pattern recognition and computer vision literature (which contain many techniques for image processing and recognition), but has its own unique challenges and tradeoffs.

Fluorescence microscopy images represent perhaps the largest class of biological images for which automation is needed. For this modality, typical problems include cell segmentation, classification of phenotypical response, or decisions regarding differentiated responses (treatment vs. control setting). This overview focuses on the problem of subcellular location determination as a running example, but the techniques discussed are often applicable to other problems.

1 Introduction

Bioimage informatics employs computational and statistical techniques to analyze images and related metadata. Bioimage informatics approaches are useful in a number of applications, such as measuring the effects of drugs on cells [1], localizing cellular proteomes [2], tracking of cellular motion and activity [3], mapping of gene expression in developing embryos [4,5] and adult brains [6], and many

others. Traditionally, bioimage analysis has been done by visual inspection, but this is tedious and error-prone. Results from visual analysis are not easily compared between papers or groups. Furthermore, as bioimage data is increasingly used to understand gene function on a genome scale, datasets of subtle phenotype changes are becoming too large for manual analysis [7,8,9]. For example, it is estimated that having a single image for every combination of cell type, protein, and timescale would require on the order of 100 billion images [10]. Over the past fourteen years, the traditional visual, knowledge-capture approach has begun to be replaced with automated, data-driven approaches [11,12,13,14,15]. Approaches to quantitatively associate image feature information with additional anatomical and ontological knowledge to generate digital atlases have also been described [16].

This brief review will focus on bioimage informatics approaches to analyzing protein subcellular patterns, with the goal of illustrating many of the principles that are also relevant to other areas of bioimage informatics. The goal of work in this area is to devise a generalizable, verifiable, mechanistic model of cellular organization and behavior that is automatically derived from images [17].

The most commonly used method for determining subcellular location is fluorescence microscopy. Images are collected by tagging a protein or other molecule so that it becomes visible under the fluorescence microscope.

2 Cell Segmentation

A first step in many analysis pipelines is segmentation, which can occur at several levels (e.g., separating nuclei, cells, tissues). This task has been an active field of research in image processing over the last 30 years, and various methods have been proposed and analysed depending on the modality, quality, and resolution of the microscopy images to analyze [18]. We only discuss two commonly used approaches, Voronoi and seeded watershed.

The two first methods require *seed regions* to be defined. These can be simply locally bright regions or can be defined by a more complex procedure. For example, nuclear segmentation, which is a difficult problem by itself [19], is often used to provide seeds for cell-level segmentation.

In Voronoi segmentation [20], a pixel is assigned to the closest seed. This is a very fast method and works well for sparse regions, but does not take into account the location of the cells and makes serious mistakes if the field is crowded.

Seeded watershed segmentation is a region growing approach [21] in which the image can be considered as a landscape with the pixel intensities as elevation. From the seeds, the basins of the landscape are flooded. When two basins are about to merge, a dam is built that represents the boundary between the two cells. This method works well if the seeds are carefully defined (see Figure 1).

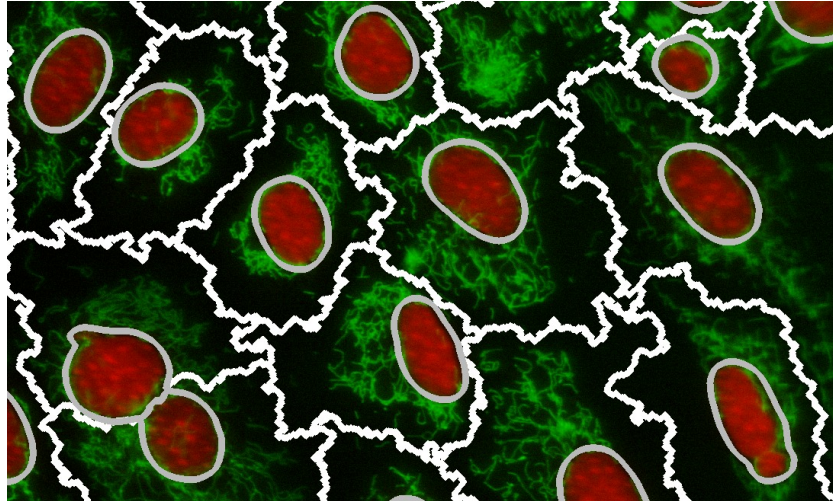


Fig. 1. Seeded Watershed. The nuclear borders (shown in grey) were first identified by a model-based algorithm [22] followed by watershed on the protein channel (resulting in the borders shown in white). Images have been contrast stretched for publication.

3 Supervised Classification

Many applications can be posed as a pattern recognition problem, i.e., given a set of examples of different classes, attempt to group other datapoints into these classes.

Given that even a small image has thousands of pixels, a direct pixel comparison is impossible. Furthermore, two images that differ only in a rigid body motion can have no common pixel values, but represent exactly the same cell state. Thus, the standard approach is to describe the image by a much smaller set of features, where a feature is a numeric function of the image. Once images have been summarized by this smaller set of features, machine learning methods can be applied to learn a classifier.

Features can be computed at several levels: directly from a field that may contain multiple cells, from cell regions (once the image has been segmented as described in the previous section), or from individual subcellular objects in the image. Certain feature classes, such as texture features, are applicable in other vision problems, but features designed specifically for this problem have also been presented (such as those that relate location of the protein of interest to the position of the nucleus) [13].

In some cases, a classifier outperforms a human expert in classification of protein location patterns. The results of one such experiment are shown in Figure 2, where the computer achieves 92% accuracy classifying ten location classes [23], while the human interpreter can only achieve 83% [24]. In another classification study, on the problem of cell detection, it was observed that computers perform

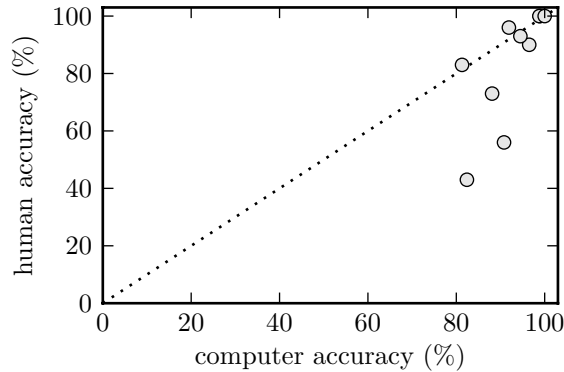


Fig. 2. A comparison of the accuracy in classification indicates that a supervised classification algorithm can perform as well or better than a human expert in recognizing subcellular patterns of ten proteins. Each circle represents one protein.

comparably to a medium-quality expert, but were still outperformed by an experienced expert. However, high variation between experts was observed [25].

4 Shape Analysis

Shapes of cells, nuclei, and organelles are critical for their function. For example, the shape of cells contribute to the overall tissue architecture that have characteristic tissue-specific functions [26]. In pathology, shape is also a useful indicator of deviations from the wild type phenotype. A classical example of a condition that can be identified by the shape of cells is sickle cell anemia, which results in deformed red blood cells. Automated analysis of nuclear shape has received the most attention given the importance of the nucleus in diagnosis.

One approach is to compute a small number of numerical features from the shape image that measure basic properties such as size or the convexity of the shape¹ [27,28].

This feature-based approach is simply an instance of the supervised classification framework as described in Section 3, but approaches that are specifically designed for shape have also been proposed. In particular, diffeomorphic methods have been applied in this area because the shape space of nuclei is non-linear [29]. By capturing how much deformation is needed to morph one shape into another, a distance function in the space of functions is defined. It is then possible to interpolate in the space of shapes [30] to generate intermediate shapes or place a collection of shapes in a low dimensional space in a way that preserves their distance relationships as well as possible. Figure 3 shows deformation distance as a starting nucleus shape is deformed to a target shape.

¹ In many applications, it is important to *remove* the influence of size by normalisation, so it depends on the application whether size (or, analogously, orientation) should be captured or normalized out.

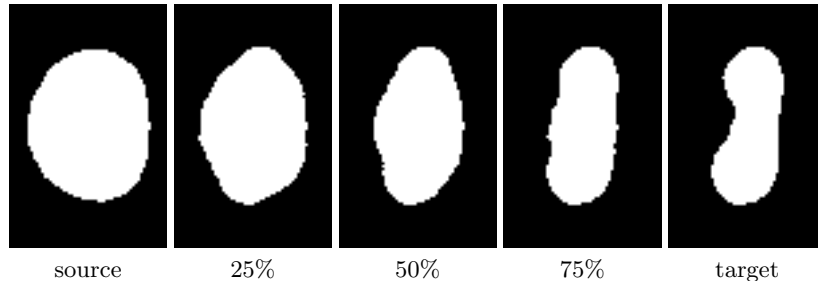


Fig. 3. The figure shows the transformation using a mapping function from one shape to another. The deformation distance of each image from the starting shape is shown as a percentage of total distance.

Analysis on cell and organelle shapes can be done in a manner similar to nuclear shape. A vast majority of methods use parametric representation of cell shapes [31,32,33].

5 Subcellular Pattern Unmixing

We have described methods to classify fluorescence images according to the depicted subcellular pattern using supervised classification approaches. These methods perform well for defined patterns, but cannot handle patterns that are composed of a mixture. For example, a protein which is partially in the plasma membrane and partially in the nucleus will exhibit a mixed pattern. A system whose model is of discrete assignments of patterns to classes will not be able to even represent the relationship between the *plasma membrane* and *nuclear patterns* and the mixed intermediates. In some situations, defining a mixed class as an extra class might be an acceptable, if inelegant, solution. However, not only does the number of classes grow combinatorially, but if a researcher is interested in quantifying the fraction of fluorescence in each compartment (for example, to study translocation as a function of time), then this solution will not be applicable.

Pattern unmixing directly models this situation. In the *supervised* case, the problem is as follows: the system is given examples of pure patterns and attempts to unmix mixed inputs (i.e., assign mixture fractions for each input condition). In the *unsupervised* form, the system is simply given a collection of images and must identify the fundamental patterns of which mixtures are made.

As one possible approach to this problem, we have described an object-based method for unmixing of subcellular patterns [34]. These methods work on the basis of identifying discrete objects in the image. An object is a contiguous set of pixels that differs in some way from its surroundings (in the simplest case, defined as being above a global threshold value). Patterns are now described by the properties of objects that they show. Mixed patterns will show objects that are characteristic of the basic patterns that compose them. Several variations of these methods have been proposed and this is still an active area of research.

To evaluate the quality of the proposed methods, the correlation between the input mixture coefficients and the underlying fraction of fluorescence in each compartment needs to be computed. In order to know the underlying fractions, Peng et al. [35] built a test set where two dyes that locate differently, but fluoresce similarly were imaged at different concentrations. For supervised unmixing, an 83% correlation with the underlying concentrations was obtained. For the unsupervised case, preliminary results of 72% correlation have been obtained [36].

6 Publicly Available Databases and Analysis of Large Datasets

In previous sections, we described the computational ways in which data is commonly processed. In this section, we give an overview of the sources of high-throughput image data that have been made available by researchers. Some of these have been the subject of automatic analysis, while others have not yet been processed in this way.

Besides the collection of images, these databases also provide annotations, typically manually assigned, describing the protein distribution within cells (see Table 1). The experimental techniques can be grouped into two major families of approaches: fusing the protein of interest to a fluorescence tag, or using fluorescent antibodies that bind to the protein of interest.

Table 1. Examples of microscopy image databases dedicated to subcellular protein location within cells and tissues

Database	Organism	Summary
PSLID-HeLa	H. Sapiens	10 protein distributions tagged by immunofluorescence, 100x magnification
LOCATE-HeLa	H. Sapiens	10 protein distributions tagged by immunofluorescence, 60x magnification
PSLID-RandTag	M. Musculus	ca. 2000 proteins with CD eGFP tagging, 40x magnification
Lifedb	C. Aethiops	ca. 1000 proteins tagged with N- and C-terminal EYFP and ECFP using 1500 cDNA whose vector is expressed in Vero cells, 63x magnification
YPL.db	S. Cerevisiae	371 GFP fusion proteins or vital staining (in 2005)
Yeastgfp	S. Cerevisiae	4,156 strains with detectable GFP signal among the 6,029 GFP positive strains, proteins tagged with GFP <i>fusion</i> at their N- and C-terminals, imaged at 100x magnification
HPA-IHC	H. Sapiens	5,000 proteins in 48 normal tissues, 20 cancer tissues, and 47 different cell lines labeled immunohistochemically (DAB and hematoxylin), imaged at 20x
HPA-IF	H. Sapiens	ca. 5,000 proteins in 3 different cell lines (A-431, U-2 OS, and U-251MG) tagged with immunofluorescence, imaged at 63x

In 2002, the first database dedicated to location proteomics, the Protein Subcellular location database (PSLID), was released. Initially composed of 2D single cell images, it has since incorporated various 2D and multicell image datasets. The most recently added dataset represents a collection of images generated by the CD-tagging protocol applied on NIH 3T3 cells [37]. PSLID and its associated software have a unique interface² which allows feature calculation, statistical comparison, clustering, classification and the creation of generative models [13].

Given its importance as a model organism, several large collection of yeast images are available. The Yeast Protein Localization database was released in 2002 and 2005 [38,39] (YPL.db²)³. The lack of information on image properties, such as the pixel resolution, limit the potential of automatic analysis of protein distributions in YPL.db². Concurrently, a collection of lines producing GFPfusion proteins for 75% of the open reading frames (ORFs) in *S. Cerevisiae* was released in 2003⁴ [40]. The assignment of each protein into one or a combination of 22 different locations was visually performed. Some fluorescent colocalization experiments with fluorescent markers were made to disambiguate uncertain cases. The same dataset was analyzed by a supervised machine learning approach that showed 81% to 95% accuracy in predicting the correct protein distribution [41]. The accuracy is calculated as the agreement between the labels found by the computer and the hand labeling. However, a closer observation of the mismatches revealed that errors were found in both approaches, human and computer.

The Database for Localization, Interaction, Functional assays and Expression of Proteins (LIFEdb) was created to collect microscopy images of fluorescent fusion proteins produced from human full-length cDNAs expressed in mammalian cell lines [42,43]⁵. The dataset was generated by creating fluorescent fusion proteins in Vero cells [44].

The Human Protein Atlas (HPA) project studies the human proteome in situ using antibodies. They collect images of proteins in stained tissues and cell lines in brightfield and fluorescence microscopies [45,46]. They aim to cover the entire human proteome (estimated to consist of ca. 20,000 non-redundant proteins) by 2014. The current release of the HPA database (version 5.0) contains images for more than 8000 antibodies targeting 7000 proteins [46]. The largest collection of images is produced by immunohistochemically (IHC) stained tissue microarrays from normal and cancerous biopsies. The database was extended with 47 different human cell lines commonly used in research and 12 primary blood cells. In addition, a collection of confocal microscopy images of 3 different human cell lines were tagged using immunofluorescence (IF) [47]. The protein distributions were assigned by visual inspection to three different compartments—nuclear, cytoplasm, membranous—for the IHC images. The higher resolution of IF images allows a finer distinction of the protein subcellular distribution into the major cellular organelles. Two machine learning based systems showed good results in generating

² Available at <http://pslid.cbi.cmu.edu/>

³ Available at <http://ypl.uni-graz.at/>

⁴ Available at <http://yeastgfp.yeastgenome.org/>

⁵ Available at <http://www.lifedb.de/>

labels on both image collections. The classification of the IF images gave 90% accuracy to distinguish 9 subcellular compartments [48] and 81% to distinguish 8 subcellular distribution from non-segmented IHC tissue images [49]. Preliminary results on identifying potential cancer biomarkers by automatically comparing protein distributions in normal and cancer images have also been presented [50].

7 Discussion

This review presented an overview of bioimage informatics, focusing on the problems in analysing fluorescence microscope images.

As a result of the continued progress of bioimage analysis, we increasingly observe domains where automated methods either outperform human analysis or are, at least, comparable with it. Thus, areas of disagreement between the two approaches cannot simply be marked as “classification error on the part of the algorithm.” Although humans still outperform computational approaches in general vision tasks by a large margin, the same is not necessarily true in the case of non-natural images of objects which humans do not encounter in their everyday lives (such as fluorescence microscopy images).

Bioimages also generate their own particular problems such as pattern unmixing or learning generative models. These are unique problems that still do not have a definite answer. Bioimage informatics still has many active research questions, in the development of entirely new methods to capture information in images, adaptation of existing ones from sister fields, or simply in solving the challenges of applying them to very large collections of data in real time with minimal user intervention. The need for such developments is underscored by the increase in publicly available data which is yet to be fully explored.

References

1. Perlman, Z.E., Slack, M.D., Feng, Y., Mitchison, T.J., Wu, L.F., Altschuler, S.J.: Multidimensional Drug Profiling By Automated Microscopy. *Science* 306(5699), 1194–1198 (2004)
2. Boland, M.V., Murphy, R.F.: A Neural Network Classifier Capable of Recognizing the Patterns of all Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells. *Bioinformatics* 17, 1213–1223 (2001)
3. Meijering, E., Smal, I., Danuser, G.: Tracking in molecular bioimaging. *IEEE Signal Processing Magazine* 23(3), 46–53 (2006)
4. Peng, H., Myers, E.W.: Comparing in situ mRNA expression patterns of drosophila embryos. In: 8th Intl. Conf. on Computational molecular biology, pp. 157–166 (2004)
5. Zhou, J., Peng, H.: Automatic recognition and annotation of gene expression patterns of fly embryos. *Bioinformatics* 23(5), 589–596 (2007)
6. Lécuyer, E., Tomancak, P.: Mapping the gene expression universe. *Current Opinion in Genetics & Development* 18(6), 506–512 (2008)
7. Boland, M.V., Murphy, R.F.: After Sequencing: Quantitative Analysis of Protein Localization. *IEEE Engineering in Medicine and Biology Magazine* 18(5), 115–119 (1999)

8. Chen, X., Murphy, R.F.: Objective Clustering of Proteins Based on Subcellular Location Patterns. *Journal Biomedical Biotechnology* 2005(2), 87–95 (2005)
9. Roques, E., Murphy, R.: Objective evaluation of differences in protein subcellular distribution. *Traffic* 3, 61–65 (2002)
10. Murphy, R.F.: Putting proteins on the map. *Nature Biotechnology* 24, 1223–1224 (2006)
11. Conrad, C., Erfle, H., Warnat, P., Daigle, N., Lörch, T., Ellenberg, J., Pepperkok, R., Eils, R.: Automatic Identification of Subcellular Phenotypes on Human Cell Arrays. *Genome Research* 14, 1130–1136 (2004)
12. Gasparri, F., Mariani, M., Sola, F., Galvani, A.: Quantification of the Proliferation Index of Human Dermal Fibroblast Cultures with the ArrayScan High-Content Screening Reader. *Journal of Biomolecular Screening* 9(3), 232–243 (2004)
13. Glory, E., Murphy, R.F.: Automated Subcellular Location Determination and High Throughput Microscopy. *Developmental Cell* 12(1), 7–16 (2007)
14. Hamilton, N.A., Pantelic, R.S., Hanson, K., Teasdale, R.D.: Fast automated cell phenotype image classification. *BMC Bioinformatics* 8, 110 (2007)
15. Huang, K., Lin, J., Gajnak, J., Murphy, R.F.: Image Content-based Retrieval and Automated Interpretation of Fluorescence Microscope Images via the Protein Subcellular Location Image Database. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 325–328 (2002)
16. Lein, E., Hawrylycz, M., Ao, N.: Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176 (2006)
17. Murphy, R.F.: Systematic description of subcellular location for integration with proteomics databases and systems biology modeling. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 1052–1055 (2007)
18. Nattkemper, T.W.: Automatic segmentation of digital micrographs: A survey. *Studies in health technology and informatics* 107(2), 847–851 (2004)
19. Coelho, L.P., Shariff, A., Murphy, R.F.: Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 518–521 (2009)
20. Jones, T.R., Carpenter, A.E., Golland, P.: Voronoi-based segmentation of cells on image manifolds. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) *CVBIA 2005*. LNCS, vol. 3765, pp. 535–543. Springer, Heidelberg (2005)
21. Beucher, S.: Watersheds of functions and picture segmentation. In: *IEEE Intl Conf. on Acoustics, Speech and Signal Processing*, Paris, pp. 1928–1931 (1982)
22. Lin, G., Adiga, U., Olson, K., Guzowski, J.F., Barnes, C.A., Roysam, B.: A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry Part A* 56A(1), 23–36 (2003)
23. Huang, K., Murphy, R.F.: Automated Classification of Subcellular Patterns in Multicell images without Segmentation into Single Cells. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 1139–1142 (2004)
24. Murphy, R., Velliste, M., Porreca, G.: Robust Numerical Features for Description and Classification of Subcellular Location Patterns in Fluorescence Microscope Images. *Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology* 35, 311–321 (2003)
25. Nattkemper, T.W., Twellmann, T., Schubert, W., Ritter, H.J.: Human vs. machine: Evaluation of fluorescence micrographs. *Computers in Biology and Medicine* 33(1), 31–43 (2003)
26. Allen, T.D., Potten, C.S.: Significance of cell shape in tissue architecture. *Nature* 264(5586), 545–547 (1976)

27. Olson, A.C., Larson, N.M., Heckman, C.A.: Classification of cultured mammalian cells by shape analysis and pattern recognition. *Proceedings of the National Academy of Sciences (USA)* 77(3), 1516–1520 (1980)
28. Pincus, Z., Theriot, J.A.: Comparison of quantitative methods for cell-shape analysis. *Journal of microscopy* 227, 140–156 (2007)
29. Rohde, G.K., Ribeiro, A.J.S., Dahl, K.N., Murphy, R.F.: Deformation-based nuclear morphometry: capturing nuclear shape variation in hela cells. *Cytometry Part A* 73A(4), 341–350 (2008)
30. Peng, T., Wang, W., Rohde, G.K., Murphy, R.F.: Instance-based generative biological shape modeling. In: *IEEE Intl. Symp. Biomedical Imaging*, vol. 1, pp. 690–693 (2009)
31. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
32. Albertini, M.C., Teodori, L., Piatti, E., Piacentini, M.P., Accorsi, A., Rocchi, M.B.L.: Automated analysis of morphometric parameters for accurate definition of erythrocyte cell shape. *Cytometry Part A* 52A(1), 12–18 (2003)
33. Lehmußola, A., Ruusuvoori, P., Selinummi, J., Huttunen, H., Yli-Harja, O.: Computational framework for simulating fluorescence microscope images with cell populations. *IEEE Trans. Medical Imaging* 26(7), 1010–1016 (2007)
34. Zhao, T., Velliste, M., Boland, M., Murphy, R.F.: Object type recognition for automated analysis of protein subcellular location. *IEEE Trans. on Image Processing* 14(9), 1351–1359 (2005)
35. Peng, T., Bonamy, G.M., Glory, E., Daniel Rines, S.K.C., Murphy, R.F.: Automated unmixing of subcellular patterns: Determining the distribution of probes between different subcellular locations. *Proceedings of the National Academy of Sciences, USA* (2009) (in press)
36. Coelho, L.P., Murphy, R.F.: Unsupervised unmixing of subcellular location patterns. In: *Proceedings of ICML-UAI-COLT 2009 Workshop on Automated Interpretation and Modeling of Cell Images (Cell Image Learning)*, Montreal, Canada (2009)
37. Garcia Osuna, E., Hua, J., Bateman, N.W., Zhao, T., Berget, P.B., Murphy, R.F.: Large-scale automated analysis of location patterns in randomly tagged 3T3 cells. *Annals of Biomedical Engineering* 35, 1081–1087 (2007)
38. Habeler, G., Natter, K., Thallinger, G.G., Crawford, M.E., Kohlwein, S.D., Trajanoski, Z.: YPL.db: the Yeast Protein Localization database. *Nucleic Acids Research* 30(1), 80–83 (2002)
39. Kals, M., Natter, K., Thallinger, G.G., Trajanoski, Z., Kohlwein, S.D.: Ypl.db²: the yeast protein localization database, version 2.0. *Yeast* 22(3), 213–218 (2005)
40. Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O’Shea, E.K.: Global analysis of protein localization in budding yeast. *Nature* 425(6959), 686–691 (2003)
41. Chen, S.C., Zhao, T., Gordon, G., Murphy, R.: Automated image analysis of protein localization in budding yeast. *Bioinformatics* 23(13), 66–71 (2007)
42. Bannasch, D., Mehrle, A., Glatting, K.H., Pepperkok, R., Poustka, A., Wiemann, S.: LIFEdb: a database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. *Nucleic Acids Research* 32, D505–D508 (2004)
43. del Val, C., Mehrle, A., Falkenhahn, M., Seiler, M., Glatting, K.H., Poustka, A., Suhai, S., Wiemann, S.: High-throughput protein analysis integrating bioinformatics and experimental assays. *Nucleic Acids Research* 32(2), 742–748 (2004)

44. Simpson, J., Wellenreuther, R., Poustka, A., Pepperkok, R., Wiemann, S.: Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO reports* 1(3), 287–292 (2000)
45. Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A.K., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M.G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skolleremo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L.L., Hober, S., Ponten, F.: A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. *Molecular & Cellular Proteomics* 4(12), 1920–1932 (2005)
46. Berglund, L., Björling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigyarto, C.A.K., Persson, A., Ottosson, J., Wernérus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Pontén, F., Uhlén, M.: A genecentric Human Protein Atlas for expression profiles based on antibodies. *Molecular & cellular proteomics* 7(10), 2019–2027 (2008)
47. Lundberg, E., Sundberg, M., Gräslund, T., Uhlén, M., Svahn, H.A.: A novel method for reproducible fluorescent labeling of small amounts of antibodies on solid phase. *Journal of Immunological Methods* 322(1-2), 40–49 (2007)
48. Newberg, J., Li, J., Rao, A., Ponten, F., Uhlen, M., Lundberg, E., Murphy, R.F.: Automated analysis of human protein atlas immunofluorescence images. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 1023–1026 (2009)
49. Newberg, J., Hua, J., Murphy, R.F.: Location Proteomics: Systematic Determination of Protein Subcellular Location. In: *Systems Biology*, vol. 500, pp. 313–332. Humana Press (2009)
50. Glory, E., Newberg, J., Murphy, R.F.: Automated comparison of protein subcellular location patterns between images of normal and cancerous tissues. In: *IEEE Intl. Symp. Biomedical Imaging*, pp. 304–307 (2008)