

Image Analysis with R: a Review

Stefan Rödiger, Hinrich Winther and Michał Burdukiewicz

Abstract Management, display, and processing of biological and medical imaging data is an important task in life sciences and medical research. R is a powerful cross-platform statistical computing environment. The aim of this mini-review is to give a brief overview about image processing software for the R statistical computing environment. When it comes to image analysis, R may appear to provide only few tools on the first sight. However, a systematic analysis of the existing packages shows that a huge potential for numerous applications.

Introduction

- Digital image processing?
- Where is it used? dffd

Scientists are using vast amount of images for qualitative data analysis. They originate from from technologies (e.g., fluorescence microscopy, microarrays) used for screening of multiple specimens of time or z-series data. For example, fluorescence microscopy data can be used to quantify the localization, localization and structure of a cell components dependent on time. Analysis of 2D and 3D digital images is a bridge technology which has been used to unravel quantitative and qualitative gene expression data (mRNA and protein level), cellular interactions and diagnostic data. In particular, immunofluorescence images were analyzed in relation to cell structures, tissues and organs (Chieco et al., 2013; Rödiger et al., 2013; Schierack et al., 2014; Willitzki et al., 2012). Other applications include fluorescence-lifetime measurements of cellular interactions (e.g., cell adhesion), Raman-imaging and more (Eliceiri et al., 2012; Schierack et al., 2014; Vogler et al., 2010). Computational digital image analysis (CDIA) is a complex process which can yield objective and quantitative measurements. It involves (1) instrument control, (2) data import and export, (3) data connectors, (4) data storage, (5) algorithms for image processing and analysis, (6) machine-learning, (7) statistical analysis and (8) report generation (Eliceiri et al., 2012). There are numerous software tools that have been made available for digital image acquisition and processing. It was shown that CDIA subtle differences within image specimens of multiple cellular phenotypes or cell cycle phases (Chieco et al., 2013; Eliceiri et al., 2012; Ljosa and Carpenter, 2009; Wiesmann et al., 2015). The scientific background of the users is deliberately broad. This includes biologists, biostatisticians, physicians and others, who fundamentally make use of the similar image analysis techniques. From discussion with peers we learned that knowledge of image processing is gained by self-study. In particular, learning of several programming languages may hamper the scientist to focus on their scientific aim.

R (R Core Team, 2012) is *de facto* the *lingua franca* of statistical bioinformatics and therefore used in numerous research disciplines (Rödiger et al., 2015a). It is a powerful tool for statistical data analysis. It comes to no surprise that software packages for digital image processing have been implemented (Frery and Perciano, 2013). There are numerous software packages for the analysis of image data (Wiesmann et al., 2015). However, R is quite functional when it comes to digital image analysis. In this review, we give an overview of the R ecosystem about which software packages exist and which deficits they may expose in comparison to other software packages. We aimed to aggregate information about R packages available on CRAN, Bioconductor (Gentleman et al., 2004), RForge or github. We performed two image processing case studies where we applied selected packages for (A) immunofluorescence image analysis and (B) RMI data.

Pixels and voxels are point samples on a grid pattern.

Multimodality imaging techniques such as CT and MRI or functional images from SPECT and PET (Martí-Bonmatí et al. (2010).

Bayesian Computation in CUDA on brain fMRI data (da Silva, 2010).

R version later than 2.11.0 have a graphics engine for rendering raster images via the functions `rasterImage()` and `grid.raster()` for better scaling of raster images, faster rendering to screen, and smaller graphics files (Murrell (2011).

Give me a title

General image processing and analysis (Ljosa and Carpenter, 2009)

An 3D spatial distribution analysis of biomarkers such as 53BP1, phosphorylated ATM, and γ H2AX is important for an image-based modeling of dynamic redistribution of DNA damage into

nuclear sub-domains (Costes et al., 2007). The biomolecules (proteins, RNA, DNA) distribution patterns within are complex. Patterns range from diffuse to punctate or microspeckled (Shiels et al., 2007; Willitzki et al., 2012). However, they all work in a coordinated and controlled manner within the nucleus (Shiels et al., 2007)

image processing capabilities of Cell-ID and data analysis by the statistical programming framework R for quantifying various cellular features (e.g., volume, total and subcellular fluorescence localization) from sets of microscope images of individual cells (Bush et al., 2012)

(Tabelow et al., 2012, 2014)

Murrel (Murrell, 2011) **mmand** (Clayden, 2016)

CRAN provides well established packages. These are **jpeg**, **png** and **tiff** to read, write and display bitmap JPEG, PNG and TIFF images. The development of the **ripa** (Frery and Perciano, 2013; Perciano and Frery, 2014) package was started in 2005 by Talita Perciano. This package can be used to processes and analyses RGB, LAN (multispectral) and AVIRIS (hyperspectral) images. Recent advances of **ripa** make it a promising tool for analysis of large datasets. The vast amount of image data is becoming more and more and essential part of Big Data analysis pipelines. R is among the frequently used for data mining and analysis. It comes to no surprise the commercial and non-commercial entities make heavy use of R (Chen et al., 2014). **EBImage** (Pau et al., 2010) is presumably the most comprehensive package and the foundation for many other R packages in the context of microscopy-based cellular assays (Gowen, 2015). This package offers tools to transform (e.g, rotate) the images, segment object (e.g., cells) and extract quantitative descriptors. The early version of **EBImage** used the Magick++ interface to the ImageMagick image processing library (Sklyar and Huber, 2006).

A recent package addition to this field of research is **imager** (Barthelme and Cecchi, 2016).

fields (Nychka et al., 2016)

General image processing and analysis

This section may contain a figure such as Figure 1.

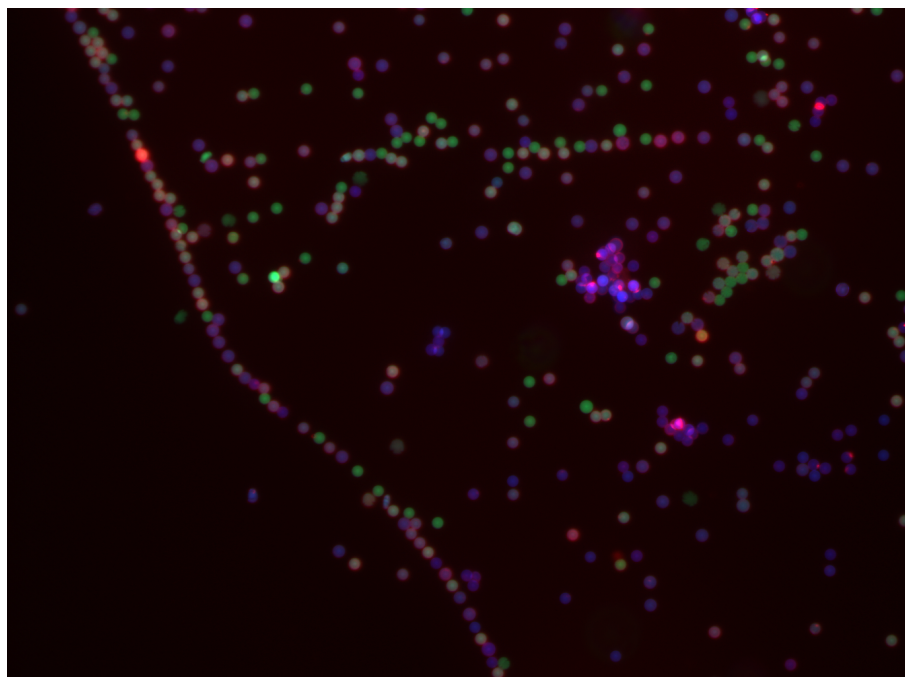


Figure 1: Image of beads.

segmentation

(Holmes et al., 2009)

adimpro is a package for manipulation of digital images and the Propagation Separation approach for smoothing digital images (Polzehl and Tabelow, 2007a). For example, image analysis is used for the

detection and quantification of cell patterns and array technologies like microarrays and bead-based assays (Dunning et al., 2006; Rödiger et al., 2013; Willitzki et al., 2012, 2013). Several software packages have been developed. `imageJ` belongs to the most used and cited tools. When it comes to R numerous packages exist, which can be readily integrated in the analysis routines (Schultze and Eggle, 2007; Frery and Perciano, 2013).

The accuracy of image segmentation is a critical step in a computer-aided diagnosis systems. The recognition of mitotic cells and the classification of fluorescent patterns is heavily dependent on this. Immunofluorescent images of cell, such as Hep-2, exhibit a high variability due a wide range of staining patterns and intensity levels (FIGURES OF CELLS), the presence of mitotic cells and artifacts. The later may be caused by uneven illumination and photo-bleaching effects (Tonti et al., 2015).

Intensity inhomogeneity (bias field) is a common artefact in magnetic resonance (MR) images, which hinders successful automatic segmentation. (Ivanovska et al., 2016)

Applications

Examples

Statistical analysis of functional magnetic resonance imaging data

Statistical analysis of functional magnetic resonance imaging (fMRI) data is a non-invasive neuroimaging technique. R has a impressive number of packages for such analysis (Tabelow et al., 2011). They are largely used in clinical routine and advanced brain research. There are various R packages that can be used to carry out analysis compare and displaying of fMRI data (Muschelli et al., 2014). `AnalyzeFMRI` (Bordier et al., 2011; Marchini, 2002) and `fmri` (Polzehl and Tabelow, 2007b) and are packages for the analysis of Magnetic Resonance Imaging (MRI) and functional Magnetic Resonance Imaging (fMRI) data, respectively. `PET` (Schulz et al., 2010)

For example this was used to compare data from from previous studies by a meta-analysis from the existing literature reults (Stocco, 2014).

Others include `dcemriS4` (Dunning et al., 2006; Frery and Perciano, 2013).

Analysis of fluorescence image data

`CRImage` package (Failmezger et al., 2012) for tumor image analysis

`gitter`, for quantification of colony sizes from plate images (Wagih and Parts, 2014)

Analysis of microbead assays

(Rödiger et al., 2013; Rödiger et al., 2014)

```
# Detailed installation instructions are available at
# http://bioconductor.org/packages/EBImage/
```

```
# Install the EBImage from Bioconductor
source("http://bioconductor.org/biocLite.R")
biocLite("EBImage")
```

```
# Load the EBImage package
require(EBImage)
```

Machine Learning, Big Data and Crowdsourced Data Preprocessing

CDIA can be used to classify cells in microscopy images automatically by machine learning for image-based screening. Sommer and Gerlich (2013) reviewed how images can be converted into a data representation for machine learning (ML). ML functionality is present in open source packages such as CellProfiler (Conrad et al., 2011; Sommer and Gerlich, 2013). CDIA R packages do not come with functionality. However, the R programming language has tools for signal processing, statistical modeling, machine learning, regression, classification, variable selection and data visualization (Abbas et al., 2014; Fuchs et al., 2010; Pau et al., 2010; Genuer et al., 2015).

The term “Big Data” was coined in 1997 by the NASA scientists Michael Cox and David Ellsworth and later adopted by other disciplines (Cox and Ellsworth, 1997). Modern biomedical research laboratories produce world-wide Big Data from multi-omics platforms, biomedical imaging platforms and electronic medical records. There is, however, no strict consensus on the term Big Data. Big Data is defined by the 4Vs. These are: (i) Volume (heterogeneous data types (nominal, ordinal, categorical)), (ii) Variety (complex structures), (iii) Value and (iv) Velocity of input/output data. Some authors argue that Big Data refers to all data where standard strategies for data storage, search, visualization and distribution no longer apply. Hypothesis generation is an important aspect in medical research. The large datasets of Big Data may be interpreted to reveal hidden patterns or find novel biomarkers and correlations, as described elsewhere (Willis and Lord, 2015). Applications of R in the context of parallel architectures and big data have been described (Das et al., 2010; Ostrouchov et al., 2012; Gandrud, 2013; Qian, 2014).

example for ML and CDIA?

Clustering is a well established and widely-used technique used in machine learning, data mining, character recognition and information retrieval. The aim is to group similar objects and separate dis-similar objects based on selected features (Szkaliczki, 2016; Hu and Qutub, 2016). Commonly used algorithms, such as k-means or hierarchical clustering, require a priori specified cluster numbers which is often not known. Therefore clustering evaluation techniques were proposed. For example, Progeny Clustering can be used to estimate the optimal number for clustering to find the cell clusters based on their imaging data (Hu and Qutub, 2016).

Expert rating of images is commonly used in several stages. For example, if new patterns need to be classified or in cases where ambiguous patterns need a consensus of the expert opinion. Biomedical research vast amounts of image data expose the scientists to a high workload. Leeper (2016) recently described the R package **MTurkR** for the Amazon Mechanical Turk (MTurk) crowdsourcing platform, which is focused on preprocessing of data for immediate use in R. In his example he used crowdsourced human intelligence for preprocessing massive “messy” data into a structured form. The image rating task involved 225 crowdsourced workers and more than 5500 images.

Conventional classification algorithms operate with datasets which have a set of independent variables (predictor), and only one variable to be predicted. However, for many biomedical image data sets classifier has to work with several outputs. The classification of picture can have a set of labels. For example, pictures can be assigned landscape, sky and forest. **mldr** package (Charte and Charte, 2015) was developed for exploratory analysis, transformation and manipulation of multilabel datasets.

Graphical User Interface for Digital Image Analysis

There exist also R graphical user interfaces (Rödiger et al., 2012) which can be used for image processing. Bio7 is an integrated development environment based on the Eclipse Rich Client Platform (RCP). The main purpose of this tool is the modeling and analysis of ecological systems. However, Bio7 is not restricted to this discipline. The application contains GUIs and plugins for simulation and analysis tasks. Interestingly, one of these plugins is an adaption image application ImageJ and another is available for a bidirectional Java connection to R. This means that data can be transferred from and to ImageJ and R.

Performance

Michał, would you like to take this section?

Requirements for recent research include the rapid processing of massive amounts of image data (Mega to Tera byte scale) that modern technologies (e.g., microscopes, MRI scanner) produce nowadays. Preferably, affordable personal desktop computers should be usable. R has several disadvantages when it comes to memory management and GPU and CPU usage ...

Orphaned R packages

In the past there have been of R packages developed which are no longer maintained. **rimage**¹ provided functions for image processing, such as sobel filter, rank filters, fft and histogram equalization. **edci**² offered edge detection based on the difference of two asymmetric M kernel estimators and

¹<https://cran.r-project.org/src/contrib/Archive/rimage/>

²<https://cran.r-project.org/src/contrib/Archive/edci/>

regression clustering (linear, circular) based on redescending M estimators. **epsi**³ offered edge and corner preserving smoothing methods for images which are based on a redescending M kernel estimator. **biOps**⁴ included functions for geometric, arithmetic, logic, morphologic operations, look-up tables, edge detection (e.g., Roberts, Sobel, Kirsch, Marr-Hildreth and Canny), convolution masks operations, fft, Isodata and k-means classification methods (standard, kd-tree, brute force methods). **RImageJ**⁵ offered R bindings for ImageJ through **rJava** to read and processes various image formats.

Summary

Image analysis within one software framework can help ensure that results are accurate, objective and reproducible. Many scientist are using R. However, it appears that only few make use of the image analysis tools currently available. We would like to raise awareness for the fact that R provides sophisticated packages for digital image analysis. The central advantage is that all analysis is conducted within the same environment across most platforms. Added values for the user are that there is less need to learn a new programming languages and that all analysis can be performed in a consistent and cross-platform environment. Table 1 gives an overview of R packages currently available. We found that functions from the packages can be easily combined to conduct manipulations and analysis (e.g., object transformations, measurements, object counting, gray level statistics, binarization) at advanced levels. The organization of such functions in customized packages is straightforward even for more complex analysis functions.

(Rödiger et al., 2015b) or combined assays of microbeads and cells (Grossmann et al., 2016; Scholz et al., 2015)

For testing purposes it is possible to make use of public image repositories as listed by Eliceiri et al. (2012).

Automated image acquisition systems enable microscopy experiments that generate large image datasets. Therefore it is important to have robust, objective and quantitative image analysis systems. Since most image analysis systems are developed for a specific type of experiment, cell type or imaging technology it is challenging to find one tool that fits all needs by manual adaption of all parameters for a specific image analysis task. Pattern recognition is a useful tool in this context. This can be achieved by trained algorithms as described by Shamir et al. (2010).

While evaluation the packages we ended up with experiences regarding the installation and maintenance. Basically all installation were well instructed. For example, the installation of sophisticated packages like **EBImage** is well documented. However, many packages depend on a number of third-party packages. This applies also to **EBImage** which requires tiff and fftwtools. Similar applies to **imager** which required the installation of fftw3 for computing Fast Fourier Transforms.

Others and we recommend to make use of packages for reproducible research (Rödiger et al., 2015a).

There exist several GUI technologies for R which make it possible to integrate R code into an easy to master point-and-click interface.

Bibliography

- S. S. Abbas, T. M. Dijkstra, and T. Heskes. A comparative study of cell classifiers for image-based high-throughput screening. *BMC Bioinformatics*, 15:342, 2014. ISSN 1471-2105. doi: 10.1186/1471-2105-15-342. [p3]
- S. Barthelme and A. Cecchi. **imager**: Image Processing Library Based on CImg, Jan. 2016. URL <https://cran.r-project.org/web/packages/imager/index.html>. [p2]
- C. Bordier, M. Dojat, and P. L. De Micheaux. Temporal and spatial independent component analysis for fMRI data sets embedded in the AnalyzeFMRI R package. *Journal of Statistical Software*, 44(9): 1–24, 2011. URL <https://hal.archives-ouvertes.fr/inserm-00659425/>. [p3]
- A. Bush, A. Chernomoretz, R. Yu, A. Gordon, and A. Colman-Lerner. Using Cell-ID 1.4 with R for Microscope-Based Cytometry. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, CHAPTER:Unit14.18, Oct. 2012. ISSN 1934-3639. doi: 10.1002/0471142727.mb1418s100. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3485637/>. [p2]

³<https://cran.r-project.org/src/contrib/Archive/epsi/>

⁴<https://cran.r-project.org/src/contrib/Archive/biOps/>

⁵<https://cran.r-project.org/src/contrib/Archive/RImageJ/>

Package	Main function	Comment	Source
adimpro	×	×	×
AnalyzeFMRI	×	×	×
CRImage	×	×	Bioconductor
dcemriS4	×	×	CRAN
dpmixsim	×	×	CRAN
EBImage	fancy stuff	well maintained	Bioconductor
gitter	×	based on EBImage	CRAN
imageHTS	fancy stuff	- d	Bioconductor
imager	×	×	CRAN
jpeg	×	×	CRAN
PET	×	×	CRAN
pixmap	×	×	CRAN
png	×	×	CRAN
ripa	×	×	×
RNiftyReg	×	×	CRAN
tiff	×	×	CRAN
rtiff	reads and writes TIFF format images	×	CRAN
videoTools	video and image analysis	-	RForge
webp	reads and writes webp format images	×	CRAN

Table 1: R packages. *videoTools*: <http://www.rforge.net/videoTools/files/>. **pixmap** can read and write bitmap images in PBM (black/white), PGM (grey) and PPM (color) format.

- F. CharTE and D. CharTE. Working with Multilabel Datasets in R: The mldr Package. *The R Journal*, 7(2): 149–162, Dec. 2015. URL <http://journal.r-project.org/archive/2015-2/charTE-charTE.pdf>. [p4]
- M. Chen, S. Mao, and Y. Liu. Big Data: A Survey. *Mobile Networks and Applications*, 19(2):171–209, Jan. 2014. ISSN 1383-469X, 1572-8153. doi: 10.1007/s11036-013-0489-0. [p2]
- P. Chieco, A. Jonker, B. A. De Boer, J. M. Ruijter, and C. J. F. Van Noorden. Image Cytometry: Protocols for 2D and 3D Quantification in Microscopic Images. *Progress in Histochemistry and Cytochemistry*, 47(4):211–333, Jan. 2013. ISSN 0079-6336. doi: 10.1016/j.proghi.2012.09.001. URL <http://www.sciencedirect.com/science/article/pii/S007963361200037X>. [p1]
- J. Clayden. *mmmand: Mathematical Morphology in Any Number of Dimensions*. 2016. URL <https://CRAN.R-project.org/package=mmmand>. R package version 1.3.0. [p2]
- C. Conrad, A. Wünsche, T. H. Tan, J. Bulkescher, F. Sieckmann, F. Verissimo, A. Edelstein, T. Walter, U. Liebel, R. Pepperkok, and J. Ellenberg. Micropilot: automation of fluorescence microscopy-based imaging for systems biology. *Nature Methods*, 8(3):246–249, Mar. 2011. ISSN 1548-7091. doi: 10.1038/nmeth.1558. URL <http://www.nature.com/nmeth/journal/v8/n3/abs/nmeth.1558.html>. [p3]
- S. V. Costes, A. Ponomarev, J. L. Chen, D. Nguyen, F. A. Cucinotta, and M. H. Barcellos-Hoff. Image-Based Modeling Reveals Dynamic Redistribution of DNA Damage into Nuclear Sub-Domains. *PLoS Computational Biology*, 3(8):e155, Aug. 2007. doi: 10.1371/journal.pcbi.0030155. [p2]
- M. Cox and D. Ellsworth. Application-controlled demand paging for out-of-core visualization. In *Proceedings of the 8th conference on Visualization'97*, pages 235–ff. IEEE Computer Society Press, 1997. URL <http://dl.acm.org/citation.cfm?id=267068>. [p4]
- A. F. da Silva. cudaBayesreg: Bayesian computation in CUDA. *The R Journal*, 2(2):48–55, 2010. URL https://www.researchgate.net/profile/Adelino_Silva2/publication/265483496_cudaBayesreg_Bayesian_Computation_in_CUDA/links/559e6fd908aeab53f8fd2760.pdf. [p1]
- S. Das, Y. Sismanis, K. S. Beyer, R. Gemulla, P. J. Haas, and J. McPherson. Ricardo: Integrating R and Hadoop. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 987–998, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0032-2. doi: 10.1145/1807167.1807275. URL <http://doi.acm.org/10.1145/1807167.1807275>. 00115. [p4]
- M. Dunning, M. Smith, N. Thorne, and S. Tavaré. beadarray: An R Package to Analyse Illumina BeadArrays. *R News*, 6(5):17–23, Dec. 2006. URL <http://CRAN.R-project.org/doc/Rnews/>. [p3]

- K. W. Eliceiri, M. R. Berthold, I. G. Goldberg, L. Ibáñez, B. S. Manjunath, M. E. Martone, R. F. Murphy, H. Peng, A. L. Plant, B. Roysam, N. Stuurman, J. R. Swedlow, P. Tomancak, and A. E. Carpenter. Biological imaging software tools. *Nature Methods*, 9(7):697–710, July 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2084. URL <http://www.nature.com/nmeth/journal/v9/n7/abs/nmeth.2084.html>. [p1, 5]
- H. Failmezger, Y. Yuan, O. Rueda, and F. Markowetz. *CRImage: CRImage a package to classify cells and calculate tumour cellularity*. 2012. R package version 1.18.0. [p3]
- A. C. Frery and T. Perciano. *Introduction to Image Processing Using R*. SpringerBriefs in Computer Science. Springer London, London, 2013. ISBN 978-1-4471-4949-1 978-1-4471-4950-7. doi: 10.1007/978-1-4471-4950-7. [p1, 2, 3]
- F. Fuchs, G. Pau, D. Kranz, O. Sklyar, C. Budjan, S. Steinbrink, T. Horn, A. Pedal, W. Huber, and M. Boutros. Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Molecular Systems Biology*, 6, June 2010. ISSN 1744-4292. doi: 10.1038/msb.2010.25. [p3]
- C. Gandrud. *Reproducible Research with R and RStudio*. Chapman and Hall/CRC, July 2013. ISBN 978-1-4665-7284-3. [p4]
- R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, and J. Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80, Sept. 2004. ISSN 1465-6906. doi: 10.1186/gb-2004-5-10-r80. URL <http://genomebiology.com/2004/5/10/R80/abstract>. 07026. [p1]
- R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. VSURF: An R package for variable selection using random forests. *The R Journal*, 7(2):19–33, Dec. 2015. URL <http://journal.r-project.org/archive/2015-2/genuer-poggi-tuleaumalot.pdf>. [p3]
- A. A. Gowen. Near infrared hyperspectral image analysis using R. Part 6: image processing using EBIImage. *NIR news*, 26(3):20–21, 2015. ISSN 0960-3360. doi: 10.1255/nirn.1526. [p2]
- K. Grossmann, N. Röber, R. Hiemann, S. Rödiger, P. Schierack, D. Reinhold, M. W. Laass, K. Conrad, and D. Roggenbuck. Simultaneous detection of celiac disease-specific IgA antibodies and total IgA. *Autoimmunity Highlights*, 7(1):1–10, Jan. 2016. ISSN 2038-0305, 2038-3274. doi: 10.1007/s13317-016-0073-2. [p5]
- S. Holmes, A. Kapelner, and P. P. Lee. An Interactive Java Statistical Image Segmentation System: GemIdent. *Journal of Statistical Software*, 30(10), June 2009. ISSN 1548-7660. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3100170/>. [p2]
- C. W. Hu and A. A. Qutub. progenyClust: an R package for Progeny Clustering. *The R Journal*, 8(1), 2016. ISSN 2073-4859. URL <https://journal.r-project.org/archive/accepted/hu-qutub.pdf>. [p4]
- T. Ivanovska, R. Laqua, L. Wang, A. Schenk, J. H. Yoon, K. Hegenscheid, H. Völzke, and V. Liebscher. An efficient level set method for simultaneous intensity inhomogeneity correction and segmentation of MR images. *Computerized Medical Imaging and Graphics*, 48:9–20, Mar. 2016. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2015.11.005. URL <http://www.sciencedirect.com/science/article/pii/S0895611115001779>. [p3]
- T. J. Leeper. Crowdsourced Data Preprocessing with R and Amazon Mechanical Turk. *The R Journal*, 8(1):NN, 2016. ISSN 2073-4859. URL <https://journal.r-project.org/archive/accepted/leeper.pdf>. [p4]
- V. Ljosa and A. E. Carpenter. Introduction to the Quantitative Analysis of Two-Dimensional Fluorescence Microscopy Images for Cell-Based Screening. *PLoS Computational Biology*, 5(12), Dec. 2009. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1000603. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2791844/>. [p1]
- J. Marchini. AnalyzefMRI: An R package for the exploration and analysis of MRI and fMRI datasets. *R News*, 2(1):17–23, Mar. 2002. URL <http://www.r-project.org/doc/2002-03/NEWS.pdf>. [p3]
- L. Martí-Bonmatí, R. Sopena, P. Bartumeus, and P. Sopena. Multimodality imaging techniques. *Contrast Media & Molecular Imaging*, 5(4):180–189, July 2010. ISSN 1555-4309. doi: 10.1002/cmmi.393. URL <http://doi.wiley.com/10.1002/cmmi.393>. [p1]

- P. Murrell. Raster Images in R Graphics. *The R Journal*, 3(1):48–54, June 2011. URL http://journal.r-project.org/archive/2011-1/RJournal_2011-1_Murrell.pdf. [p1, 2]
- J. Muschelli, E. Sweeney, and C. Crainiceanu. brainR: Interactive 3 and 4D images of high resolution neuroimage data. *The R Journal*, 6(1):41–48, June 2014. URL <http://journal.r-project.org/archive/2014-1/muschelli-sweeney-crainiceanu.pdf>. [p3]
- D. Nychka, R. Furrer, J. Paige, and S. Sain. fields: Tools for Spatial Data, Jan. 2016. URL <https://cran.r-project.org/web/packages/fields/index.html>. [p2]
- G. Ostrouchov, W.-C. Chen, D. Schmidt, and P. Patel. Programming with Big Data in R, 2012. URL <http://r-pbd.org/>. [p4]
- G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. EBIImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7):979–981, 2010. doi: 10.1093/bioinformatics/btq046. [p2, 3]
- T. Perciano and A. C. Frery. ripa: R Image Processing and Analysis, May 2014. URL <https://cran.r-project.org/web/packages/ripa/index.html>. [p2]
- J. Polzehl and K. Tabelow. Adaptive Smoothing of Digital Images: The R Package adimpro. *Journal of Statistical Software*, 19(1):1–17, 2007a. ISSN 1548-7660. URL <http://www.jstatsoft.org/index.php/jss/article/view/v019i01>. [p2]
- J. Polzehl and K. Tabelow. fmri: A Package for Analyzing fmri Data. *R News*, 7(2):13–17, Oct. 2007b. URL <http://www.r-project.org/doc/2007-10/rnews/fmri.pdf>. [p3]
- H. Qian. Pivotalr: A package for machine learning on big data. *The R Journal*, 6(1):57–67, 2014. URL <http://mirror.bjtu.edu.cn/cran/web/packages/PivotalR/vignettes/pivotalr.pdf>. [p4]
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL <http://www.R-project.org/>. ISBN 3-900051-07-0. [p1]
- S. Rödiger, T. Friedrichsmeier, P. Kapat, and M. Michalke. RKWard: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R. *Journal of Statistical Software*, 49(9):1–34, 2012. ISSN 1548-7660. URL <http://www.jstatsoft.org/v49/i09>. [p4]
- S. Rödiger, P. Schierack, A. Böhm, J. Nitschke, I. Berger, U. Frömmel, C. Schmidt, M. Ruhland, I. Schimke, D. Roggenbuck, W. Lehmann, and C. Schröder. A highly versatile microscope imaging technology platform for the multiplex real-time detection of biomolecules and autoimmune antibodies. *Advances in Biochemical Engineering/Biotechnology*, 133:35–74, 2013. ISSN 0724-6145. doi: 10.1007/10_2011_132. [p1, 3]
- S. Rödiger, C. Liebsch, C. Schmidt, W. Lehmann, U. Resch-Genger, U. Schedler, and P. Schierack. Nucleic acid detection based on the use of microbeads: a review. *Microchimica Acta*, 181(11-12): 1151–1168, Aug. 2014. ISSN 0026-3672, 1436-5073. doi: 10.1007/s00604-014-1243-4. [p3]
- S. Rödiger, M. Burdukiewicz, K. A. Blagodatskikh, and P. Schierack. R as an Environment for the Reproducible Analysis of DNA Amplification Experiments. *The R Journal*, 7(2):127–150, 2015a. URL <http://journal.r-project.org/archive/2015-1/RJ-2015-1.pdf>. 00000. [p1, 5]
- S. Rödiger, T. Kramer, U. Frömmel, J. Weinreich, D. Roggenbuck, S. Guenther, K. Schaufler, C. Schröder, and P. Schierack. Intestinal *Escherichia coli* colonization in a mallard duck population over four consecutive winter seasons. *Environmental Microbiology*, pages 1–10, 2015b. ISSN 1462-2920. doi: 10.1111/1462-2920.12807. 00000. [p5]
- P. Schierack, S. Rödiger, R. Kolenda, R. Hiemann, E. Berger, K. Grzymajło, A. Swidsinski, T. Juretzek, D. Meissner, K. Mydlak, D. Reinhold, L. K. Nolan, and D. Roggenbuck. Species-specific and pathotype-specific binding of bacteria to zymogen granule membrane glycoprotein 2 (GP2). *Gut*, pages gutjnl-2014-307854, July 2014. ISSN , 1468-3288. doi: 10.1136/gutjnl-2014-307854. URL <http://gut.bmj.com/content/early/2014/07/29/gutjnl-2014-307854>. 00003. [p1]
- J. Scholz, K. Grossmann, I. Knütter, R. Hiemann, M. Sowa, N. Röber, S. Rödiger, P. Schierack, D. Reinhold, D. P. Bogdanos, P. L. Meroni, A. Radice, K. Conrad, and D. Roggenbuck. Second generation analysis of antinuclear antibody (ANA) by combination of screening and confirmatory testing. *Clinical Chemistry and Laboratory Medicine (CCLM)*, 0(0):1–12, May 2015. ISSN 1437-4331. doi: 10.1515/cclm-2015-0083. URL <http://www.degruyter.com/view/j/cclm.ahead-of-print/cclm-2015-0083/cclm-2015-0083.xml>. [p5]

- J. L. Schultze and D. Eggle. IlluminaGUI: Graphical User Interface for analyzing gene expression data generated on the Illumina platform. *Bioinformatics*, 23(11):1431–1433, June 2007. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btm101. URL <http://bioinformatics.oxfordjournals.org/content/23/11/1431>. [p3]
- J. Schulz, P. Toft, J. J. Jensen, and P. Philipsen. PET: Simulation and Reconstruction of PET Images, Aug. 2010. URL <https://cran.r-project.org/web/packages/PET/index.html>. [p3]
- L. Shamir, J. D. Delaney, N. Orlov, D. M. Eckley, and I. G. Goldberg. Pattern Recognition Software and Techniques for Biological Image Analysis. *PLoS Computational Biology*, 6(11):e1000974, Nov. 2010. doi: 10.1371/journal.pcbi.1000974. [p5]
- C. Shiels, N. M. Adams, S. A. Islam, D. A. Stephens, and P. S. Freemont. Quantitative Analysis of Cell Nucleus Organisation. *PLoS Computational Biology*, 3(7):e138, July 2007. doi: 10.1371/journal.pcbi.0030138. [p2]
- O. Sklyar and W. Huber. Image Analysis for Microscopy Screens. *R News*, 6(5):12–16, Dec. 2006. URL <http://CRAN.R-project.org/doc/Rnews/>. [p2]
- C. Sommer and D. W. Gerlich. Machine learning in cell biology – teaching computers to recognize phenotypes. *J Cell Sci*, 126(24):5529–5539, Dec. 2013. ISSN 0021-9533, 1477-9137. doi: 10.1242/jcs.123604. URL <http://jcs.biologists.org/content/126/24/5529>. [p3]
- A. Stocco. Coordinate-Based Meta-Analysis of fMRI Studies with R. *The R Journal*, 6(2):5–15, Dec. 2014. URL <http://journal.r-project.org/archive/2014-2/stocco.pdf>. [p3]
- T. Szkaliczki. clustering.sc.dp: Optimal Clustering with Sequential Constraint by Using Dynamic Programming. *The R Journal*, 8(1), 2016. ISSN 2073-4859. URL <https://journal.r-project.org/archive/accepted/szkaliczki.pdf>. [p4]
- K. Tabelow, B. Whitcher, and others. Special volume on magnetic resonance imaging in R. *Journal of Statistical Software*, 44(1):1–6, 2011. URL <http://www.jstatsoft.org/htaccess.php?volume=44&type=i&issue=01&paper=true>. [p3]
- K. Tabelow, H. U. Voss, and J. Polzehl. Modeling the orientation distribution function by mixtures of angular central Gaussian distributions. *Journal of Neuroscience Methods*, 203(1):200–211, Jan. 2012. ISSN 0165-0270. doi: 10.1016/j.jneumeth.2011.09.001. URL <http://www.sciencedirect.com/science/article/pii/S0165027011005231>. [p2]
- K. Tabelow, J. Polzehl, and F. Anker. dti: Analysis of diffusion weighted imaging (DWI) data, Dec. 2014. URL <https://cran.r-project.org/web/packages/dti/index.html>. [p2]
- S. Tonti, S. D. Cataldo, A. Bottino, and E. Ficarra. An automated approach to the segmentation of HEp-2 cells for the indirect immunofluorescence ANA test. *Computerized Medical Imaging and Graphics*, 40:62–69, Mar. 2015. ISSN 0895-6111. doi: 10.1016/j.compmedimag.2014.12.005. URL <http://www.medicalimagingandgraphics.com/article/S089561111400202X/abstract>. [p3]
- N. Vogler, T. Bocklitz, M. Mariani, V. Deckert, A. Markova, P. Schelkens, P. Rösch, D. Akimov, B. Dietzek, and J. Popp. Separation of CARS image contributions with a Gaussian mixture model. *JOSA A*, 27(6):1361–1371, June 2010. ISSN 1520-8532. doi: 10.1364/JOSAA.27.001361. URL <http://www.osapublishing.org/abstract.cfm?uri=josaa-27-6-1361>. [p1]
- O. Wagih and L. Parts. gitter: A Robust and Accurate Method for Quantification of Colony Sizes From Plate Images. *G3: Genes | Genomes | Genetics*, 4(3):547–552, Mar. 2014. ISSN , 2160-1836. doi: 10.1534/g3.113.009431. URL <http://www.g3journal.org/content/4/3/547>. [p3]
- V. Wiesmann, D. Franz, C. Held, C. Münzenmayer, R. Palmisano, and T. Wittenberg. Review of free software tools for image analysis of fluorescence cell micrographs. *Journal of Microscopy*, 257(1): 39–53, Jan. 2015. ISSN 1365-2818. doi: 10.1111/jmi.12184. [p1]
- J. C. D. Willis and G. M. Lord. Immune biomarkers: the promises and pitfalls of personalized medicine. *Nature Reviews Immunology*, 15(5):323–329, 2015. ISSN 1474-1733. doi: 10.1038/nri3820. URL <http://www.nature.com/nri/journal/v15/n5/abs/nri3820.html>. [p4]
- A. Willitzki, R. Hiemann, V. Peters, U. Sack, P. Schierack, S. Rödiger, U. Anderer, K. Conrad, D. P. Bogdanos, D. Reinhold, and D. Roggenbuck. New platform technology for comprehensive serological diagnostics of autoimmune diseases. *Clinical & developmental immunology*, 2012:284740, 2012. ISSN 1740-2530. doi: 10.1155/2012/284740. [p1, 2, 3]

A. Willitzki, S. Lorenz, R. Hiemann, K. Guttek, A. Goihl, R. Hartig, K. Conrad, E. Feist, U. Sack, P. Schierack, L. Heiserich, C. Eberle, V. Peters, D. Roggenbuck, and D. Reinhold. Fully automated analysis of chemically induced γ H2AX foci in human peripheral blood mononuclear cells by indirect immunofluorescence. *Cytometry. Part A: the journal of the International Society for Analytical Cytology*, 83(11):1017–1026, Nov. 2013. ISSN 1552-4930. doi: 10.1002/cyto.a.22350. [p3]

Stefan Rödiger (corresponding author)
orcid.org/0000-0002-1441-6512
Faculty of Natural Sciences
Brandenburg University of Technology Cottbus–Senftenberg
Senftenberg
Germany
Stefan.Roediger@b-tu.de

Hinrich Winther
Affiliation
Address
Country
author@work

Michał Burdukiewicz
University of Wrocław
Faculty of Biotechnology
Department of Genomics
Wrocław
Poland michalburdukiewicz@gmail.com