Methods 102 (2016) 20–25

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Bioimaging for quantitative phenotype analysis

Weiyang Chen, Xian Xia, Yi Huang, Xingwei Chen, Jing-Dong J. Han *

Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 27 December 2015
Accepted 6 January 2016
Available online 2 February 2016

Keywords:
Bioimaging
Phenotype
Quantitative analysis

ABSTRACT

With the development of bio-imaging techniques, an increasing number of studies apply these techniques to generate a myriad of image data. Its applications range from quantification of cellular, tissue, organismal and behavioral phenotypes of model organisms, to human facial phenotypes. The bio-imaging approaches to automatically detect, quantify, and profile phenotypic changes related to specific biological questions open new doors to studying phenotype-genotype associations and to precisely evaluating molecular changes associated with quantitative phenotypes. Here, we review major applications of bioimage-based quantitative phenotype analysis. Specifically, we describe the biological questions and experimental needs addressable by these analyses, computational techniques and tools that are available in these contexts, and the new perspectives on phenotype-genotype association uncovered by such analyses.

Contents

The development of imaging technology and quantitative imaging analysis provide a promising new perspective to examine biological problems. Quantitative phenotype analysis based on imaging analysis is an important area of bioimage informatics [1]. Its applications range from quantification of cellular [2–4], tissue and organismal phenotypes to behavior analysis of model organisms [5–8], which will be summarized below.

1. Cellular phenotypes

Image-based assays have the advantage of unbiasedly and quantitatively observing and recording experimental results. Multiple fluorescent labels and morphology phenotypes can be quantified at the same time by imaging techniques and analysis methods [4].

Combining RNA interference with image analysis is one of the attempts to uncover the molecular underpinnings of cellular phenotypes. Neumann et al. developed a phenotypic screening platform to identify genes required for basic functions of life [9,10]. For cells processed by RNA interference of each of the 21,000 human protein-coding genes, they took real-time images by 2 days live imaging of fluorescently labeled chromosomes. Then they used their image processing package to automatically identify each nucleus and quantified ~200 features, such as gray level and shape features, for each nucleus. They classified each nucleus into one of 16 morphological classes by support vector machines [11,12] based on quantified features. These classes, describing chromosome configurations, include prometaphase, binuclear (two

* Corresponding author.
E-mail address: jdhan@picb.ac.cn (J.-D.J. Han).

http://dx.doi.org/10.1016/j.ymeth.2016.01.017
1046-2023/

nuclei), grape (many micronuclei) and polylobed (multilobed nuclei), etc. They then used the percentage of nuclei in each morphological class to quantify and describe the response to RNAi of each gene. Therefore, every time-lapse experiment can be quantified as a time series of percentage values. By testing the significance of deviation between each gene and control, and clustering of the genes' time-resolved phenotypic profiles, they identified hundreds of genes involved in basic cellular functions, such as cell division, migration and survival. Kirsanova et al. developed a cellular phenotype database to manage phenotype data automatically or manually quantified based on live cell imaging [13]. The database has collected 93,578 RNA interference data sets from 10 independent studies on 135 phenotype terms, and provides a platform for querying and browsing cellular phenotypes. It can be queried by gene, gene attribute, reagent, phenotype, study keywords or ontology terms.

Horn et al. developed a method to identify genetic interactions by double-RNAi [14]. They performed about 70,000 pairwise perturbations of signaling factors by *Drosophila melanogaster* cells. They acquired the fluorescent images and extracted features, such as the number of cells, intensity and area, from these images. These quantified phenotypes of cells under single-RNAi or double-RNAi were used to identify potential interactions. The interaction identification criterion is: if the phenotypes of the two genes simultaneously targeted deviate strongly from the cumulative distribution of the two single-knockdown phenotypes. Deviations between the expected results based on single-RNAi and the experimentally observed double-RNAi effects were used to reveal negative (aggravating for genes with single-RNAi effect) or positive (alleviating) genetic interactions. Finally they identified more than 600 interactions affecting the quantified phenotypes. Laufer et al. presented a method to study genetic interactions based on high-throughput imaging and RNAi [15]. They quantified many phenotypes of human cancer cells, such as cell number, eccentricity and nuclear area. In their experiments, they identified 2376 gene pairs' interactions affecting the quantified phenotypes. In these two studies, Student's *t*-test between the observed values and the non-interacting model was used to evaluate the significance of an interaction.

Boland et al. developed method to quantify the morphological and textual features, then they used these features to classify different cellular protein localization patterns [16]. They quantified two sets of numeric features, they are Zernike moments [17] and Haralick texture features [18]. These features were then used to train a classification tree and a neural network. They tested their classification performance on independent image sets.

Recognition and segmentation of individual cells is a critical operation in image analysis. Because almost all previous segmentation methods are dependent on specific analytic applications, it is difficult to directly use these methods for other applications. Held et al. developed an automatic parameter optimization method based on genetic algorithm to determine the parameters of different image segmentation methods, such as maximum-intensity linking method and watershed image segmentation method [19]. Bensch et al. developed a method based on the min-cut approach [20] where they applied the notion that at all cell borders there is a dark-to-bright transition to improve the segmentation results.

Ramm et al. presented an automated image analysis system to analyze neurite outgrowth [21]. They automatically quantified several features, such as cell area, number of neuritis and neurite length. Their results showed that their automatically quantified results are similar to manually measured results.

Wang et al. developed a method to quantitatively analyze pollen tube growth based on images [22]. They quantified many features, such as pollen grain area, longest backbone length and width. These quantified features were then used to classify pollen tubes into six groups.

For cellular phenotype analysis, Carpenter et al. developed the CellProfiler, a versatile package in Python, to help biologists analyze cell-based experimental images [2,3,23–25]. It can quantify many cellular phenotypes, such as cell number, cell size and morphological phenotypes of cells. Fiji [26] is a batteries-included package of ImageJ written in Java [27], bundling many plugins for 2D and 3D image analysis. Other platforms are summarized and compared in [28]. These platforms along with their modules or plugins enable quantification of cellular phenotypes, including size, intensity, and texture of single cell, and counts, topology and relative distance of cell colonies [2,24]. More subtle phenotypes such as the distribution pattern of peripheral chromatin in the nuclei of cells [29] and molecule colocalization [30] are also implemented in plugins of ImageJ, such as JACoP [31].

After quantifying image features, data mining or pattern recognition of the phenotypes then becomes more important [32]. A classifier can be trained based on a feature vector extracted from an image, using different machine learning algorithms. CellProfiler Analyst [25,3] is a machine learning package designed for analyzing features extracted from CellProfiler, using the GentleBoosting algorithm [33]. Many algorithms, including nearest neighbor [34], support vector machines [35], random forest [36] and neural network [37] have been used to train classifiers. ImageJ/Fiji and BIOCAT [38] are two platforms that provide several pattern recognition models. The latter also provides algorithm comparisons and thus enables selection for the most effective model.

2. Spatial and temporal information from tissue images

Development biology questions are often about location and/or time. Tissue images help to quantitatively measure spatial and temporal phenotypes, especially gene expression, a molecular phenotype. Microarray and sequencing techniques *per se* take advantage of fluorescent signal detection and image processing [39,40], however, their interpretation relies on tissue dissection, such as laser microdissection [41] or fluorescence activated cell sorting [42], and therefore generally have low spatial resolution. RNA *in situ* hybridization (ISH) [43] can provide high-resolution expression signals from *in situ* images, even at single-cell level. Nonetheless, there was little need for computational quantitation, until the recent booming of high-throughput tissue ISH images by automatic pipelines [44–46]. Among them, Eurexpress project, Allen Institute for Brain Science, St. Jude Brain Gene Expression Map, GenitoUrinary Development Molecular Anatomy Project and so on have generated or maintained large-scale ISH image resources [47–53]. Eurexpress project provides ISH images for >18,000 genes in E14.5 mouse embryo, and manual spatial annotations based on the EMAP reference anatomy tree [54]. This manual curation approach is robust across various image qualities, yet low-throughput and low resolution with a few discrete categorical levels. At the meantime, Allen Institute for Brain Science provides brain images at multiple developmental stages for mouse and human, and quantifies expression with an automated pipeline. The pipeline constructs reference 3D models at specific stages that agree with MRI images, registers series of ISH images to the corresponding models by mutual information alignment, and digitally summarize the signal intensities within each "voxel" [49,51,55]. In addition, automated methods mining these resources were mostly developed based on mutual information [56,57]. To take advantage of the robustness of manual curation and high efficiency of image processing, semi-automated approaches were also developed to systematically retrieve molecular phenotypes. For example, Hill and Baldock suggested a relatively fast constrained

distance transform (CDT) method with mesh and geodesic distance, for simultaneously interactive image registration [58]. Either automated or semi-automated registration with deformation at the expense of resolution may not be suitable for some particular biological questions. Huang et al. attempted to simplify the model to 1D at very high spatial resolutions – at the single-cell level, and proposed a semi-automated method to digitize high-resolution signals from images, which captures single-cell level gene expression during embryonic neuronal radial migration in mouse brain [59]. Such high-resolution spatial profiles allowed to identify synchronized expression of chromatin remodeling genes in the intermediate zone [59]. Besides ISH images, with the need for improving resolution and throughput together, new techniques are emerging, such as *in situ* sequencing [60,61], a technology that sequence RNAs on site at the original location of the sample, and generating more valuable spatial and temporal images for tissues.

3. Quantitative phenotypes of *Caenorhabditis elegans*

C. elegans is a powerful genetic model organism in studying development [62], apoptosis [63,64], aging [65–67], RNA-mediated interference [68] and drug effects [69]. In high-throughput phenotyping of *C. elegans*, manual quantification is subjective, laborious and sometime impossible. Thus effective automated nematode image processing and quantification methods are urgently needed.

Quantitative image analysis of *C. elegans* can be intuitively divided into two categories: static and dynamic image analysis. On the static side, common phenotypes include size, shape, fertility, area, and distinguishing head from tail.

Size and shape can be accessed quantitatively by many software packages, such as QuantWorm [70] or by plug-ins of widely-used platforms like WormSizer [71]. Fertility is another important phenotypes of *C. elegans* and thus auto-counting of eggs is implemented by QuantWorm [70]. QuantWorm first detects some single eggs, which are distinguished by object edge detection, image segmentation and region labeling. These detected single eggs are then used to compute the egg parameters, such as the average size and average gray value of eggs to find egg candidates on binary images.

Head identification can be achieved by fluorescent markers [72], direction of locomotion [73], or support vector machine classifier [74]. Features, such as area, perimeter and radial distribution, for each of the detected foreground particles have been used to train support vector machine models to distinguish pharyngeal grinder from other particles, then the worm head is identified by the presence of pharyngeal grinder.

Some interesting works on static image analysis also stem from experimental needs, such as straightening worms [75] and untangling worms [76] for comparison between individual worms. The straightening worm method invented by Peng et al. facilitates the comparability of worms with diverse shapes [75]. Untangling worm developed in WormToolbox aims to untangle worms to facilitate analysis of individual worms [76].

Using confocal image stacks, Long et al. built a 3D digital nuclear atlas of the first larval stage (L1) of *C. elegans* at single-cell resolution [77]. DAPI (4,6-diamidino-2-phenylindole) was used to stain the nuclei of cells in each *C. elegans*, and a *myo-3: GFP* transgene was used to mark some anchor cells of body wall muscle. Then, based on the staining/fluorescence information, they developed a support vector machine classifier to improve the segmentation results of each individual nucleus, and a method to automatically identify 357 nuclei of L1 stage *C. elegans* using a series of machine learning approaches, such as random sample consensus approach and bipartite matching algorithm with high accuracy. They thus quantified 93 genes' expression profiles in 363 cells [78].

As for the dynamic imaging analysis, since *C. elegans* has been extensively used in aging studies [65–67], lifespan is one of the most important phenotype. Traditionally, lifespan assay is done on agar surface. However, for drug screen experiments, liquid culture is widely used [69] because it needs smaller amount of drugs, maintains more accurate drug concentration and is more high-throughput than solid media. Manual lifespan quantification in either liquid or solid media is laborious and time consuming.

To automate lifespan assays, tools have been developed for different types of nematode cultivation [79]. For worms on agar surface, the *C. elegans* lifespan machine [80] and WormScan [81] can automate lifespan assay and plot survival curves. For worms in microfluidic devices, WormFarm achieves a similar function [82]. For worms in liquid media, the WormToolbox [76] of CellProfiler [2,24] can identify the dead worm by their straight shape, which in practice is not precise enough to tell the live and dead worms, especially when worms died of natural causes, such as aging, rather than being killed by toxic reagents.

Worm tracking is important for scoring phenotypes in neuroscience studies, as complicated locomotory and bending behavior is a indicator of various gene functions [83]. Such a challenging task can be solved by using centroid position based trackers [84] or skeleton-based trackers [85] for single or multiple worms [86,87]. In the tracking process, posture, forward or reverse movement, speed and angular velocity can be measured as quantitative phenotypes [88] to screen for stochastic behaviors [89] from stored videos or real time feed-in images.

4. Behavior analysis of model organisms

Quantifying behavior is important for dissecting gene and drug functions in cognition. Temporal dynamics of behaviors reflect the way internal and external stimuli are memorized and integrated. A common way to characterize behaviors of model organisms is through visually analyzing videos of experiments. However, human visual inspection is difficult to give a quantitative assessment and cannot follow complex behaviors.

Rihel et al. developed a video-based behavior screen for small molecules. They used zebrafish behavioral profiles to quantitatively screen for drugs that change the behavior phenotypes of larval zebrafish [6]. They cultured larval zebrafish in 96-well plate with different drugs. The movement of each larva over 3 days was recorded using a video tracking system. Then from the activity videos, they quantified many behavioral parameters, such as the duration and number of rest bouts, waking activity, and rest latency, by their in-house image analysis package. They screened 5648 compounds totally, and found 547 compounds significantly altered the behavior of zebrafish.

Ant societies have division of labor. To learn about how the division is generated, Mersch et al. developed an automated video tracking system which can continuously monitor and distinguish each ant in one ant colony [7]. They marked each ant with unique matrix code. These code markers enabled their system to identify each ant unambiguously. They recorded the activities of 6 ant colonies over 41 days, and quantified ants' positions and interactions data. These quantified data indicate that colonies are structured in three social groups (nurses, foragers, and cleaners), and their social role changed from nursing to cleaning to foraging as they age.

Similarly, de Chaumont et al. developed a computerized method to study the social interactions of mice [8]. They defined and quantified more than 20 behavioral phenotypes between two mice from experiment videos, such as follow behavior and stop behavior.

These quantified behavior events are useful for examination of decision-making and complex behaviors in mouse models of neurological diseases.

5. Human facial quantitative phenotype analysis

Human face is probably the most basic biometric identifier that is normally used by humans to recognize people. Human face is a morphological structure with complex traits. Human face has been widely studied in many fields, including anthropology, psychology, aging, biometrics, medicine and forensics. However, most of previous studies have been largely based on manual quantifications. Visual assessment, subjectively given by an observer, is sensitive to the judgment of observer. As acquisition of high content facial images becomes easier, methods based on image analysis become increasingly important to enable high-throughput quantitative analysis of facial traits.

Normalization of different facial photos is essential for comparison of different faces [90]. To compare 2D facial images of different individuals, a commonly used normalization method is based on landmarks and relative geometric relationships to describe face shapes and quantify the differences. For example, Turaga et al. used 2D geometric attributes described by a set of facial landmarks to perform 2D facial alignment and comparison [91]. They manually labeled 47 facial landmarks (around mouth, nose, eyes and facial contour, etc.) on each face, and used these landmarks to describe the geometric attributes of each face. As using the 2-dimensional coordinates of these landmarks to describe face shapes is sensitive to affine transformation, Grassmann manifold $G_{2,47}$ (47 facial landmarks, each landmark has 2 coordinates from plane coordinate system) defined as the space spanned by these landmarks was used to interpret different facial shapes. Then each face will be a point on the Grassmann manifold, and the average face or reference face was calculated as the 'Karcher' mean using training faces [92]. Subsequently, new faces were quantified by a matrix indicating direction and speed of geodesic flow from the average face. The matrix was then used as the normalized shape signature.

Du et al. used 2D human facial images to study facial expressions of emotion [93]. They collected facial expressions of 230 subjects, and then defined the feature space of face shape by a series of quantified distances between fiducial points. They used these quantified distance features to classify emotions with high accuracy.

Vernon et al. used facial image analysis to uncover the information driving our judgments of first impressions [94]. They quantified 65 facial features, such as head area, head width, nose width, from each facial image. Then they characterized the relationship between facial features and perceived social traits. They also built prediction model from these quantified facial features to predict the perceived social trait scores of each face.

Compared with 2D facial data, 3D facial data consists of spatial coordinates, and retains more facial information. The 3D facial image alignment method is also landmark-guided. These landmarks are manually labeled [95] or automatically recognized by machine-learning methods [96]. A reference face is warped to register each individual 3D face under the guidance of landmarks by thin-plate spline transformation. Then Generalized Procrustes analysis (GPA) can be used to align all the registered faces into a common coordinate system. After alignment, all 3D vertices can be used as high density landmarks to be compared across all individuals [97].

Hutton et al. trained a face growth model based on 400 3D human facial images [95]. Their results showed that face size changed greatly before adulthood, and no big changes in subjects older than 18 years.

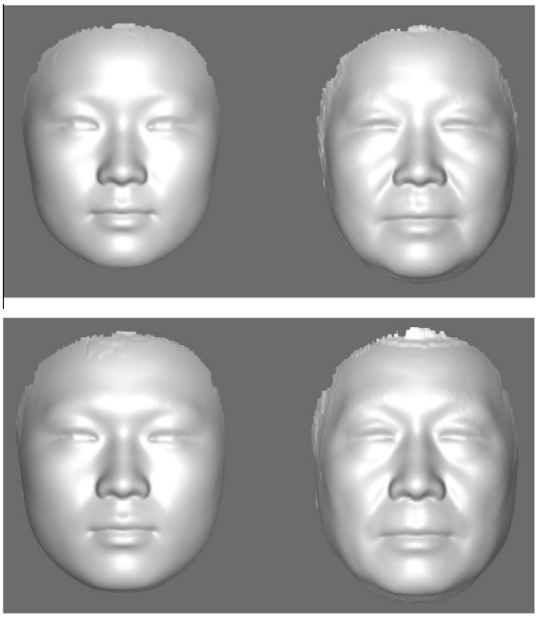

Fig. 1. The average 3D images of young and old female (upper panel) and male (lower panel) faces, respectively.

To uncover facial phenotypes indicative of general health status and age, Chen et al. collected more than 300 3D human facial images; by analyzing these facial images, they identified quantitative facial features highly associated with age, such as mouth width, nose width increase with aging, while eye corners droop with aging. They visualized the quantified 3D image-based facial aging phenotypes. The average 3D human facial models reflect the general facial aging patterns clearly (Fig. 1). They also constructed an age predictor based on 3D human facial images. The identified slow and fast agers compared with subjects' actual age are well supported by their facial appearances and the levels of health indicators [98].

6. Discussions

Bioimaging approaches to automatically detect, quantify, and profile the phenotypic changes related to specific biological question are becoming increasingly useful for modern biology studies. In summary, quantitative phenotype analysis is not only a young field that enables precise descriptions of a biological phenomenon, but also an important platform that links different levels of biological data, moving us closer to a more comprehensive understanding of systematic change related to a specific biological question.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China 91329302 and 31210103916; China Ministry of Science and Technology 2015CB964803 and 2011CB504206; Chinese Academy of Sciences XDA01010303 and YZ201243 to J.D.J.H.

References

[1] H. Peng, Bioinformatics 24 (2008) 1827–1836.
[2] A.E. Carpenter, T.R. Jones, M.R. Lamprecht, C. Clarke, I.H. Kang, O. Friman, D.A. Guertin, J.H. Chang, R.A. Lindquist, J. Moffat, P. Golland, D.M. Sabatini, Genome Biol. 7 (2006) R100.
[3] T.R. Jones, A.E. Carpenter, M.R. Lamprecht, J. Moffat, S.J. Silver, J.K. Grenier, A.B. Castoreno, U.S. Eggert, D.E. Root, P. Golland, D.M. Sabatini, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 1826–1831.

[4] A.E. Carpenter, Nat. Chem. Biol. 3 (2007) 461–465.
[5] A.G. White, B. Lees, H.L. Kao, P.G. Cipriani, E. Munarriz, A.B. Paaby, K. Erickson, S. Guzman, K. Rattanakorn, E. Sontag, D. Geiger, K.C. Gunsalus, F. Piano, IEEE Trans. Med. Imaging 32 (2013) 1791–1803.
[6] J. Rihel, D.A. Prober, A. Arvanites, K. Lam, S. Zimmerman, S. Jang, S.J. Haggarty, D. Kokel, L.L. Rubin, R.T. Peterson, A.F. Schier, Science 327 (2010) 348–351.
[7] D.P. Mersch, A. Crespi, L. Keller, Science 340 (2013) 1090–1093.
[8] F. de Chaumont, R.D. Coura, P. Serreau, A. Cressant, J. Chabout, S. Granon, J.C. Olivo-Marin, Nat. Methods 9 (2012) 410–417.
[9] B. Neumann, T. Walter, J.K. Heriche, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Cetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wunsche, V. Satagopam, M.H. Schmitz, C. Chapuis, D.W. Gerlich, R. Schneider, R. Eils, W. Huber, J.M. Peters, A.A. Hyman, R. Durbin, R. Pepperkok, J. Ellenberg, Nature 464 (2010) 721–727.
[10] T. Walter, M. Held, B. Neumann, J.K. Heriche, C. Conrad, R. Pepperkok, J. Ellenberg, J. Struct. Biol. 170 (2010) 1–9.
[11] M. Pirooznia, Y. Deng, BMC Bioinf. 7 (Suppl. 4) (2006) S25.
[12] C.-C. Chang, C.-J. Lin, ACM Trans. Intell. Syst. Technol. 2 (2011) 1–27.
[13] C. Kirsanova, A. Brazma, G. Rustici, U. Sarkans, Bioinformatics 31 (2015) 2736–2740.
[14] T. Horn, T. Sandmann, B. Fischer, E. Axelsson, W. Huber, M. Boutros, Nat. Methods 8 (2011) 341–346.
[15] C. Laufer, B. Fischer, M. Billmann, W. Huber, M. Boutros, Nat. Methods 10 (2013) 427–431.
[16] M.V. Boland, M.K. Markey, R.F. Murphy, Cytometry 33 (1998) 366–375.
[17] A. Khotanzad, Y.H. Hong, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 489–497.
[18] R.M. Haralick, K. Shanmugam, I. Dinstein, IEEE Trans. Syst. Man Cybern. 3 (1973) 610–621.
[19] C. Held, R. Palmisano, L. Haberle, M. Hensel, T. Wittenberg, Cytometry A 79 (2011) 933–945.
[20] R. Bensch, O. Ronneberger, Proc. IEEE Int. Symp. Biomed. Imaging (2015) 1220–1223.
[21] P. Ramm, Y. Alexandrov, A. Cholewinski, Y. Cybuch, R. Nadon, B.J. Soltys, J. Biomol. Screen. 8 (2003) 7–18.
[22] C. Wang, C.P. Gui, H.K. Liu, D. Zhang, A. Mosig, J. Integr. Plant Biol. 55 (2013) 131–141.
[23] L. Kamentsky, T.R. Jones, A. Fraser, M.A. Bray, D.J. Logan, K.L. Madden, V. Ljosa, C. Rueden, K.W. Eliceiri, A.E. Carpenter, Bioinformatics 27 (2011) 1179–1180.
[24] M.R. Lamprecht, D.M. Sabatini, A.E. Carpenter, Biotechniques 42 (2007) 71–75.
[25] T.R. Jones, I.H. Kang, D.B. Wheeler, R.A. Lindquist, A. Papallo, D.M. Sabatini, P. Golland, A.E. Carpenter, BMC Bioinf. 9 (2008) 482.
[26] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D.J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Nat. Methods 9 (2012) 676–682.
[27] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, Nat. Methods 9 (2012) 671–675.
[28] D. Laksameethanasan, R.Z. Tan, G.W.L. Toh, L.-H. Loo, BMC Bioinf. 14 (2013) S4.
[29] S. Basu, S. Kolouri, G.K. Rohde, Proc. Natl. Acad. Sci. 111 (2014) 3448–3453.
[30] T. Lagache, N. Sauvonnet, L. Danglot, J.-C. Olivo-Marin, Cytometry 87 (2015) 568–579.
[31] S. Bolte, F.P. Cordelières, J. Microsc. 224 (2006) 213–232.
[32] C. Sommer, D.W. Gerlich, J. Cell Sci. 126 (2013) 5529–5539.
[33] J. Friedman, T. Hastie, R. Tibshirani, Ann. Stat. 28 (2000) 337–374.
[34] L. Shamir, N. Orlov, D.M. Eckley, T. Macura, J. Johnston, I.G. Goldberg, Source Code Biol. Med. 3 (2008) 13.
[35] M. Wang, X. Zhou, F. Li, J. Huckins, R.W. King, S.T.C. Wong, Bioinformatics 24 (2008) 94–101.
[36] A. Kreshuk, C.N. Straehle, C. Sommer, U. Koethe, M. Cantoni, G. Knott, F.A. Hamprecht, PLoS ONE 6 (2011) e24899.
[37] M.V. Boland, R.F. Murphy, Bioinformatics 17 (2001) 1213–1223.
[38] J. Zhou, S. Lamichhane, G. Sterne, B. Ye, H. Peng, BMC Bioinf. 14 (2013) 291.
[39] L.H. Augenlicht, D. Kobrin, Cancer Res. 42 (1982) 1088–1093.
[40] Z. Wang, M. Gerstein, M. Snyder, Nat. Rev. Genet. 10 (2009) 57–63.
[41] M.R. Emmert-Buck, R.F. Bonner, P.D. Smith, R.F. Chuaqui, Z. Zhuang, S.R. Goldstein, R.A. Weiss, L.A. Liotta, Science 274 (1996) 998–1001.
[42] M.H. Julius, T. Masuda, L.A. Herzenbe, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 1934.
[43] J.G. Gall, M.L. Pardue, Proc. Natl. Acad. Sci. U.S.A. 63 (1969) 378–383.
[44] U. Herzig, C. Cadenas, F. Sieckmann, W. Sierralta, C. Thaller, A. Visel, G. Eichele, Novart. Found. Symp. 239 (2001) 129–149.
[45] J.P. Carson, C. Thaller, G. Eichele, Curr. Opin. Neurobiol. 12 (2002) 562–565.
[46] A. Visel, C. Thaller, G. Eichele, Nucleic Acids Res. 32 (2004) D552–D556.
[47] G. Diez-Roux, S. Banfi, M. Sultan, L. Geffers, S. Anand, D. Rozado, A. Magen, E. Canidio, M. Pagani, I. Peluso, N. Lin-Marq, M. Koch, M. Bilio, I. Cantiello, R. Verde, C. De Masi, S.A. Bianchi, J. Cicchini, E. Perroud, S. Mehmeti, E. Dagand, S. Schrinner, A. Nurnberger, K. Schmidt, K. Metz, C. Zwingmann, N. Brieske, C. Springer, A.M. Hernandez, S. Herzog, F. Grabbe, C. Sieverding, B. Fischer, K. Schrader, M. Brockmeyer, S. Dettmer, C. Helbig, V. Alunni, M.A. Battaini, C. Mura, C.N. Henrichsen, R. Garcia-Lopez, D. Echevarria, E. Puelles, E. Garcia-Calero, S. Kruse, M. Uhr, C. Kauck, G. Feng, N. Milyaev, C.K. Ong, L. Kumar, M. Lam, C.A. Semple, A. Gyenesei, S. Mundlos, U. Radelof, H. Lehrach, P. Sarmientos, A. Reymond, D.R. Davidson, P. Dolle, S.E. Antonarakis, M.L. Yaspo, S. Martinez, R.A. Baldock, G. Eichele, A. Ballabio, PLoS Biol. 9 (2011) e1000582.
[48] E.S. Lein, M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A.F. Boe, M. S. Boguski, K.S. Brockway, E.J. Byrnes, L. Chen, L. Chen, T.-M. Chen, M.C. Chin, J. Chong, B.E. Crook, A. Czaplinska, C.N. Dang, S. Datta, N.R. Dee, A.L. Desaki, T. Desta, E. Diep, T.A. Dolbeare, M.J. Donelan, H.-W. Dong, J.G. Dougherty, B.J. Duncan, A.J. Ebbert, G. Eichele, L.K. Estin, C. Faber, B.A. Facer, R. Fields, S.R. Fischer, T.P. Fliss, C. Frensley, S.N. Gates, K.J. Glattfelder, K.R. Halverson, M.R. Hart, J.G. Hohmann, M.P. Howell, D.P. Jeung, R.A. Johnson, P.T. Karr, R. Kawal, J. M. Kidney, R.H. Knapik, C.L. Kuan, J.H. Lake, A.R. Laramee, K.D. Larsen, C. Lau, T. A. Lemon, A.J. Liang, Y. Liu, L.T. Luong, J. Michaels, J.J. Morgan, R.J. Morgan, M.T. Mortrud, N.F. Mosqueda, L.L. Ng, R. Ng, G.J. Orta, C.C. Overly, T.H. Pak, S.E. Parry, S.D. Pathak, O.C. Pearson, R.B. Puchalski, Z.L. Riley, H.R. Rockett, S.A. Rowland, J.J. Royall, M.J. Ruiz, N.R. Sarno, K. Schaffnit, N.V. Shapovalova, T. Sivisay, C.R. Slaughterbeck, S.C. Smith, K.A. Smith, B.I. Smith, A.J. Sodt, N.N. Stewart, K.-R. Stumpf, S.M. Sunkin, M. Sutram, A. Tam, C.D. Teemer, C. Thaller, C.L. Thompson, L.R. Varnam, A. Visel, R.M. Whitlock, P.E. Wohnoutka, C.K. Wolkey, V.Y. Wong, M. Wood, M.B. Yaylaoglu, R.C. Young, B.L. Youngstrom, X.F. Yuan, B. Zhang, T.A. Zwingman, A.R. Jones, Nature 445 (2007) 168–176.
[49] M.J. Hawrylycz, E.S. Lein, A.L. Guillozet-Bongaarts, E.H. Shen, L. Ng, J.A. Miller, L.N. van de Lagemaat, K.A. Smith, A. Ebbert, Z.L. Riley, C. Abajian, C.F. Beckmann, A. Bernard, D. Bertagnolli, A.F. Boe, P.M. Cartagena, M.M. Chakravarty, M. Chapin, J. Chong, R.A. Dalley, B.D. Daly, C. Dang, S. Datta, N. Dee, T.A. Dolbeare, V. Faber, D. Feng, D.R. Fowler, J. Goldy, B.W. Gregor, Z. Haradon, D.R. Haynor, J.G. Hohmann, S. Horvath, R.E. Howard, A. Jeromin, J.M. Jochim, M. Kinnunen, C. Lau, E.T. Lazarz, C. Lee, T.A. Lemon, L. Li, Y. Li, J.A. Morris, C.C. Overly, P.D. Parker, S.E. Parry, M. Reding, J.J. Royall, J. Schulkin, P.A. Sequeira, C.R. Slaughterbeck, S.C. Smith, A.J. Sodt, S.M. Sunkin, B.E. Swanson, M.P. Vawter, D. Williams, P. Wohnoutka, H.R. Zielke, D.H. Geschwind, P.R. Hof, S.M. Smith, C. Koch, S.G.N. Grant, A.R. Jones, Nature 489 (2012) 391–399.
[50] J.A. Miller, S.L. Ding, S.M. Sunkin, K.A. Smith, L. Ng, A. Szafer, A. Ebbert, Z.L. Riley, J.J. Royall, K. Aiona, J.M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R.A. Dalley, N. Dee, T.A. Dolbeare, B. A. Facer, D. Feng, T.P. Fliss, G. Gee, J. Goldy, L. Gourley, B.W. Gregor, G. Gu, R.E. Howard, J.M. Jochim, C.L. Kuan, C. Lau, C.K. Lee, F. Lee, T.A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P.D. Parker, S.E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N.V. Shapovalova, E.H. Shen, N. Sjoquist, C.R. Slaughterbeck, M. Smith, A.J. Sodt, D. Williams, L. Zollei, B. Fischl, M.B. Gerstein, D.H. Geschwind, I.A. Glass, M.J. Hawrylycz, R.F. Hevner, H. Huang, A.R. Jones, J.A. Knowles, P. Levitt, J.W. Phillips, N. Sestan, P. Wohnoutka, C. Dang, A. Bernard, J.G. Hohmann, E.S. Lein, Nature 508 (2014) 199–206.
[51] C.L. Thompson, L. Ng, V. Menon, S. Martinez, C.K. Lee, K. Glattfelder, S.M. Sunkin, A. Henry, C. Lau, C. Dang, R. Garcia-Lopez, A. Martinez-Ferre, A. Pombero, J.L. Rubenstein, W.B. Wakeman, J. Hohmann, N. Dee, A.J. Sodt, R. Young, K. Smith, T.N. Nguyen, J. Kidney, L. Kuan, A. Jeromin, A. Kaykas, J. Miller, D. Page, G. Orta, A. Bernard, Z. Riley, S. Smith, P. Wohnoutka, M.J. Hawrylycz, L. Puelles, A.R. Jones, Neuron 83 (2014) 309–323.
[52] S. Magdaleno, P. Jensen, C.L. Brumwell, A. Seal, K. Lehman, A. Asbury, T. Cheung, T. Cornelius, D.M. Batten, C. Eden, S.M. Norland, D.S. Rice, N. Dosooye, S. Shakya, P. Mehta, T. Curran, PLoS Biol. 4 (2006) 497–500.
[53] E.W. Brunskill, B.J. Aronow, K. Georgas, B. Rumballe, M.T. Valerius, J. Aronow, V. Kaimal, A.G. Jegga, S. Grimmond, A.P. McMahon, L.T. Patterson, M.H. Little, S. S. Potter, Dev. Cell 15 (2008) 781–791.
[54] T.F. Hayamizu, M.N. Wicks, D.R. Davidson, A. Burger, M. Ringwald, R.A. Baldock, J. Biomed. Semant. 4 (2013) 15.
[55] M. Hawrylycz, R.A. Baldock, A. Burger, T. Hashikawa, G.A. Johnson, M. Martone, L. Ng, C. Lau, S.D. Larson, J. Nissanov, L. Puelles, S. Ruffins, F. Verbeek, I. Zaslavsky, J. Boline, PLoS Comput. Biol. 7 (2011) e1001065.
[56] M. Jagalur, C. Pal, E. Learned-Miller, R.T. Zoeller, D. Kulp, BMC Bioinf. 8 (Suppl. 10) (2007) S5.
[57] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, IEEE Trans. Med. Imaging 16 (1997) 187–198.
[58] B. Hill, R.A. Baldock, BMC Bioinf. 16 (2015) 90.
[59] Y. Huang, X. Yu, N. Sun, N. Qiao, Y. Cao, J.D. Boyd-Kirkup, Q. Shen, J.D. Han, Genome Res. 25 (2015) 570–581.
[60] R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wahlby, M. Nilsson, Nat. Methods 10 (2013) 857–860.
[61] J.H. Lee, E.R. Daugharthy, J. Scheiman, R. Kalhor, J.L. Yang, T.C. Ferrante, R. Terry, S.S.F. Jeanty, C. Li, R. Amamoto, D.T. Peters, B.M. Turczyk, A.H. Marblestone, S.A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach, G.M. Church, Science 343 (2014) 1360–1363.
[62] S. Brenner, Genetics 77 (1974) 71–94.
[63] M.O. Hengartner, H.R. Horvitz, Philos. Trans. R. Soc. Lond. B Biol. Sci. 345 (1994) 243–246.
[64] M.O. Hengartner, H. Robert Horvitz, Curr. Opin. Genet. Dev. 4 (1994) 581–586.
[65] C. Kenyon, J. Chang, E. Gensch, A. Rudner, R. Tabtiang, Nature 366 (1993) 461–464.
[66] P.L. Larsen, P.S. Albert, D.L. Riddle, Genetics 139 (1995) 1567–1583.
[67] K.D. Kimura, H.A. Tissenbaum, Y. Liu, G. Ruvkun, Science 277 (1997) 942–946.
[68] A. Fire, S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, C.C. Mello, Nature 391 (1998) 806–811.
[69] L.P. O'Reilly, C.J. Luke, D.H. Perlmutter, G.A. Silverman, S.C. Pak, Adv. Drug Deliv. Rev. 69–70 (2014) 247–253.
[70] S.-K. Jung, B. Aleman-Meza, C. Riepe, W. Zhong, PLoS ONE 9 (2014) e84830.
[71] B.T. Moore, J.M. Jordan, L.R. Baugh, PLoS ONE 8 (2013) e57142.

[72] J. Larsch, D. Ventimiglia, C.I. Bargmann, D.R. Albrecht, Proc. Natl. Acad. Sci. 110 (2013) E4266–E4273.
[73] K.-M. Huang, P. Cosman, W.R. Schafer, J. Neurosci. Methods 158 (2006) 323–336.
[74] M. Zhan, M.M. Crane, E.V. Entchev, A. Caballero, D.A. Fernandes de Abreu, Q. Ch'ng, H. Lu, PLoS Comput. Biol. 11 (2015) e1004194.
[75] H. Peng, F. Long, X. Liu, S.K. Kim, E.W. Myers, Bioinformatics 24 (2008) 234–242.
[76] C. Wahlby, L. Kamentsky, Z.H. Liu, T. Riklin-Raviv, A.L. Conery, E.J. O'Rourke, K. L. Sokolnicki, O. Visvikis, V. Ljosa, J.E. Irazoqui, P. Golland, G. Ruvkun, F.M. Ausubel, A.E. Carpenter, Nat. Methods 9 (2012) 714–716.
[77] F. Long, H. Peng, X. Liu, S.K. Kim, E. Myers, Nat. Methods 6 (2009) 667–672.
[78] X. Liu, F. Long, H. Peng, S.J. Aerni, M. Jiang, A. Sanchez-Blanco, J.I. Murray, E. Preston, B. Mericle, S. Batzoglou, E.W. Myers, S.K. Kim, Cell 139 (2009) 623–633.
[79] F.R.G. Amrit, R. Ratnappan, S.A. Keith, A. Ghazi, Methods 68 (2014) 465–475.
[80] N. Stroustrup, B.E. Ulmschneider, Z.M. Nash, I.F. Lopez-Moyado, J. Apfeld, W. Fontana, Nat. Methods 10 (2013) 665–670.
[81] M.D. Mathew, N.D. Mathew, P.R. Ebert, PLoS ONE 7 (2012) e33483.
[82] B. Xian, J. Shen, W. Chen, N. Sun, N. Qiao, D. Jiang, T. Yu, Y. Men, Z. Han, Y. Pang, M. Kaeberlein, Y. Huang, J.-D.J. Han, Aging Cell 12 (2013) 398–409.
[83] S.J. Wang, Z.-W. Wang, PLoS ONE 8 (2013) e69653.
[84] J.T. Pierce-Shimomura, T.M. Morse, S.R. Lockery, J. Neurosci. 19 (1999) 9557–9569.
[85] C.J. Cronin, J.E. Mendel, S. Mukhtar, Y.-M. Kim, R.C. Stirbl, J. Bruck, P.W. Sternberg, BMC Genet. 6 (2005) 5.
[86] E. Fontaine, J. Burdick, A. Barr, Conf. Proc. IEEE Eng. Med. Biol. Soc. 1 (2006) 3716–3719.
[87] K.M. Huang, P. Cosman, W.R. Schafer, Proc. IEEE Int. Symp. Biomed. Imaging (2007).
[88] D. Ramot, B.E. Johnson, T.L. Berry, L. Carnell, M.B. Goodman, PLoS ONE 3 (2008) e2208.
[89] N.A. Swierczek, A.C. Giles, C.H. Rankin, R.A. Kerr, Nat. Methods 8 (2011) 592–598.
[90] G. Guo, Y. Fu, C.R. Dyer, T.S. Huang, IEEE Trans. Image Process. 17 (2008) 1178–1188.
[91] P. Turaga, S. Biswas, R. Chellappa, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2010) 946–949.
[92] H. Karcher, Commun. Pure Appl. Math. 30 (1977) 509–541.
[93] S. Du, Y. Tao, A.M. Martinez, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) E1454–E1462.
[94] R.J. Vernon, C.A. Sutherland, A.W. Young, T. Hartley, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) E3353–E3361.
[95] T.J. Hutton, B.F. Buxton, P. Hammond, H.W. Potts, IEEE Trans. Med. Imaging 22 (2003) 747–753.
[96] J. Guo, X. Mei, K. Tang, BMC Bioinf. 14 (2013) 232.
[97] S. Peng, J. Tan, S. Hu, H. Zhou, J. Guo, L. Jin, K. Tang, PLoS Comput. Biol. 9 (2013) e1003375.
[98] W. Chen, W. Qian, G. Wu, B. Xian, X. Chen, Y. Cao, C.D. Green, F. Zhao, K. Tang, J. D. Han, Cell Res. 25 (2015) 574–587.