(**Supplementary Table 1**). The average bead displacement and the ratio between correspondence candidates and true correspondences is a quantitative measure of the reconstruction success, which is crucial for automatic validation of registration results in long time-lapse recordings. The beads can be removed optically or computationally from the sample (**Supplementary Methods**).

We applied the bead-based registration framework to SPIM recordings of early *Drosophila melanogaster* embryos, which are very challenging samples for multiview reconstruction owing to the scattering of the yolk that severely limits the overlap between views. We imaged *Drosophila* embryos expressing ubiquitous His-YFP from five and seven views in an extended time-lapse recording covering early embryonic development. We registered each time point separately and then registered all time points to each other compensating for minor drift during image acquisition (**Supplementary Methods**). We combined content-based fusion with nonlinear blending[5] to compensate for brightness differences at boundaries between views (**Supplementary Fig. 4**). The reconstructed multiview acquisition of the specimen showed, in contrast to the single view, comparable lateral and axial resolution (**Fig. 1e**,**f**). We never imaged the anterior and posterior poles of the embryo with full lateral resolution in this acquisition, and yet the cells were clearly distinguishable, demonstrating the precision of the multiview reconstruction (**Fig. 1g–i**). In the middle of the specimen, the resolution was lower because only some views contributed high-content information whereas other views were blocked by the yolk (**Fig. 1h**). The reconstructed time-lapse recording provided an unprecedented four-dimensional view of *Drosophila* embryogenesis (**Supplementary Videos 4** and **5**).

The bead-based registration framework is sample-independent (**Supplementary Data** and **Supplementary Fig. 5**) and enables fully unguided registration without prior knowledge of the arrangement of the views (**Supplementary Video 4**). The software outperforms intensity-based registration approaches[6,7] in terms of precision and speed, enabling accurate registration of large, multiview acquisitions in minutes (**Supplementary Data**, **Supplementary Fig. 6** and **Supplementary Table 1**). The run time of the bead-based registration framework is comparable to the time it takes to acquire the multiview data, and thus, to our knowledge, it is currently the only solution allowing robust, real-time registration of time-lapse SPIM recordings. Moreover, the bead-based registration framework is applicable to other optical sectioning microscopy techniques (**Supplementary Fig. 7** and **Supplementary Video 6**), considerably expanding the possible applications in biology. We provide our bead-based registration algorithm to the bioimaging community as an open-source plugin for Fiji (**Supplementary Fig. 8** and **Supplementary Software**; http://pacific.mpi-cbg.de/wiki/index.php/SPIM_Registration).

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

ACKNOWLEDGMENTS

We thank Carl Zeiss Microimaging for access to the SPIM demonstrator; R.K. Ejsmont (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden) for His-YFP flies; D. White, E. Dimitrova, M. Sarov and P. Campinho for help with imaging and B. Schmid for help with 3D viewer programming. S.P. and S.S. were supported by a Dresden International Graduate School for Biomedicine and Bioengineering doctorate stipend. J.S. and P.T. acknowledge funding from the Human Frontier Science Program Research grant RGY0084.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Stephan Preibisch[1,2], Stephan Saalfeld[1,2], Johannes Schindelin[1] & Pavel Tomancak[1]

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.
[2]These authors contributed equally to this work.
e-mail: tomancak@mpi-cbg.de

1. Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E.H. *Science* **305**, 1007–1009 (2004).
2. Huisken, J. & Stainier, D.Y. *Development* **136**, 1963–1975 (2009).
3. Lindeberg, T. *J. Appl. Stat.* **21**, 224–270 (1994).
4. Fischler, M.A. & Bolles, R.C. *Commun. ACM* **24**, 381–395 (1981).
5. Preibisch, S., Saalfeld, S. & Tomancak, P. *Bioinformatics* **25**, 1463–1465 (2009).
6. Preibisch, S., Rohlfing, T., Hasak M.P. & Tomancak P. *SPIE Medical Imaging 2008* (eds., Reinhardt, J.M. & Pluim, J.P.W.) **6914,** 69140E-69140E-8 (2008).
7. Swoger, J. *et al. Opt. Express* **15**, 8029–8042 (2007).

Partitioning biological data with transitivity clustering

To the Editor: Clustering is a common computational technique for data analysis in the life sciences. Essentially one defines clustering as a partitioning of arbitrary data objects into groups, such that the objects in each group, or cluster, have common traits, with respect to a similarity function. Ideally, objects from the same cluster are more similar to each other than to objects from different clusters. A density parameter controls the size and the number of resulting clusters.

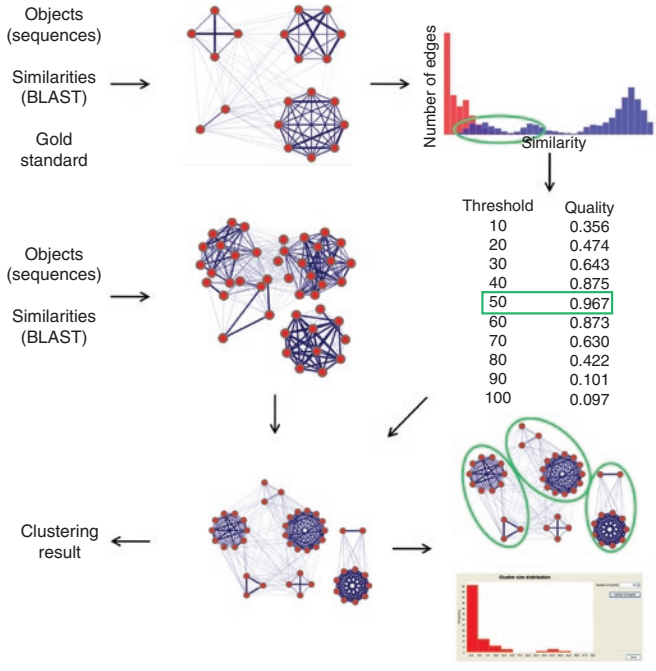

Figure 1 | Data analysis and clustering workflow. After importing and visualizing a subset of the data, a histogram of the distribution of similarities within and between gold standard clusters is created to estimate a promising region for the density parameter. Iterative clustering with varying thresholds and subsequent comparison against the gold standard identifies the 'best threshold'. After importing and visualizing the whole dataset, this threshold is used to detect and report meaningful cluster assignments, which may be subject to subsequent integrated analyses.

Here we present 'transitivity clustering', an alternative way to partition biological data, with easy-to-use interfaces: (i) a web interface for a quick analysis of medium-sized datasets, (ii) a powerful stand-alone Java implementation for large-scale data clustering, and (iii) a collection of Cytoscape[1] plug-ins that also provide methods to answer typical follow-up questions. Transitivity clustering is embedded in an integrated data-analysis framework, TransClust. We applied TransClust for protein sequence clustering, a long-standing challenge in computational biology[2,3]. Using the TransClust Cytoscape plug-ins we followed the typical data-analysis workflow (**Fig. 1**) to reconstruct protein families given the amino acid sequences. For this task, we found that transitivity clustering outperformed typical clustering approaches, such as connected component analysis, Markov clustering, spectral clustering, affinity propagation and hierarchical clustering (**Supplementary Methods**).

The typical process incorporates several data analysis steps: (i) identification of a similarity function, (ii) computation and postprocessing of a similarity matrix, (iii) visualization as a similarity network, (iv) analysis of this network, that is, estimation of meaningful density parameters based on the similarity values (**Supplementary Methods**), (v) clustering of the similarity network, (vi) comparison with given gold standards or toy examples, (vii) fine-tuning of the clustering by varying the clustering tool parameters, mainly the density parameter, (viii) visualization of the clustering results and (ix) follow-up analysis of the results regarding underlying real-world questions. Although these steps are important to draw conclusions from the input data, most clustering frameworks solely concentrate on step v. Mainly two important functionalities are neglected: a semiautomatic estimation of 'good' density parameters (step iv; **Supplementary Fig.1**) and a graphical postprocessing of the clustering results regarding specific biological follow-up questions (step viii; **Supplementary Table 1**). With transitivity clustering and its implementation TransClust, we closed these gaps. The project website is available at http://transclust.cebitec.uni-bielefeld.de/.

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

ACKNOWLEDGMENTS

This work was supported by the German Academic Exchange Service (DAAD), the Boehringer Ingelheim Fonds, US National Institutes of Health grant P41-RR01081, the German Research Foundation (DFG) Graduate School (GK) 635, the German National Genome Research Network (NGFN) and the DFG-funded Cluster of Excellence for Multimodal Computing and Interaction.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Tobias Wittkop[1–3], Dorothea Emig[1,4], Sita Lange[5], Sven Rahmann[6], Mario Albrecht[4], John H Morris[7], Sebastian Böcker[8], Jens Stoye[2,9] & Jan Baumbach[1,4,10]

[1]International Computer Science Institute, University of California at Berkeley, Berkeley, California, USA. [2]Center for Biotechnology, Bielefeld University, Bielefeld, Germany. [3]Buck Institute for Age Research, Novato, California, USA. [4]Max Planck Institute for Informatics, Saarbrücken, Germany. [5]Department of Computer Science, Albert Ludwigs University, Freiburg, Germany. [6]Technische Universität Dortmund, Dortmund, Germany. [7]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California, USA. [8]Institute for Informatics, Friedrich Schiller Universität Jena, Jena, Germany. [9]Faculty of Technology, Bielefeld University, Bielefeld, Germany. [10]Saarland University, Cluster of Excellence for Multimodal Computing and Interaction, Saarbrücken, Germany. e-mail: jbaumbac@mpi-inf.mpg.de

1. Cline, M.S. *et al. Nat. Protoc.* **2**, 2366–2382 (2007).
2. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
3. Wittkop, T., Baumbach, J., Lobo, F.P. & Rahmann, S. *BMC Bioinformatics* **8**, 396 (2007).