community to aggregate, store, and track biologically important cancer variants with provenance supported by the literature.

A variety of somatic cancer variant databases exist that help identify important variants, including gene-level[1], variant-level[2,3], and clinically focused variant interpretation databases[4–6]. These resources have greatly increased our understanding of the landscape of clinically and biologically relevant cancer variants, and when used in aggregate they provide an understanding of the relevance of specific variants. DoCM is a curated repository that facilitates the aggregation of gene and variant information for variants with prognostic, diagnostic, predictive, or functional roles from these resources as well as from individually curated publications (**Supplementary Fig. 1** and **Supplementary Table 1**). DoCM's scope and its batch submission process (**Supplementary Results** and **Supplementary Figs. 2–4**) place it at a critical intersection between the two major tradeoffs of curated resources: comprehensiveness of variants and curation burden (**Fig. 1**). The automated batch submission and the review system allow DoCM curations to scale easily.

Curation of the literature to produce a high-quality set of pathogenic somatic variants is not trivial, on account of the large number of papers and laborious curation process (**Supplementary Fig. 5**). Hence, we designed DoCM as an open resource that can coordinate contributions from research and clinical practitioners. Once important variants are identified, curation efforts are required to format, standardize, and structure the variants for inclusion in DoCM (**Supplementary Methods** and **Supplementary Fig. 6**). A set of such curated variants can be contributed to DoCM by batch submission at http://docm.info/variant_submission, whereupon it is reviewed and evaluated by DoCM editors for possible inclusion. DoCM is licensed under the creative commons attribution license (CC BY 4.0), allowing academic and industry researchers unencumbered access to the content.

We performed a focused knowledge-based variant discovery study to identify pathogenic variants missed in 1,833 cases across four TCGA projects (**Supplementary Methods** and **Supplementary Fig. 7**). Validation sequencing data from 93 of these cases showed that at least one functionally important variant in DoCM was recovered in 41% of cases (**Supplementary Results**, **Supplementary Data 1** and **2**, **Supplementary Figs. 7–9**, and **Supplementary Tables 2–4**). As genomics evolves in the era of precision medicine, and our understanding of the etiology of molecular lesions grows, community curation along with our ongoing efforts will allow DoCM to adapt, refine, and expand with the field.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge L. Trani, J. Hodges, and A. Wollam for efforts in manual review. T. Ley, R. Bose, R. Govindan, and S. Devarakonda provided valuable curation input. D. Larson provided valuable analysis input. M.G. was supported by the National Human Genome Research Institute (NIH NHGRI K99HG007940). O.L.G. was supported by the National Cancer Institute (NIH NCI K22CA188163). This work was supported by a grant to R.K.W. from the National Human Genome Research Institute (NIH NHGRI U54HG003079).

**AUTHOR CONTRIBUTIONS**

B.J.A. wrote the manuscript, supervised creation of the resource, and performed all analysis. M.N.K.C., M.G., O.L.G., E.R.M., and A.H.W. contributed text and revised the manuscript. A.C.C. designed and implemented the web interface, database, and API. B.J.A., A.C.C., M.G., A.H.W., and J.F.M. made contributions to the code. J.F.M. was the lead user experience web developer. M.N.K.C., J.K., and A.H.W. curated publications. R.S.F. supervised validation sequencing. M.G., R.K.W., O.L.G., and E.R.M. supervised the project.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

**Benjamin J Ainscough[1,2], Malachi Griffith[1–3], Adam C Coffman[1], Alex H Wagner[1,2], Jason Kunisaki[1], Mayank NK Choudhary[3], Joshua F McMichael[1], Robert S Fulton[1–3], Richard K Wilson[1–4], Obi L Griffith[1–4] & Elaine R Mardis[1–4]**

[1]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA. [2]Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, USA. [3]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA. [4]Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA.
e-mail: obigriffith@wustl.edu or mgriffit@wustl.edu

1. Van Allen, E.M. *et al. Nat. `Med.* **20**, 682–688 (2014).
2. Forbes, S.A. *et al. Nucleic Acids Res.* **43**, D805–D811 (2015).
3. Zhang, J. *et al. Database (Oxford)* **2011**, bar026 (2011).
4. Yeh, P. *et al. Clin. Cancer Res.* **19**, 1894–1901 (2013).
5. Dienstmann, R. *et al. Mol. Oncol.* **8**, 859–873 (2014).
6. MacConaill, L.E. *et al. J. Mol. Diagn.* **16**, 660–672 (2014).

## Micro-Magellan: open-source, sample-adaptive, acquisition software for optical microscopy

**To the Editor:** The past decade has seen an explosion of new techniques in optical microscopy with the potential to reveal the complex orchestra of biological processes across large scales of space and time[1]. Scalable usability of any new optical technique requires widespread dissemination of hardware and software for data acquisition and of software for data analysis. Commercial software and, more recently, open-source software packages have begun to meet many needs for visualization and analysis of terabyte-sized imaging volumes[2–4]. As the options for data analysis continue to expand and evolve, it becomes increasingly difficult to make ideal use of the full capabilities of widefield, confocal, and multiphoton microscopes already present in many labs because of a lack of automated and customizable acquisition software[5].

To fill this gap, we have developed μMagellan, open-source software (**Supplementary Note 1**) for reproducible, high-throughput imaging of biological samples across heterogeneous scales of space and time. μMagellan provides several high-level automation capabilities for the collection of tiled 3D volumes (**Supplementary Note 2**) that dramatically reduce the amount of time and effort researchers must expend to perform complex experiments. Because it utilizes the hardware abstraction layer of μManager[6], μMagellan can be used with a diverse set of components or complete microscopes from different vendors (**Supplementary Note 3**), enabling many new types of experiments with thousands of existing instruments. In addition, data written by μMagellan can be read directly or imported easily by several commercial and open-source software packages for the visualization and analysis of large volumetric imaging

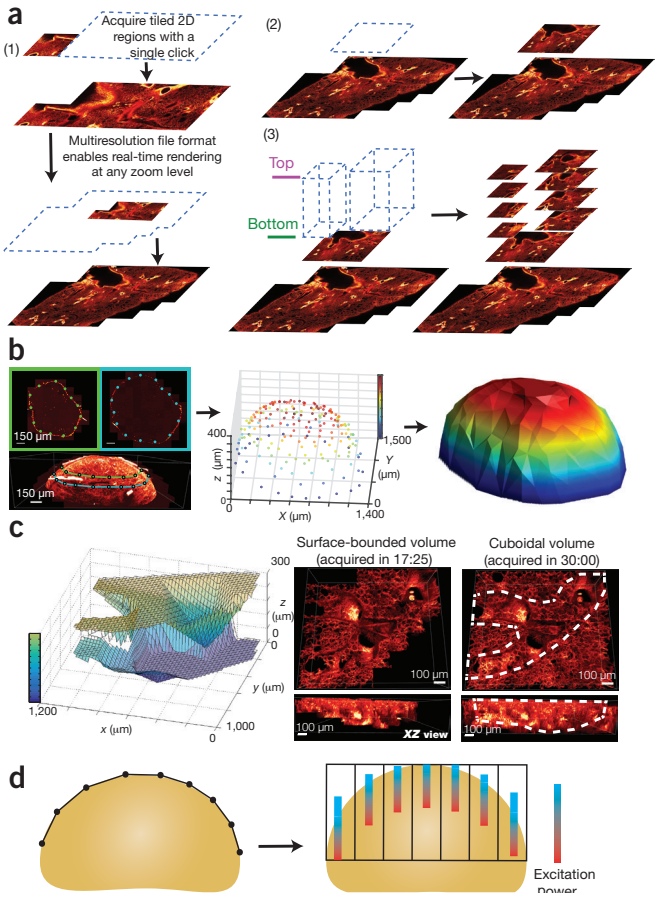


**Figure 1** | Overview of µMagellan's capabilities for exploring samples and running automated, sample-adaptive acquisitions. (**a**) 'Explore acquisitions' present a simple click-and-drag interface for acquisition of tiled high-resolution images. A multiresolution file format allows image panning and zooming in real time. Upper and lower focus limits can be set to collect tiled *z*-stacks. (1), (2), and (3) indicate sequential acquisitions, beginning with exploration of a 2D slice and subsequently tracing the branching airways in a mouse lung in 3D without imaging irrelevant areas of the sample. (**b**) The user can create interpolated surfaces to encode sample morphology (the cortex of a mouse popliteal lymph node is shown here) by marking points on 2D slices to build a 3D point distribution, which is then interpolated into a smooth surface. (**c**) Two surfaces used to bound the acquisition volume of an airway (white dashed lines) in a 400-mm lung slice, compared to the entire bounding cuboidal volume. (**d**) Surfaces and covaried settings used to set gradients of increasing excitation power at different *xy* positions, which begin at the surface marking the top of the sample for each position. Red indicates higher excitation power.

data, including BigDataViewer, Vaa3D/TeraFly, and Imaris (**Supplementary Note 4** and **Supplementary Video 1**).

µMagellan enables users to efficiently map biological samples of unusual shape and unknown spatial organization in three dimensions with 'explore acquisitions'. These present an interactive Google Maps–like user interface that enables fast sample navigation and high-resolution tiled imaging in user-specified shapes and directions (**Fig. 1a** and **Supplementary Video 2**). µMagellan's multiresolution pyramid file format (**Supplementary Note 5**) allows users to pan and zoom through 2D slices of samples in real time. After sample exploration, volumes can be specified for conventional acquisition of 3D cuboids using tiled *z*-stacks, or user-generated surfaces can be specified to bound acquisition volumes of arbitrary shape (**Fig. 1b**,**c** and **Supplementary Videos 3** and **4**). Surfaces are interpolated from user-specified points using an algorithm based on Delaunay triangulations, which allows both arbitrary precision and sublinear scaling of calculation time with the number of points (**Supplementary Fig. 1** and **Supplementary Note 6**).

In addition to allowing for complex, non-cuboidal imaging volumes, µMagellan permits surfaces to be used to automatically control imaging parameters on the basis of sample morphology using covariant pairings. These pairings enable automated variation of a particular hardware setting (such as excitation power, detector gain, exposure, etc.) based on either another hardware setting or, in relation to sample morphology, a calculation involving the geometry of a particular surface (**Fig. 1d**, **Supplementary Note 7**, and **Supplementary Video 5**). Coupling morphological information with acquisition settings in such a way facilitates reproducibility and comparisons across heterogeneous biological samples.

µMagellan is designed to be able to adapt to dynamic biological processes and allows almost all settings (such as spatial regions, time point spacing, automated excitation calculations, etc.) to be altered during acquisition (**Supplementary Note 8** and **Supplementary Video 6**). To compensate for focus drift, which often occurs during long time-lapse experiments, µMagellan provides an algorithm for automated drift compensation based on a designated fiducial channel (**Supplementary Fig. 2** and **Supplementary Note 9**).

Finally, µMagellan provides automation to run multiple acquisitions in series or in parallel. This can be used, for example, to image multiple tissue sections sequentially on a single slide for hours or days at a time, or to monitor multiple sites in a sample or multiple organs from the same animal simultaneously to compare conditions while minimizing biological variability.

µMagellan fills an important niche in the open-source bioimaging software ecosystem by empowering many existing microscopes for automated, reproducible, high-throughput applications. Its open-source code also makes it an ideal platform for the development and dissemination of new technologies, thereby increasing the ease with which they can be put into practice to reveal the mysteries of biological systems. µMagellan comes bundled with µManager and can be accessed in the plug-ins menu under "Acquisition tools." The µMagellan source code and user guide can be found in the supplementary material (**Supplementary Software** and **Supplementary Software Guide**). The latest versions of these materials can be found at https://micro-manager.org/wiki/MicroMagellan.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS**

We thank K. Thorn, M. Tsuchida, C. Weisiger, J.O. Edwards III, and M. Foxxe for helpful conversations during development; M. Headley for beta testing; E. Oswald for figure data; T. Pietzsch, C. Rueden, and G. Iannello for help with integration with BigDataViewer, FIJI, and Vaa3d/TeraFly, respectively; and C. Norberg for support. This work was supported in part by the US National Institutes of Health (grants R01EB007187 (to R.V.), R01AI52116 (to M.F.K.), and U19A1077439-06 (to M.F.K.)) and the Sandler Basic Asthma Research Center (M.F.K.).

**AUTHOR CONTRIBUTIONS**

H.P. and M.F.K. conceived of the project. H.P. created the software with guidance from N.S. H.P. and K.C. performed beta testing. H.P., K.C., and M.F.K. wrote the manuscript. H.P. and K.C. wrote the user guide and recorded screencasts. M.F.K. and R.V. provided administrative and financial support.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

**Henry Pinkard[1–3], Nico Stuurman[4,5], Kaitlin Corbin[1,2], Ronald Vale[4,5] & Matthew F Krummel[1,2]**

[1]Department of Pathology, University of California, San Francisco, San Francisco, California, USA. [2]Biological Imaging Development Center, University of California, San Francisco, San Francisco, California, USA. [3]Computational Biology Graduate Group, University of California, Berkeley, Berkeley, California, USA. [4]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, California, USA. [5]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California, USA.
e-mail: hbp@berkeley.edu

1. Scherf, N. & Huisken, J. *Nat. Biotechnol.* **33**, 815–818 (2015).
2. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. *Nat. Methods* **12**, 481–483 (2015).
3. Peng, H. *et al. Nat. Commun.* **5**, 4342 (2014).
4. Royer, L.A. *et al. Nat. Methods* **12**, 480–481 (2015).
5. Eliceiri, K.W. *et al. Nat. Methods* **9**, 697–710 (2012).
6. Edelstein, A.D. *et al. J. Biol. Methods* **1**, e10 (2014).