bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Brief Communication

CloudReg: Automatic Terabyte-Scale Cross-Modal Brain Volume Registration

Vikram Chandrashekhar BS[1], Daniel J Tward PhD[2], Devin Crowley BS[1], Ailey K Crow PhD[3], Matthew A Wright MD PhD[4], Brian Y Hsueh AB[3,4], Felicity Gore PhD[3,4], Timothy A Machado PhD[3,4], Audrey Branch PhD[5], Jared S Rosenblum MD[6], Karl Deisseroth MD PhD[3,4], Joshua T Vogelstein PhD[1]*

From

[1]Center for Imaging Science, Johns Hopkins University, Baltimore, MD, 21218, United States. [2]Department of Computational Medicine and Neurology, University of California Los Angeles, Los Angeles, CA 90095, United States. [3]CNC Program, Stanford University, Stanford, CA, 94305, United States. [4]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States. [5]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, United States. [6]Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, United States.

To whom correspondence should be addressed:

Joshua Tzvi Vogelstein, Ph.D.
Center for Imaging Science
Clark Hall, Rm 314D
3400 N. Charles St
Baltimore, MD 21218
Phone: 410-516-8775
jovo@jhu.edu

Running title: CloudReg
Key Words: CLARITY, iDISCO, cloud, registration, image processing, terascale

Abstract: 70
Total word Count: 1268
Figure(s): 2
Supplemental Material(s): 5 supplemental figures, 1 supplemental video, 1 supplemental table, and 1 supplemental note

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Quantifying terabyte-scale multi-modal human and animal imaging data requires scalable analysis tools. We developed CloudReg, an open-source, automatic, terabyte-scale, cloud-based image analysis pipeline that pre-processes and registers cross-modal volumetric datasets with artifacts via spatially-varying polynomial intensity transform. CloudReg accurately registers the following datasets to their respective atlases: *in vivo* human and *ex vivo* macaque brain magnetic resonance imaging, *ex vivo* mouse brain micro-computed tomography, and cleared murine brain light-sheet microscopy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Main

Modern imaging methods can generate intact, whole brain data from a variety of modalities including magnetic resonance imaging (MRI), computed tomography (CT), and light-sheet microscopy (LSM) of cleared tissue samples. Each of these methods provides specific information about an individual sample based on the physical principles of the technique, also producing artifacts unique to each technique. MRI can provide detailed anatomic or functional information but can be limited by intensity inhomogeneity due to magnetic field bias.[1] CT can provide detailed anatomic information but can be limited by beam hardening.[2] LSM, in combination with tissue clearing methods, can provide anatomic, functional, and molecular information at subcellular resolution,[3] but can be limited by intensity inhomogeneity due to microscope optics.

Clearing methods including CLARITY (Clear Lipid-exchanged Anatomically Rigid Imaging/immunostaining-compatible Tissue Hydrogel),[3] SHIELD (Stabilization to Harsh conditions via Intramolecular Epoxide Linkages to prevent Degradation),[4] and iDISCO (immunolabeling-enabled three-Dimensional Imaging of Solvent-Cleared Organs)[5] can generate terabytes of data per sample.[6] High-resolution, multi-field-of-view (mFOV) datasets require pre-processing to remove artifacts, stitching into a complete volume, registration to a reference atlas, and visualization in order to perform quantitative analyses.[7,8]

Each of these image processing steps presents unique challenges. First, aligning and stitching every FOV acquired into a complete volume requires significant compute power and is time-intensive. Second, pre-processing imaging data requires correcting artifacts unique to each modality and sample such as intensity inhomogeneity in LSM and MRI. Third, registration methods are frequently intra-modal, have manual components, and are limited by artifacts introduced by specimen preparation and imaging.[9,10] Finally, visualization of these terabyte-scale datasets on a local machine is compute-intensive, slow, and expensive.[11]

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

To address these challenges, we present CloudReg, an automatic, cross-modal, cloud-based pipeline consisting of local and global intensity correction,[12,13] alignment and stitching,[8] image registration with nonlinear methods,[14,15] and interactive online visualization through Neuroglancer (https://github.com/google/neuroglancer).[16] We specifically developed algorithms for distributed local intensity correction and cross-modal registration while leveraging existing state-of-the-art, open-source tools.[12,13] We applied CloudReg to various datasets including *in vivo* human brain MRI,[17] *ex vivo* macaque brain MRI,[18,19] *ex vivo* in situ mouse brain micro-CT,[20,21] and LSM-imaged cleared mouse and rat brains.[3,4,5]

Figure 1 shows an overview of the CloudReg pipeline. Data is uploaded to a web-accessible cloud storage provider; CloudReg is launched from a local machine (Figure 1A). The pipeline is run automatically in the cloud. CloudReg starts a cloud computing instance with sufficient RAM to perform pre-processing and registration, downloads raw data onto that server, corrects local intensity, stitches, corrects global intensity, and uploads pre-processed data to cloud storage for online visualization and analysis with Neuroglancer (Figure 1B). Next, registration is started by providing an affine initialization to roughly align an atlas to the input data and is automatically updated to include non-linear deformation via an expectation-maximization optimization process. Then, the atlas anatomic parcellations are automatically transformed to the input data at high resolution for visualization (Figure 1C). All the data is stored in the cloud, and then routed through a content delivery network and firewall to facilitate efficient and secure visualization via Neuroglancer (Figure 1D). The resulting data can be shared by sending a Universal Resource Locator (URL) and visualized from anywhere with a web browser and internet connection (example URL in Supplemental Note 1). Our deployment of Neuroglancer enables instant visualization of multi-channel, terabyte-scale datasets and provision of a shortened URL with one-click to share data views and analysis results (Supplemental Figure 1).[16]

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In the next two paragraphs, we elaborate on the pre-processing and registration steps (Figure 1B and 1C, respectively). We initially developed CloudReg using high-resolution, LSM-imaged CLARITY mouse brain data[6] and used the Allen Reference Atlas (ARA) Common Coordinate Frame Version 3 (CCFv3) as the reference atlas (Figure 2, rows 1 and 2).[22]

Tissue clearing procedures and optics of LSM introduce sample-specific artifacts and intensity inhomogeneity in the imaged samples, which we aimed to correct with our mFOV-based pre-processing algorithm. Intensity inhomogeneity manifests as a decay of intensity away from the center of each FOV and, for mFOV samples, the center of the stitched whole brain. This makes automatic, intensity-based registration a significant challenge. To minimize this per-FOV artifact, we developed a parallelized intensity correction algorithm based on the hypothesis that the introduced intensity inhomogeneity is the same in each FOV. To efficiently compute this correction, we uniformly subsample the mFOV data in three dimensions, compute the mean across subsampled FOVs in parallel, and apply the N4 bias correction algorithm[12] to the resulting mean FOV (Supplemental Figure 2A). Our intensity correction algorithm accounts for differences in tissue scattering from different clearing methods by estimating the intensity correction directly from the data. This pre-processed data is then automatically aligned and stitched using Terasticher, an open-source tool for stitching teravoxel microscopy datasets.[8] To minimize intensity inhomogeneity at the whole-brain scale, we apply the N4 bias correction algorithm to the whole stitched volume directly (Supplemental Figure 2B-C).

The fully pre-processed sample, which can be acquired from a variety of modalities in a number of species, is then registered to a corresponding reference atlas. To enable registration of LSM-imaged tissue samples, we developed a spatially-varying polynomial intensity transform, expanding the scope of samples that can be automatically registered (Figure 2; Supplemental Figure 3, Supplemental Video 1). CloudReg computes affine and nonlinear transformations by building on the Expectation-Maximization Large Deformation Diffeomorphic Metric Mapping (EM-LDDMM) registration algorithm that we previously developed.[15] Our extension of EM-

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

LDDMM built into CloudReg enables cross-modal registration of a diversity of brain volume samples with artifacts, tears, and deformations (Supplemental Figure 4).

To assess registration accuracy, we used Target Registration Error (TRE) by computing the Euclidean distance between 19 landmarks (Supplemental Figure 5) placed by experts on the ARA and our samples, where possible. TRE for samples 1, 2, and 3 was 2.27 ± 0.86 voxels, 2.42 ± 1.31 voxels, and 2.74 ± 0.88 voxels, respectively (Supplemental Table 1). These voxels are 100 µm indicating the resolution at which the registration was performed. Measurements and summary statistics can be found in Supplemental Table 1. Examining landmark error by brain region shows regional error differences, reflecting gross regional deformations. Mean landmark error for midbrain, cortical, and cerebellar regions across all samples are 1.47, 2.32, and 2.65 voxels, respectively. The region with the greatest error, the cerebellum, was most often subject to gross deformation including rotation and translation relative to the rest of the brain. By comparison, the midbrain, a relatively fixed structure in the brain and scan, had the lowest error. However, we found that the midbrain is most subject to the two artifacts typical of tissue clearing and LSM methods: intensity inhomogeneity and hydrogel-based deformation. Our TRE demonstrates that CloudReg handles artifacts typical of hydrogel-based tissue clearing methods including intensity inhomogeneity, hyperlocal structural deformation (nonuniform, micron-scale deviation from true anatomic position), and local missing tissue exceedingly well. CloudReg achieves this by relying on a rough affine initialization. Thus, in gross regions of the brain that have been displaced relative to the rest of the intact brain, the error will increase.

Registration is a crucial first step in analyzing a single or cohort of samples but can be more informative if combined with additional downstream analysis methods including cell and axon detection. A potential extension of our current work will be to accelerate the registration component of the code by leveraging our existing C++ implementations (https://github.com/InsightSoftwareConsortium/ITKNDReg).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

CloudReg can correct intensity, align and stitch, register, and visualize terabyte-scale brain volumes with artifacts and tears. CloudReg is immediately applicable to brain volumes spanning a variety of species and imaging modalities including mouse, rat, monkey, and human brain imaging.

Methods

All methods and additional supplementary information are available in the online e-methods.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgments

The authors would like to thank NeuroNex and Microsoft Research for supporting this work.

Author contributions

VC conceived the study, drafted the original manuscript, performed data analysis, built reproducible pipeline, revised the manuscript. DJT conceived the study, revised the manuscript, performed data analysis, and supervised data analysis. DC ported registration algorithm to Python, performed data analysis, and revised the manuscript. AKC, MAW, BYH, FG, TAM generated CLARITY/iDISCO LSM data and revised the manuscript. JSR helped draft the original manuscript, revised the manuscript, performed experiments, generated data, and placed landmarks for determining registration accuracy. KD conceived the study, supervised data analysis, and revised the manuscript. JTV conceived the study, revised the manuscript, and supervised data analysis.

Competing interest statement

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. Zhuo, J & Gullapalli, RP. MR Artifacts, Safety, and Quality Control. Radiographics. 26, 275–297 (2006).
2. Barrett, JF & Keat, N. Artifacts in CT: Recognition and Avoidance. Radiographics. 24, 1679–1691 (2004).
3. Chung, K. et al. Structural and molecular interrogation of intact biological systems. Nature 497, 332–337 (2013).
4. Park, Y.-G. et al. Protection of tissue physicochemical properties using polyfunctional crosslinkers. Nat. Biotechnol. 37, 73–83 (2018).
5. Renier, N. et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell 159, (2014).
6. Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. Nat. Protoc. 9, (2014).
7. Chung, K. & Deisseroth, K. CLARITY for mapping the nervous system. Nat. Methods 10, (2013).
8. Bria, A. & Iannello, G. TeraStitcher - a tool for fast automatic 3D-stitching of teravoxel-sized microscopy images. BMC Bioinformatics 13, (2012).
9. Goubran, M. et al. Multimodal image registration and connectivity analysis for integration of connectomic data from microscopy to MRI. Nat. Commun. 10, (2019).
10. Renier, N. et al. Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell 165, (2016).
11. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. BigDataViewer: visualization and processing for large image data sets. Nat. Methods 12, (2015).
12. Tustison, N. J. et al. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

13. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, (1998).
14. Faisal Beg, M., Miller, M. I., Trouvé, A. & Younes, L. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. Int. J. Comput. Vis. 61, 139–157 (2005).
15. Tward, D. et al. Diffeomorphic Registration With Intensity Transformation and Missing Data: Application to 3D Digital Pathology of Alzheimer's Disease. Front. Neurosci. 14, (2020).
16. Vogelstein, J. T. et al. A community-developed open-source computational ecosystem for big neuro data. Nat. Methods 15, (2018).
17. Mori, S., Wu, D., Ceritoglu, C., Li, Y., Kolasny, A., Vaillant, M.A., Faria, A.V., Oishi, K. and Miller, M.I., 2016. "MRICloud: Delivering High-Throughput MRI Neuroinformatics as Cloud-Based Software as a Service". Computing in Science & Engineering, 18(5), pp.21-35. doi:10.1109/MCSE.2016.93
18. Feng, L et al. Population-averaged macaque brain atlas with high-resolution ex vivo DTI integrated into in vivo space. Brain Struct Funct. 2017 Dec;222(9):4131-4147. doi: 10.1007/s00429-017-1463-6. Epub 2017 Jun 20. PMID: 28634624; PMCID:PMC5687984.
19. Ratnanather JT et al. Cortico-cortical, cortico-striatal, and cortico-thalamic white matter fiber tracts generated in the macaque brain via dynamic programming. Brain Connect. 2013;3(5):475-90. doi:10.1089/brain.2013.0143. Epub 2013 Sep 18. PMID: 23879573; PMCID:PMC3798182
20. Rosenblum, JS et al. Neuraxial dysraphism in EPAS1-associated syndrome due to improper mesenchymal transition. Neurol Genet Jun 2020, 6 (3) e414; DOI: 10.1212/NXG.0000000000000414
21. Lesciotto, KM et al. Phosphotungstic acid-enhanced microCT: Optimized protocols for embryonic and early postnatal mice. Dev. Dyn. 249(4), 573–585 (2020).
22. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell 181, (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure Legends

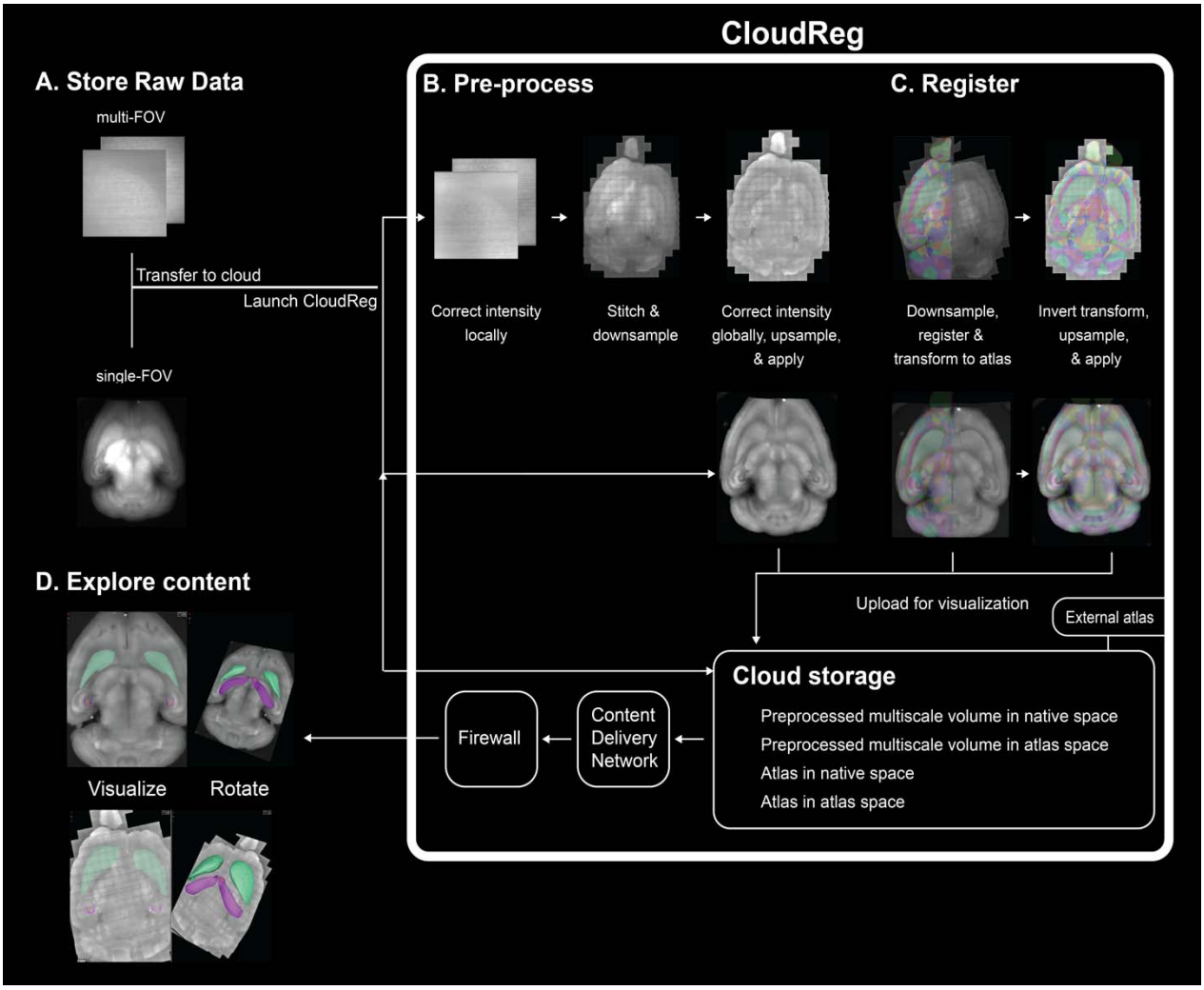


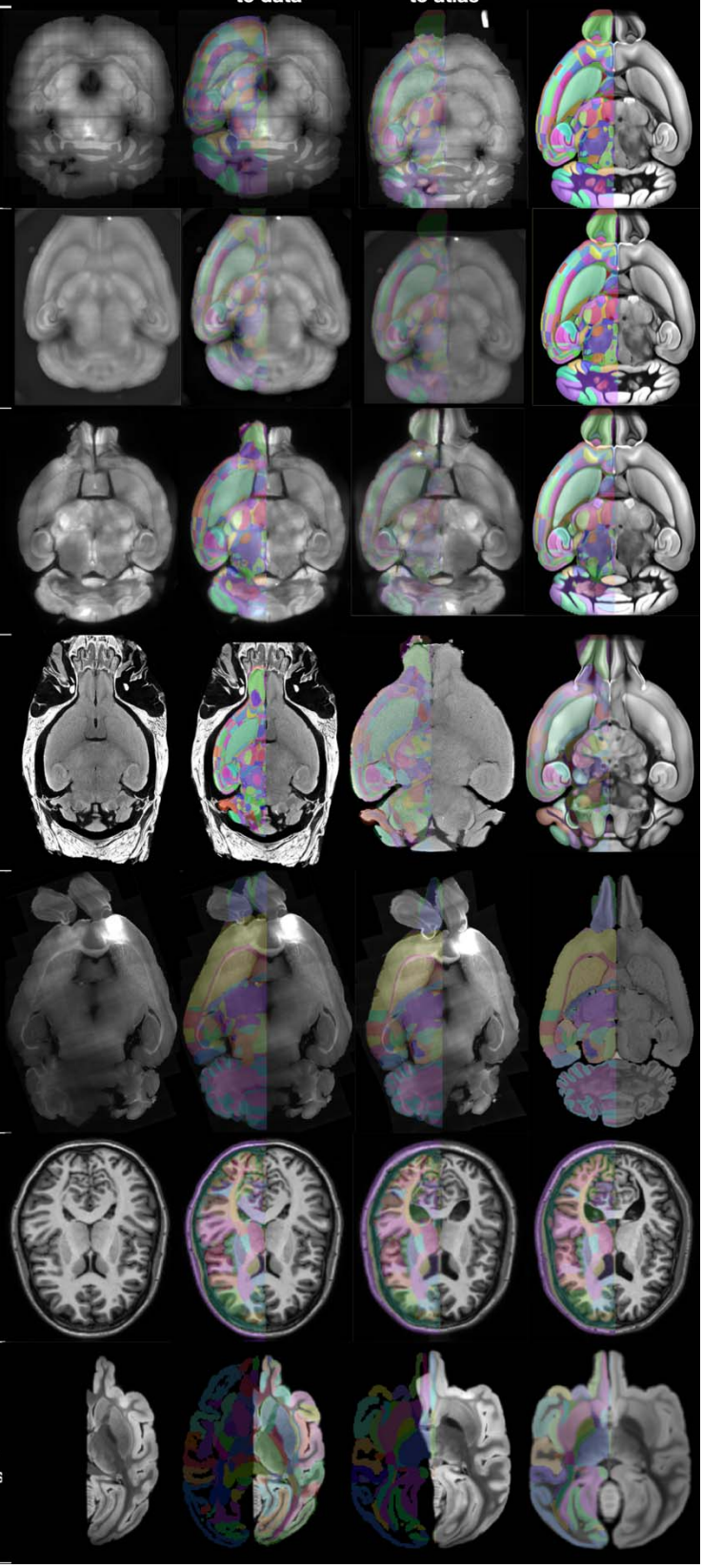
Figure 1. CloudReg pipeline schematic, example outputs at each step. A: Store raw data. Raw data can be either multi-field-of-view (mFOV) or single-field-of-view (sFOV). After raw data is generated, it is transferred to the cloud and CloudReg is launched. *B:* Pre-process. All of the following steps are performed on a cloud compute server that has access to the raw data. First, local intensity inhomogeneity is corrected per-FOV using an algorithm we developed. Next, mFOV data is aligned and stitched into a complete volume using Terastitcher; if the data is sFOV, this step is not performed. Then, global intensity inhomogeneity is corrected on the stitched whole brain by computing an intensity correction on the downsampled data, upsampling the result, and applying it at native resolution; this step applies to both mFOV and sFOV data.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The pre-processed data is then downsampled and uploaded to cloud storage for visualization. *C*: Register. A downsampled version of the pre-processed data is registered to the ARA CCFv3 and the computed transformations are saved. These transformations are invertible so the atlas can be transformed to the data space and the data can be transformed to the atlas space. The transformations are upsampled and applied to the ARA anatomic parcellations, and input data and are uploaded to cloud storage for visualization through Neuroglancer. To maintain privacy of imaging data, visualization is restricted to authorized users by using a content delivery network and firewall. *D*: Explore content. Left column shows ARA parcellations transformed to the s- and mFOV data based on the computed transformations from the registration. Right column shows 2D axial slice from the pre-processed input data and 3D rendering of Caudoputamen and Hippocampal regions selected from the transformed ARA visualized in Neuroglancer. ARA CCFv3, Allen Reference Atlas Common Coordinate Frame version 3.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Modality	Species	Data size, Voxel size (X,Y,Z)	Reference atlas name
CLARITY, COLM	Mouse	1.11 TB, 0.585 µm x 0.585 µm x 5 µm	Allen Reference Atlas
CLARITY, LaVision	Mouse	16.07 GB, 5.16 µm x 5.16 µm x 5.16 µm	Allen Reference Atlas
iDISCO, COLM	Mouse	166.80 GB, 1.46 µm x 1.46 µm x 5 µm	Allen Reference Atlas
in situ MicroCT	Mouse	33.33 GB, 13.53µm x 13.53µm x 13.53µm	Allen Reference Atlas
CLARITY, COLM	Rat	291.45 GB, 2.92 µm x 2.92 µm x 5 µm	Waxholm atlas
MRI	Human	14.22 MB, 1 mm x 1mm x 1 mm	MRICloud atlas
MRI	Macaque	16.29 MB, 400 µm x 400 µm x 400 µm	Macaque population-average atlas



1

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.26.428355; this version posted January 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2. CloudReg pipeline registration outputs from multiple species imaged with various modalities. Each row demonstrates registration of either mouse, rat, macaque, or human brain imaging data to the corresponding atlas using CloudReg. The leftmost column of images shows the input data; the data from the autofluorescence channel is used for samples imaged with a light-sheet microscope (LSM). The rightmost column shows the atlas parcellations overlaid on one hemisphere of the atlas image data. The second and third columns show the respective atlas parcellations transformed to and overlaid on the original samples and vice-versa, respectively. CLARITY, Clear Lipid-exchanged Anatomically Rigid Imaging/immunostaining-compatible Tissue Hydrogel; COLM, CLARITY-Optimized Light-sheet Microscopy; GB, Gigabyte; iDISCO, immunolabeling-enabled three-dimensional imaging of solvent-cleared organs; MB, Megabyte; Micro-CT, Micro-Computed Tomography; TB, Terabyte.