CellPress

Neuron
NeuroView

Big Science, Team Science, and Open Science for Neuroscience

Christof Koch[1,*] and Allan Jones[1]
[1]Allen Institute for Brain Science, Seattle, WA 98109, USA
*Correspondence: christofk@alleninstitute.org
http://dx.doi.org/10.1016/j.neuron.2016.10.019

The Allen Institute for Brain Science is a non-profit private institution dedicated to basic brain science with an internal organization more commonly found in large physics projects—large teams generating complete, accurate and permanent resources for the mouse and human brain. It can also be viewed as an experiment in the sociology of neuroscience. We here describe some of the singular differences to more academic, PI-focused institutions.

> But the greatest paradox of the sport has to do with the psychological makeup of the people who pull the oars … Great crews may have men or women of exceptional talent and strength; they may have outstanding coxswains or stroke oars or bowmen; but they have no stars. The team effort—the perfectly synchronized flow of muscle, oars, boat and water; the single, whole, unified and beautiful symphony that a crew in motion becomes—is all that matters. Not the individual, not the self.—Daniel James Brown (*The Boys in the Boat*)

The field of biomedical science enjoys worldwide prestige, notable triumphs, and significant funding. It has set itself the goal of understanding biological organisms small and large, in health and in disease. But the field also faces two acute and severe challenges. The first—the vast complexity of organisms that it seeks to understand—is a fundamental feature of evolved organisms. The second challenge is the growing recognition that many of the conclusions drawn from biomedical research are unreliable, and cannot be reproduced. This replication crisis is pernicious and could endanger the long-term public support, particularly given the prevailing anti-intellectual and anti-expert political environment.

Successfully tackling both challenges is of the essence, in particular if the promise underpinning the various international brain projects to diagnose, ameliorate, and ultimately cure mental diseases and disorders is to be realized in our lifetime. This will require complementing the traditional, small laboratory culture that has hitherto driven almost all biological discovery and that reward individual investigators and their fecundity in publishing papers, with large, multi-disciplinary teams working under highly reproducible standards and making all of their methods, data, and metadata publicly available. This latter model is what the Allen Institute for Brain Science seeks to achieve with its focus on Big Science, Team Science, and Open Science. We here describe some of our experiences and lessons learned.

Biomedical science seeks to answer an array of diverse questions, such as deciphering the fates of individuals from their genomes, linking the microbiome to lifestyle and disease, diagnosing and developing rational therapies for autism spectrum disorders, and understanding how the conscious mind emerges from the flickering activity of a dizzying number of heterogeneous nerve cells. Addressing these questions requires the design, construction, and operation of an armamentarium of sophisticated tools and methods from a variety of scientific and technological disciplines, in addition to software engineers and computer scientists to grapple with the massive data streams to extract and analyze key parameters, and data scientists and theoreticians to fit these data to digital simulacra and theories. Biology has undergone a dramatic transformation since Gregor Mendel, the founder of genetics, used gardening tools to discover the laws of Mendelian inheritance. Just 150 years later, the Human Genome Project involved a far-flung coalition of technicians and scientists with their sophisticated machinery. This singular project brought the era of Big Science to biology.

It was the construction of large particle accelerators at the Radiation Laboratory in Berkeley under Ernest Lawrence in the 1930s that gave birth to Big Science. Growing up in the shadow of the Manhattan project, Big Science came of age post-Sputnik, during the Cold War, fueled by massive government programs. Its rationalization was the discovery of unknown, exotic high energy particles ("new" physics) and the testing of specific theories, such as the Standard Model at the LHC or General Relativity at LIGO. In astronomy, large telescopes and planetary probes opened up new windows into the cosmos or investigated planets, moons, or other denizens of the solar system. This motto of Big Science as accessing new frontiers and surveying the landscape of possibilities carried into the Human Genome Project and its follow-up ENCODE project. This is also the spirit in which the Allen Institute views its contribution to neuroscience.

Big Science is imposed by the complexity of the phenomena investigated and by the prolificacy and intricacies of modern instruments. Thus, developing tests to detect and therapies to stop and perhaps even reverse the ravages visited upon an adolescent brain by schizophrenia, one gifted professor working with her graduate student and postdoctoral fellow in isolation will not tame the vast beast that is the genome and

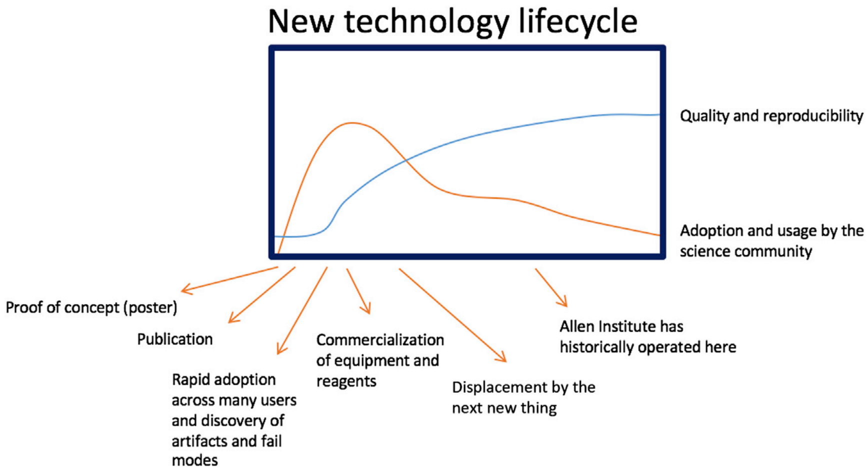

Figure 1. Life Cycle of a Generic Technology in Basic Science
An important lesson we have learned many times is the need to pick state-of-the-art instrumentation technologies that are sensitive and reliable, with well-understood operational procedures that can operate stably over years for our large neuroscience data products. This may preclude adopting the latest cutting-edge technique that has just been published as a proof of principle. This is schematically illustrated with the time line for some of the canonical events in the life cycle of some relevant technology. The ordinate plots penetration (usage and adoption by the science community) in red and quality and reproducibility of the technology in blue.

the brain (see, for instance, the long author list in Sekar et al. 2016). Understanding how brains work in health, and how they break down in disease, demands Big Science to complement the traditional discovery process. Big Science, though, is a significant departure from the investigator-focused culture of small and autonomous laboratories and requires fresh thinking on how individual scientists are rewarded, teams are motivated, and research is planned, paid for, and executed.

Big Science

At the beginning of the millennium, Paul Allen, co-founder of Microsoft, assembled a group of biologists to discuss the future of neuroscience and what could be done to accelerate its course. From these meetings emerged the idea of generating a complete atlas detailing gene expression throughout the brain of the most popular mammalian model system, the laboratory mouse. Funded by an initial gift of $50 million by Mr. Allen, a team of about 50 people accomplished this feat in 2006, under budget and on time, using in situ hybridization for mapping the expression of 20,000 genes in sectioned brain tissue. The annotated data were, and continue to be, freely available to anybody with an internet connection at http://brain-map.org. This effort was powered by a "structured science" approach adopted from the biotechnology industry, including robotic instrumentation, standardized operating procedures to reduce variability, high quality control and quality assurance standards and extensive program management, supplemented by external scientific advisory boards.

The Allen Mouse Brain Atlas is now the de facto standard for visualizing gene expression patterns in the mouse brain, with more than 2,000 citations to the associated publication (Lein et al. 2007) and more than 15,000 unique monthly visitors to the online resource. Indeed, it could be argued that the Allen Mouse Brain Atlas satisfies Sydney Brenner's famous CAP criteria pertaining to large-scale surveys: *complete* coverage of all genes throughout the brain, *accurate* probes (i.e., with high sensitivity and specificity), and maintained as a curated and *permanent* resource.

Building on the success of the Allen Mouse Brain Atlas, the Allen Institute refined and adapted this high-throughput, scalable, and robust platform to create additional cartographies, comprehensively mapping transcriptional expression patterns within a three-dimensional coordinate system in the developing mouse brain (Thompson et al., 2014), the adult mouse spinal cord, the adult human brain (Hawrylycz et al., 2012), the developing human brain (Miller et al., 2014), the adult macaque brain (Bernard et al., 2012), and the developing macaque brain (Bakken et al., 2016). These massive online resources are complemented by a mesoscopic mouse connectivity atlas (Oh et al., 2014) and a cellular-resolution (1 μm/pixel) annotated atlas of an adult human brain (Ding et al., 2016).

These large-scale and comprehensive projects could not have thrived within the academic ecosystem, with its need for rapid results and specific hypotheses to sustain funding and multiple first-authored publications for participating graduate students to obtain their PhDs. The constant refinement and adjustment of tools, instruments, and analysis methods by multiple actors without documentation is also a hindrance for big science projects. Of course, technology development propels scientific discovery, but if the dynamics at which such improvements occur are faster than the project timeline, the goal of obtaining a canonical and reproducible product that can serve as a gold standard becomes elusive.

We have learned several valuable lessons over the years. Critical to success is testing competing technologies and instruments under realistic conditions to assess their robustness and reliability and insisting on firm deadlines. Bringing a technology platform from proof-of-concept to maturity requires planning for an extended time course. Finally, it is important to resist the siren song of constantly switching to the latest "hot" method in favor of staying the course with a well-proven, older technology until large-scale data production is completed (Figure 1).

Team Science

Large science projects require a highly specialized workforce that works together, day after day, hand-in-glove, across the organization. Consider the team that operates our latest online offering, the Allen Brain Observatory. Its goal is to record the cellular level activity of thousands and ultimately millions of neurons in functionally identified regions of the mouse brain as the animal is engaged in a stereotyped behavior. Our first release (http://observatory.brain-map.org/) in the open-source *Neurodata Without Borders* data format (Teeters et al., 2015) features the calcium activity of more than 30,000 individual neurons responding to a battery of visual stimuli, including static and moving gratings, sparse noise, natural scenes, and the movie *Touch of Evil*. Activity is recorded via two-photon microscopy in primary visual cortex and nearby regions in different transgenic mice and mapped to a high-resolution neuroanatomical coordinate framework. Together with associated metadata, including eye movements, video of the running mice and so on, this first dataset exceeds 30TB.

The team that built and operates the Allen Brain Observatory numbers one hundred specialists and technicians (not all of whom work full-time on this project)—technicians to care for and train the animals, neurobiologists to plan and execute the behavioral experiment, electrical and optical engineers to construct and maintain the microscopes to identify functional regions and image cellular-level activity, neurosurgeons to precisely place transparent windows into murine skulls, mechanical engineers to build the gimbals, reticules, and other widgets to reproducibly return to the same set of neurons over multiple trials, anatomists and annotators to localize responses within the 3D brain, software engineers to harness and massage the massive data stream from all instruments into a common laboratory information management system, data scientists to extract the tiny fraction of relevant information, and modelers and theoreticians to carry out statistical analyses and build neural network and other models to analyze and replicate the data. Four years in the making, the project has been fully funded by the generosity of one individual, Paul Allen.

The difficulties in operating such an observatory are not just those associated with setting up any advanced instrumentation suite, but also include significant sociological and organizational challenges.

We all treasure a sense of autonomy. Yet inherent to any team, whether operating on the battle field, on the water racing a scull, in a start-up, or at an institute with production deadlines, is the imperative to align on the agreed-upon goals and the attendant need to submerge the ego, the self, to the group as a whole. These conflicting demands have important implications for internal decision making and highlight the absolute need for the development and maintenance of trust across the entire team. Achieving these goals requires a sophisticated meeting culture (e.g., taking minutes, specific agenda, follow-up, starting and ending on time, no distracting smartphone or laptop usage, and so on). Regular team meetings are the most visible difference to academic culture.

These challenges grow as the size of the team grows. Our anecdotal evidence suggests that above a hundred members, group cohesion appears to become weaker with the appearance of semi-autonomous cliques and sub-groups. This may relate to the postulated limit on the number of meaningful social interactions humans can sustain given the size of their brain (Dunbar, 1992).

Teams are built on the strengths and abilities of their members. While alpha-type personalities pursuing their own idea thrive in an academic setting, we are more dependent on team players with a highly cooperative style of give and take in an environment that unleashes their energy and creativity.

Within this team-oriented context, how should individual contributors, on whose creativity, brilliance, dedication, discipline, and hard work the entire enterprise rests, be rewarded? As a non-profit institution, we do not offer stock options or outsized salaries to reward strong performers. Likewise, the promise of first or senior authorship only acts as a weak draw given the large number of contributors.

This means that internal motivators are key—team members value the knowledge that they are participating in an historic mission at the frontier of science that will unearth new knowledge to the benefit of all of humankind.

Such considerations require a judicious recruitment process, ample opportunities for growth, and promotion within the organization to reward performance, the nurturing of a sense of being part of something larger than oneself, and distinct tracks for employees interested in careers in basic research (scientists I, II, senior scientist, and investigators), structured science (managers and directors of distinct rankings), or science management (project and program managers).

Fortunately, the physics community has shown the way by assembling colossal teams. The two independent collaborations at the Large Hadron Collider (LHC) at CERN in Geneva that successfully hunted for the Higgs Boson—CMS and ATLAS—each has about 3,000 participants with elaborate, point-based rules for authorship and who can speak for the group. The authors list of the LIGO consortium publication announcing the first detection of gravity waves emitted by two merging black holes was strictly alphabetical and included about a thousand individuals from 133 institutions (Abbott et al., 2016). Note that these large organizations happily and productively coexist with traditional academia. Thus, while there can be upward of 13,000 people on site at CERN in Geneva, only about 2,200 of these are CERN employees; the rest are subcontractors, students, fellows, and visiting faculty.

We have learned from these and other organizations that to build a well-functioning team with a strong esprit de corps, it is critical to jointly and cooperatively align on specific common goals, build trust by open and transparent decision-making, a maximal sharing of both responsibilities and credit, and nourish morale by a dense web of formal and informal means of communication. We have also learned the importance of compromise among conflicting demands and experts from different scientific traditions. This is really nothing but the art of the achievable under time and budgetary constraints—succinctly expressed by Otto von Bismarck's dictum, "Politics is the art of the possible."

Open Science

From the early days of the Institute, we have made the fruits of our investigations available to anybody with an internet connection through our repository at http://brain-map.org. These massive resources totaling more than 3,000 Terabytes of data are actively maintained and curated. No login or registration step is needed to browse, search, view, or download any of this wealth of information via our dedicated web tools. Data releases occur well ahead of the publication

Box 1. Open Science Recommendation

To accelerate the rate of discovery and to ameliorate the replication crisis, it is imperative that the neuroscience community moves toward an open science ethic. After all, almost all basic science is funded, directly or indirectly, by citizens via their taxes. Therefore, the fruits of these labors ought to be publicly, freely, and widely shared. We offer some specific suggestions.

- The detailed code for the complete analysis of data and statistical procedures should accompany every publication. The most convenient form is a *jupyter* notebook (https://jupyter.org/), a Python-based web application for the creation and sharing of documents that contains live code, equations, figures, and explanatory text. This allows researchers to easily replicate—and vary—the conclusion of the study. This imposes a low burden on authors.
- We do know from our own experience that making data freely available imposes a significant burden (for a thoughtful discussion of the costs and the benefits, see Choudhury et al. 2014). Data and relevant metadata need to be formatted into a common data format and placed online. This repository needs to be curated. Increasingly, funding agencies are receptive to such initiatives and the attendant costs. With rare exceptions for singular findings, the era of illustrating discoveries via nothing but a flat PDF file, with ''representative results'' that are often the most expected or cleanest responses, ought to be coming to a close.

of associated platform papers. Our current 10-year plan calls for continual thrice yearly releases of data. We also share white papers describing in minute detail our operating procedures and methods and many of our tools, such as transgenic animals and viruses (e.g., Madisen et al., 2010). Indeed, The Jackson Laboratory has shipped more than 20,000 of our transgenic mice to customers worldwide.

Why do we pursue such an ''information-wants-to-be-free'' policy? The simplest answer is that this is the vision of Paul Allen. His intent is that Institute resources should accelerate neuroscience discovery. He also believes that the Institute should act as an example of what every laboratory should be doing.

For millennials, communicating via texts and images is part of their social online experience growing up. Thus, it comes natural to young scientists to freely and openly share data, computer code, and manuscripts. Indeed, open science has been advocated for and practiced by a small number of neuroscientists for more than a decade (e.g., Van Horn and Gazzaniga, 2013; Glasser et al., 2016). Regrettably, however, the vast majority of data and metadata in the neurosciences continues to remain inaccessible.

It is urgent that the field as a whole move toward an open science policy, as it will alleviate some of the root causes of the replication crisis so pervasive in biomedical research. A recent review on this topic (Button et al., 2013) opens with an eye-catching, ''It has been claimed and demonstrated that many (and possibly most) of the conclusions drawn from biomedical research are probably false.'' Estimates for the fraction of false findings range from a simple majority to 80% and higher (Open Science Collaboration, 2015). Given that replication is one of the key steps in the scientific process, its systematic failure for so many published findings constitute a striking departure from good scientific practice that could come back to haunt the field. This cannot be eluded by blithely ignoring it as most scientists are wont to do.

There are two broad categories of causes for the replication crisis.

First, even simple organisms have vastly more degrees of freedom than the natural, non-evolved systems physicists typically deal with, such as elementary particles, gravity waves, or exo-planets. Thus, while a Higgs Boson has the same signature no matter where on Earth it is detected, a C57BL/6J mouse in Seattle will not be the same as a laboratory mouse in Boston (Crabbe et al., 1999). What is true for standard breeds of mice kept under controlled conditions is vastly more so for ''neuro-typical'' volunteers or patients. Mitigating this problem will not be easy but must include enforcing, to the maximal extent possible, standardization of model systems, procedures, tools, and instruments, as well as meticulously publicizing all logistical and methodological details. Unbiased surveys eliminate some sources of these statistical infelicities and have found widespread adoptions in some fields. For example, within astronomy, generations of *Sloan Digital Sky* surveys have resulted in more than 5,000 refereed publications and 200 million SQL queries to the relevant database and revolutionized the field in the process (Burns et al., 2014).

A second set of causes relate to lack of statistical ''hygiene,'' including low statistical power, incorrect noise models, publication bias (or ''the file drawer problem''), hypothesizing after the results are known (HARKing), and p-hacking. These phenomena have been much commented upon (Ioannidis 2005; Button et al. 2013). The latest scandal follows the re-analysis of resting-state fMRI data from the public database *Functional Connectomes Project* (Eklund et al., 2016), which revealed that widely used analysis software packages yield false-positive rates of up to 70%. This may affect more than 3,000 published fMRI studies, funded at great expense by the public purse.

There is no question that making all data and metadata, together with all computational and statistical procedures of every publication freely and publicly available would address some of these shortcomings (see Box 1). Such a policy allows researchers to easily test whether the published conclusions are valid, particularly when they conflict with other studies. Given the potential threat to the reputation of the authors, Open Science is likely to lead to statistically more valid results than otherwise. Open access also allows for analysis of published data using different algorithms and assumptions to gain additional or alternative insights, or to ask questions that were not conceived by the original authors. Finally, open data sharing enables the aggregation and contrasting of data across multiple studies for meta-analyses to extract broader insights. Quite simply, Open Science is the right thing to do.

The Allen Institute for Brain Science has proven that Big Science, Team Science, and Open Science can be harnessed to create extraordinary neuroscientific resources that benefit all. With bright eyes, we look toward a future in which we, together with the world-wide community of researchers from individual laboratories at universities and independent research institutes, will decipher the most highly organized piece of excitable matter in the known universe: the human brain.

ABOUT THE AUTHORS

Christof Koch is the President and Chief Scientific Officer of the Allen Institute for Brain Science. After spending 27 years as a faculty member at the California Institute of Technology, he joined the Institute in 2011. He seeks to understand the workings of the mammalian cerebral cortex and how it generates conscious experience. Allan Jones is the President and Chief Executive Officer of the Allen Institute. He joined the Institute from biotech/pharma at the beginning of the Allen Brain Atlas project in 2003 and is passionate about understanding complexity in biology.

ACKNOWLEDGMENTS

We thank Anton Arkhipov, Mike Hawrylycz, Hannah Krakauer, Ed Lein, John Phillips, Erica Sessle, and Hongkui Zeng for thoughtful comments and feedback. None of this work could have been accomplished without the vision and the enduring support of our founders, Paul G. Allen and Jody Allen. A.J. is the head of the overall Allen Institute, and C.K. is the head of its daughter, the Allen Institute for Brain Science.

REFERENCES

Abbott, B.P., Abbott, R., Abbott, T.D., Abernathy, M.R., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R.X., et al.; LIGO Scientific Collaboration and Virgo Collaboration (2016). Phys. Rev. Lett. *116*, 061102.

Bakken, T.E., Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., et al. (2016). Nature *535*, 367–375.

Bernard, A., Lubbers, L.S., Tanis, K.Q., Luo, R., Podtelezhnikov, A.A., Finney, E.M., McWhorter, M.M., Serikawa, K., Lemon, T., Morgan, R., et al. (2012). Neuron *73*, 1083–1099.

Burns, R., Vogelstein, J.T., and Szalay, A.S. (2014). Neuron *83*, 1249–1252.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., and Munafò, M.R. (2013). Nat. Rev. Neurosci. *14*, 365–376.

Choudhury, S., Fishman, J.R., McGowan, M.L., and Juengst, E.T. (2014). Front. Hum. Neurosci. *8*, 239.

Crabbe, J.C., Wahlsten, D., and Dudek, B.C. (1999). Science *284*, 1670–1672.

Ding, S.-L., Royall, J.J., Sunkin, S.M., Ng, L., Facer, B.A., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T.A., et al. (2016). J. Comp. Neurol. *524*, 3127–3481.

Dunbar, R.I.M. (1992). J. Hum. Evol. *22*, 469–493.

Eklund, A., Nichols, T.E., and Knutsson, H. (2016). Proc. Natl. Acad. Sci. USA *113*, 7900–7905.

Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L., Auerbach, E.J., Behrens, T.E., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., et al. (2016). Nat. Neurosci. *19*, 1175–1187.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). Nature *489*, 391–399.

Ioannidis, J.P.A. (2005). PLoS Med. *2*, e124.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Nature *445*, 168–176.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). Nat. Neurosci. *13*, 133–140.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Nature *508*, 199–206.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). Nature *508*, 207–214.

Open Science Collaboration (2015). Science *349*, http://dx.doi.org/10.1126/science.aac4716.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2016). Nature *530*, 177–183.

Teeters, J.L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., et al. (2015). Neuron *88*, 629–634.

Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., et al. (2014). Neuron *83*, 309–323.

Van Horn, J.D., and Gazzaniga, M.S. (2013). Neuroimage *82*, 677–682.