[1]Eric Kan, [2]Julia Anglin, [3]Michael Borich, [4]Neda Jahanshad, [4]Paul Thompson, [5]Sook-Lei Liew*

Report from Brainhack

Facilitating Big Data Meta-Analyses for Clinical Neuroimaging through ENIGMA Wrapper Scripts

Project URL: https://github.com/npnl/ENIGMA-Wrapper-Scripts

[1]Eric Kan, [2]Julia Anglin, [3]Michael Borich, [4]Neda Jahanshad, [4]Paul Thompson, [5]Sook-Lei Liew*

[1]The Saban Research Institute of Children's Hospital, Los Angeles, California, USA, Department of Pediatrics of the Keck School of Medicine, University of Southern California, Los Angeles, California, USA [2]Chan Division of Occupational Science and Occupational Therapy, USC [3]Division of Physical Therapy, Department of Rehabilitation Medicine, Emory University School of Medicine [4]Imaging Genetics Center, Laboratory of Neuro Imaging, Keck School of Medicine of USC, University of Southern California [5]Chan Division of Occupational Science and Occupational Therapy, Division of Biokinesiology and Physical Therapy, Department of Neurology of the Keck School of Medicine, USC
*Correspondence: sliew@usc.edu Neural Plasticity and Neurorehabilitation Laboratory, Chan Division of Occupational Science and Occupational Therapy, Division of Biokinesiology and Physical Therapy, Keck School of Medicine Department of Neurology, University of Southern California, Los Angeles, 1540 Alcazar Street, CHP 133 MC 9003, 90089, California, USA

1 Introduction

A vast number of clinical disorders may involve changes in brain structure that are correlated with cognitive function and behavior (e.g., depression, schizophrenia, stroke, etc.). Reliably understanding the relationship between specific brain structures and relevant behaviors in worldwide clinical populations could dramatically improve healthcare decisions around the world. For instance, if a reliable relationship between brain structure after stroke and functional motor ability was established, brain imaging could be used to predict prognosis/recovery potential for individual patients. However, high heterogeneity in clinical populations in both individual neuroanatomy and behavioral outcomes make it difficult to develop accurate models of these potentially subtle relationships.

Large neuroimaging studies (n>10,000) would provide unprecedented power to successfully relate clinical neuroanatomy changes with behavioral measures. While these sample sizes might be difficult for any one individual to collect, the ENIGMA Center for

Worldwide Medicine, Imaging, and Genomics has successfully pioneered meta- and mega-analytic methods to accomplish this task. ENIGMA brings together a global alliance of over 500 international researchers from over 35 countries to pool together neuroimaging data on different disease states in hopes of discovering critical brain-behavior relationships (http://enigma.ini.usc.edu; Thompson et al., 2015). Individual investigators with relevant data run ENIGMA analysis protocols on their own data and send back an output folder containing the analysis results to be combined with data from other sites for a meta-analysis. In this way, large sample sizes can be acquired without the hassle of large-scale data transfers or actual neuroimaging data sharing.

ENIGMA protocols were initially developed to harmonize processing methods of imaging researchers around the world and they require a moderate level of familiarity with several programming languages and environments. However, ENIGMA's recent success has attracted greater interest in collaborative neuroimaging and protocols must be adjusted to allow for all levels of experience, as, the success of this approach depends on individual collaborators running these

User Level	Ease of Use (1-10) 1 = hardest, 10 = easies	Time to implement scripts	Notes
Novice 1	3	00:35:25	Required walk through support
Novice 2	3	00:52:55	Required support for basic terminal commands only; then was able to complete independently
Moderate 1	8	00:23:45	Required no support
Moderate 2	7	00:22:10	Required no support
Expert 1	8	00:11:34	Required no support
Expert 2 - different enviornment	8	02:00:00	Getting scripts to run took several minutes but reorganizing data and troubleshooting with freesurfer took significant time
Expert 3 - different enviornment	9	01:15:00	Required walk through support

ENIGMA protocols on their data. Here, we worked on simplifying these protocols so even a novice programmer could use them. In this way, we hope to expand the feasibility of collecting critical clinical data from collaborators who may have less experience with neuroimaging techniques.

2 Approach

The current ENIGMA protocols (http://enigma.ini.usc.edu/protocols/) for structural neuroimaging analyses consist of a number of different word documents with embedded links to different scripts and snippets of code that use R, bash scripting, Matlab, FSL, and Freesurfer. Each step must be run sequentially, costing the user time during the implementation to wait for each step to finish before beginning the next. In addition, the number of different scripts, programming languages and software environments can be challenging for a novice user and introduces numerous instances where the individual may make errors in implementing the code. To address this, we created a 3 easy-to-use wrapper scripts that automate the implementation of the ENIGMA protocols for both subcortical and cortical structural MRI analyses. These wrapper scripts reduce over 40 steps down to 3 quick steps for the user. We also created a user-friendly readme file that includes screenshots of the code implementation.

To examine the ease of use and time to implement the new scripts, we tested each of them on 6 users who had different levels of familiarity with programming and neuroimaging (novice users (no programming experience), moderate users (basic-to-intermediate programming experience), and expert users (extensive programming)). To explore additional factors relating to implementation, two of the expert users had to use the scripts in a different environment (e.g., organize the data, install the software, etc.).

3 Results/Discussion

Overall, moderate and expert level users found the scripts extremely easy to implement and required less than 25 minutes to get all three scripts running (excluding the run time of each script; see Table 1 for individual results). The two novice users required greater support to understand basic elements (e.g., what is a terminal), but with support, were able to complete all the steps in less than 1 hour. Finally, the expert users who had to reorganize the environment took 2-3 hours, but about 10 minutes was script implementation and the remainder of the time was troubleshooting Freesurfer errors.

The wrapper scripts made the implementation of the ENIGMA protocols quick and feasible even for novice users. However, there were still three main barriers to participation that required significant time, computational resources, and some expertise: 1) data organization (depending on previous data structure), 2) running freesurfer (~12 hrs/subj), and 3) installation of the required software (e.g., Freesurfer, FSL, R). Future projects may look at ways to streamline these areas for a more seamless user experience in order to facilitate greater sharing of clinical neuroimaging data through ENIGMA.

All scripts and user guides can be found at: https://github.com/npnl/ENIGMA-Wrapper-Scripts

4 References

ENIGMA and the Individual: Predicting Factors that Affect the Brain in 35 Countries Worldwide.Paul M. Thompson, Ole A. Andreassen, Alejandro Arias-Vasquez, Carrie E. Bearden, Premika S. Boedhoe, Rachel M. Brouwer, Randy L. Buckner, Jan K. Buitelaar, Kazima B. Bulaeva, Dara M. Cannon, Ronald A. Cohen, Patricia J. Conrod, Anders M. Dale, Ian J. Deary, Emily L. Dennis, Marcel A. de Reus, Sylvane Desrivieres, Danai Dima, Gary Donohoe, Simon E. Fisher, Jean-Paul Fouche, Clyde Francks, Sophia Frangou, Barbara Franke, Habib Ganjgahi, Hugh Garavan, David C Glahn, Hans J Grabe, Tulio Guadalupe, Boris A Gutman, Ryota Hashimoto,

Derrek P Hibar, Dominic Holland, Martine Hoogman, Hilleke E Hulshoff Pol, Norbert Hosten, Neda Jahanshad, Sinead Kelly, Peter Kochunov, William S Kremen, Phil H Lee, Scott Mackey, Nicholas G Martin, Bernard Mazoyer, Colm McDonald, Sarah E Medland, Rajendra A Morey, Thomas E Nichols, Tomas Paus, Zdenka Pausova, Lianne Schmaal, Gunter Schumann, Li Shen, Sanjay M Sisodiya, Dirk JA Smit, Jordan W Smoller, Dan J Stein, Jason L Stein, Roberto Toro, Jessica A Turner, Martijn van den Heuvel, Odile A van den Heuvel, Theo GM van Erp, Daan van Rooij, Dick J Veltman, Henrik Walter, Yalin Wang, Joanna M Wardlaw, Christopher D Whelan, Margaret J Wright, Jieping Ye, for the ENIGMA Consortium.NeuroImage, in press(in press), pp. 1-7. 2015.

Common genetic variants influence human subcortical brain structures. Derrek P. Hibar, Jason L. Stein, Miguel E. Renteria, Alejandro Arias-Vasquez, Sylvane Desrivis, Neda Jahanshad, Roberto Toro, Katharina Wittfeld, Lucija Abramovic, Micael Andersson, Benjamin S. Aribisala, Nicola J. Armstrong, Manon Bernard, Marc M. Bohlken, Marco P. Boks, Janita Bralten, Andrew A. Brown, M. Mallar Chakravarty, Qiang Chen, Christopher R. K. Ching, Gabriel Cuellar-Partida, Anouk den Braber, Sudheer Giddaluru, Aaron L. Goldman, Oliver Grimm, Tulio Guadalupe, Johanna Hass, Girma Woldehawariat, Avram J. Holmes, Martine Hoogman, Deborah Janowitz, Tianye Jia, Sungeun Kim, Marieke Klein, Bernd Kraemer, Phil H. Lee, Loes M. Olde Loohuis, Michelle Luciano, Christine Macare, Karen A. Mather, Manuel Mattheisen, Yuri Milaneschi, Kwangsik Nho, Martina Papmeyer, Adaikalavan Ramasamy, Shannon L. Risacher, Roberto Roiz-Santia Emma J. Rose, Alireza Salami, Philipp G. Snn, Lianne Schmaal, Andrew J. Schork, Jean Shin, Lachlan T. Strike, Alexander Teumer, Marjolein M. J. van Donkelaar, Kristel R. van Eijk, Raymond K. Walters, Lars T. Westlye, Christopher D. Whelan, Anderson M. Winkler, Marcel P. Zwiers, Saud Alhusaini, Lavinia Athanasiu, Stefan Ehrlich, Marina M. H. Hakobjan, Cecilie B. Hartberg, Unn K. Haukvik, Angelien J. G. A. M. Heister, David Hoehn, Dalia Kasperaviciute, David C. M. Liewald, Lorna M. Lopez, Remco R. R. Makkinje, Mar Matarin, Marlies A. M. Naber, D. Reese McKay, Margaret Needham, Allison C. Nugent, Benno Ptz, Natalie A. Royle, Li Shen, Emma Sprooten, Daniah Trabzuni, Saskia S. L. van der Marel, Kimm J. E. van Hulzen, Esther Walton, Christiane Wolf, Laura Almasy, David Ames, Sampath Arepalli, Amelia A. Assareh, Mark E. Bastin, Henry Brodaty, Kazima B. Bulayeva, Melanie A. Carless, Sven Cichon, Aiden Corvin, Joanne E. Curran, Michael Czisch, Greig I. de Zubicaray, Allissa Dillman, Ravi Duggirala, Thomas D. Dyer, Susanne Erk, Iryna O. Fedko, Luigi Ferrucci, Tatiana M. Foroud, Peter T. Fox, Masaki Fukunaga, J. Raphael Gibbs, Harald H. H. Gring, Robert C. Green, Sebastian Guelfi, Narelle K. Hansell, Catharina A. Hartman, Katrin Hegenscheid, Andreas Heinz, Dena G. Hernandez, Dirk J. Heslenfeld, Pieter J. Hoekstra, Florian Holsboer, Georg Homuth, Jouke-Jan Hottenga, Masashi Ikeda, Clifford R. Jack Jr, Mark Jenkinson, Robert Johnson, Ryota Kanai, Maria Keil, Jack W. Kent Jr, Peter Kochunov, John B. Kwok, Stephen M. Lawrie, Xinmin Liu, Dan L. Longo, Katie L. McMahon, Eva Meisenzahl, Ingrid Melle, Sebastian Mohnke, Grant W. Montgomery, Jeanette C. Mostert, Thomas W. Mhleisen, Michael A. Nalls, Thomas E. Nichols, Lars G. Nilsson, Markus M. Nthen, Kazutaka Ohi, Rene L. Olvera, Rocio Perez-Iglesias, G. Bruce Pike, Steven G. Potkin, Ivar Reinvang, Simone Reppermund, Marcella Rietschel, Nina Romanczuk-Seiferth, Glenn D. Rosen, Dan Rujescu, Knut Schnell, Peter R. Schofield, Colin Smith, Vidar M. Steen, Jessika E. Sussmann, Anbupalam Thalamuthu, Arthur W. Toga, Bryan J. Traynor, Juan Troncoso, Jessica A. Turner, Maria C. Valdernez, Dennis van ?t Ent, Marcel van der Brug, Nic J. A. van der Wee, Marie-Jose van Tol, Dick J. Veltman, Thomas H. Wassink, Eric Westman, Ronald H. Zielke, Alan B. Zonderman, David G. Ashbrook, Reinmar Hager, Lu Lu, Francis J. McMahon, Derek W. Morris, Robert W. Williams, Han G. Brunner, Randy L. Buckner, Jan K. Buitelaar, Wiepke Cahn, Vince D. Calhoun, Gianpiero L. Cavalleri, Benedicto Crespo-Facorro, Anders M. Dale, Gareth E. Davies, Norman Delanty, Chantal Depondt, Srdjan Djurovic, Wayne C. Drevets, Thomas Espeseth, Randy L. Gollub, Beng-Choon Ho, Wolfgang Hoffmann, Norbert Hosten, Ren Kahn, Stephanie Le Hellard, Andreas Meyer-Lindenberg, Bertram Mller-Myhsok, Matthias Nauck, Lars Nyberg, Massimo Pandolfo, Brenda W. J. H. Penninx, Joshua L. Roffman, Sanjay M. Sisodiya, Jordan W. Smoller, Hans van Bokhoven, Neeltje E. M. van Haren, Henry Vlzke, Henrik Walter, Michael W. Weiner, Wei Wen, Tonya White, Ingrid Agartz, Ole A. Andreassen, John Blangero, Dorret I. Boomsma, Rachel M. Brouwer, Dara M. Cannon, Mark R. Cookson, Eco J. C. de Geus, Ian J. Deary, Gary Donohoe, Guillernez, Simon E. Fisher, Clyde Francks, David C. Glahn, Hans J. Grabe, Oliver Gruber, John Hardy, Ryota Hashimoto, Hilleke E. Hulshoff Pol, Erik G. Jnsson, Iwona Kloszewska, Simon Lovestone, Venkata S. Mattay, Patrizia Mecocci, Colm McDonald, Andrew M. McIntosh, Roel A. Ophoff, Tomas Paus, Zdenka Pausova, Mina Ryten, Perminder S. Sachdev, Andrew J. Saykin, Andy Simmons, Andrew Singleton, Hilkka Soininen, Joanna M. Wardlaw, Michael E. Weale, Daniel R. Weinberger, Hieab H. H. Adams, Lenore J. Launer, Stephan Seiler, Reinhold Schmidt, Ganesh Chauhan, Claudia L. Satizabal, James T. Becker, Lisa Yanek, Sven J. van der Lee, Maritza Ebling, Bruce Fischl, W. T. Longstreth Jr, Douglas Greve, Helena Schmidt, Paul Nyquist, Louis N. Vinke, Cornelia M. van Duijn, Luting Xue, Bernard Mazoyer, Joshua C. Bis, Vilmundur Gudnason, Sudha Seshadri, M. Arfan Ikram, The Alzheimer?s Disease Neuroimaging Initiative, The CHARGE Consortium, EPIGEN, IMAGEN, SYS, Nicholas G. Martin, Margaret J. Wright, Gunter Schumann, Barbara Franke, Paul M. Thompson Sarah E. MedlandNature, vol. 520(7546), pp. 224-229. 2015. Journal Paper