

IMAGING GENOMICS

LI SHEN

*Center for Neuroimaging, Department of Radiology and Imaging Sciences
Center for Computational Biology and Bioinformatics
Indiana University School of Medicine
355 West 16th Street Suite 4100, Indianapolis, IN 46202
E-mail: shenli@iu.edu*

LEE A.D. COOPER

*Department of Biomedical Informatics, Emory University School of Medicine
Department of Biomedical Engineering, Georgia Institute of Technology
PAIS Building, 36 Eagle Row, 5th Floor South, Atlanta, GA 30322
E-mail: lee.cooper@emory.edu*

Imaging genomics is an emerging research field, where integrative analysis of imaging and omics data is performed to provide new insights into the phenotypic characteristics and genetic mechanisms of normal and/or disordered biological structures and functions, and to impact the development of new diagnostic, therapeutic and preventive approaches. The Imaging Genomics Session at PSB 2017 aims to encourage discussion on fundamental concepts, new methods and innovative applications in this young and rapidly evolving field.

1. Introduction

Imaging genomics¹⁻⁹ is an emerging research field that arises with the recent advances in acquiring high throughput omics data and multimodal imaging data. Its major task is to perform integrative analysis of genomics data and structural, functional and molecular imaging data. Bridging imaging and genomic factors and exploring their connections have the potential to provide important new insights into the phenotypic characteristics and genetic mechanisms of normal and/or disordered biological structures and functions, which in turn will impact the development of new diagnostic, therapeutic and preventive approaches.

Binformatics strategies for imaging genomics, which is a relatively young field,¹⁻⁴ have been rapidly evolving. Early studies started with the simplest strategy to examine pairwise univariate associations^{10,11} between genetic markers and imaging phenotypes. To identify more flexible associations involving multiple genetic markers and multiple imaging phenotypes, recent studies employed multiple regression and multivariate models,¹² sometimes coupled with powerful machine learning approaches¹³ and valuable prior knowledge¹⁴ to discover relevant imaging and genomic features. To increase statistical power and reduce false positives, meta-analysis studies^{15,16} were performed to quantitatively synthesize imaging genomic findings from multiple independent analyses. To hunt for “missing heritability”, epistatic studies¹⁷ were performed to examine genetic interaction effects on imaging phenotypes. To identify biologically meaningful findings with increased statistical power, imaging genetic enrichment analysis¹⁸ was proposed to mine set level associations in both imaging and genomic domains.

The topic of imaging genomics has recently been addressed in several medical imaging and bioinformatics conferences. The most focused one is the International Imaging Genetics

Conference (IIGC, <http://www.imaginggenetics.uci.edu/>), which is an annual meeting organized at the UC Irvine since 2005. The MICCAI Workshop on Imaging Genetics (MICGen, <http://micgen.csail.mit.edu/>) has been held twice in conjunction with the major medical image computing conference MICCAI in 2014 and 2015. An educational course on “Introduction to Imaging Genetics” has been offered at the annual meeting of the Organization for Human Brain Mapping (OHBM) since 2009. The topic of imaging genomics has also been covered in the following two events in the bioinformatics field: (1) ACM BCB 2015 Workshop on The Computational Pathology: Linking Tissue Phenotypes with Genomics and Clinical Outcomes, and (2) ICIBM 2015 Tutorial in Bioimage Informatics and Integrative Genomics.

As the field of imaging genomics contains a significant genomics (or omics in general) component in addition to biomedical imaging, we feel that it is timely for a major bioinformatics conference such as PSB to address this important, relevant and emerging topic. We believe that PSB offers an ideal and timely opportunity to bring together people with different expertise and shared interests in this rapidly evolving field. Specifically, the computational biology and bioinformatics expertise of the PSB and ISCB communities can provide important new perspective, complementary to the expertise of the IIGC, MICCAI, OHBM, ACM BCB and ICIBM communities, and thus can help contribute new concepts, methods, and applications to the analysis of emerging imaging and genomic data.

The scale and complexity of multidimensional imaging and omics data provide us unprecedented opportunities in enhancing mechanistic understanding of complex disorders such as neurological diseases¹⁹⁻²¹ and cancers,^{22,23} which can benefit public health outcomes by facilitating diagnostic and therapeutic progress. However, due to the extremely high dimensionality and complex structure of these data sets, this field is facing major computational and bioinformatics challenges. The technological advance in this field is urgently needed and has the potential to significantly contribute to multiple national health priority areas including *the Precision Medicine Initiative*,²⁴ *the Brain Initiative*,²⁵ and *the Big Data to Knowledge Initiative*.²⁶

The objective of this Imaging Genomics Session at PSB 2017 is to encourage discussion on fundamental concepts, novel methods and innovative applications. We hope that this session will become a forum for researchers to exchange ideas, data, and software, in order to speed up the development of innovative technologies for hypothesis testing and data-driven discovery in Imaging Genomics.

2. Session Summary

This session includes an invited lecture and five accepted presentations with peer-reviewed papers. Three presentations will be delivered as platform talks and the other two as posters.

2.1. Invited Talk

Our invited lecture will be given by Dr. Paul Thompson, a world renowned pioneer in imaging genomics. Dr. Thompson is from the University of Southern California (USC). At USC, he is a Professor of Neurology, Psychiatry, Radiology, Pediatrics, Engineering, and Ophthalmology, the director of the USC Imaging Genetics Center, and the director of the ENIGMA

Center for Worldwide Medicine, Imaging & Genomics – an \$11M NIH Center of Excellence in Big Data Computing. Dr. Thompson’s major contributions to the field of imaging genomics and to the science in general can be summarized by the following text quoted from <http://keck.usc.edu/faculty/paul-m-thompson/>:

Paul Thompson directs the ENIGMA Consortium, a global alliance of 307 scientists in 33 countries who conduct the largest studies of 10 major brain diseases – ranging from schizophrenia, depression, ADHD, bipolar illness and OCD, to HIV and addictions on the brain. ENIGMA’s genomic screens of over 31,000 people’s brain scans and genome-wide data (published in *Nature Genetics*, 2012; *Nature*, 2015) have brought together experts from 185 institutions to unearth genetic variants that affect brain structure, disease risk, and brain connectivity. Collaborating with imaging labs around the world, Dr. Thompson and his students have published over 1,300 publications (h-index: 116) describing novel mathematical and computational strategies for analyzing brain image databases, for detecting pathology in individual patients and groups, and for creating disease-specific atlases of the human brain.

2.2. Papers

In *Integrative analysis for lung adenocarcinoma predicts morphological features associated with genetic variations*, **Wang et al.** analyzed an imaging genomic data set downloaded from the TCGA portal, containing 201 patients with lung adenocarcinoma (LUAD). The data includes clinical information, mRNA expression profiles, and histopathologic whole slide images of the patients. On the imaging end, the authors calculated 283 morphological features from histopathologic images, and identified features strongly correlated with patient survival outcome. On the genomic end, the authors constructed the gene co-expression network and extracted gene co-expression clusters. To relate imaging with genomics, the authors regressed the outcome-relevant morphological feature on multiple co-expressed gene clusters using Lasso. The study identified gene clusters highly associated with DNA copy number variations. These observations may lead to new insight on lung cancer development, suggesting biological pathways from genetic variations, gene transcription, cancer morphology to survival outcome.

In *Identification of discriminative imaging proteomics associations in Alzheimer’s disease via a novel sparse canonical correlation model*, **Yan et al.** analyzed an imaging proteomic data set downloaded from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. Participants include 42 healthy controls, 67 patients with mild cognitive impairment (MCI), and 67 patients with Alzheimer’s disease (AD). The data includes clinical information, magnetic resonance imaging (MRI) scans, and expression data of 229 proteomic analytes (83 from cerebrospinal fluid and 146 from plasma). The authors developed a novel machine learning model, called discriminative sparse canonical correlation analysis (DSCCA), and applied it to the joint analysis of imaging, proteomic and diagnostic data. This analysis yielded a strong imaging proteomic association so that the identified imaging and proteomic components had also high discriminative power. Such an outcome-relevant imaging proteomic pattern has the potential to improve mechanistic understanding of the disease.

In *Enforcing co-expression in multimodal regression framework*, **Zille et al.** analyzed an imaging genomic data set collected by Mind Clinical Imaging Consortium (MCIC). Participants include 116 controls and 92 schizophrenia patients. The data includes clinical information, functional MRI (fMRI) scans, and genotyping data. The authors developed a new machine learning model, called MT-CoReg, by combining sparse regression with canonical correlation analysis; and applied it to the analysis of the MCIC data. The analysis identified imaging and genomic markers that not only induce a strong imaging genomic association but also can jointly predict the outcome.

In *Adaptive testing of SNP-brain functional connectivity association via a modular network analysis*, **Gao et al.** analyzed an imaging genomic data set downloaded from the ADNI database. Participants include 162 ADNI subjects: 73 with no *APOE* E4 allele, 67 with one copy of the *APOE* E4 allele, and 22 with two copies of the *APOE* E4 allele. The authors analyzed the resting-state fMRI data to identify modular structures in brain functional networks, using a weighted gene co-expression network analysis (WGCNA) framework, coupled with topological overlap matrix (TOM) elements in hierarchical clustering. After that, they employed an adaptive association test based on the proportional odds model to identify distinct modular structures in brain functional networks in relation to different *APOE* E4 groups.

In *Exploring brain transcriptomic patterns: a topological analysis using spatial expression networks*, **Kuncheva et al.** analyzed whole genome whole brain gene expression data downloaded from the Allen Human Brain Atlas (AHBA). Participants include six AHBA donors. The authors focused on 16,906 genes selected based on a previous study, and 105 brain regions where at least one measurement in all 6 brains were available. A Spatial Expression Network (SEN) was extracted for each gene to quantify co-expression patterns amongst several anatomical locations. After that, network similarity measures were computed and used to quantify the topological resemblance between pairs of SENs and identify naturally occurring clusters. The analysis identified three stable clusters, including one with genes specifically involved in the nervous system, and the other two representing immunity, transcription and translation.

2.3. Discussion

Most of these studies were facilitated by and conducted using the Big Data resources available in the open science domain, including TCGA analyzed in (**Wang et al.**), ADNI analyzed in (**Yan et al. & Gao et al.**), and AHBA analyzed in (**Kuncheva et al.**). The imaging data investigated by these studies ranged from histological whole slide images of cancer specimens in (**Wang et al.**), structural MRI scans in (**Yan et al.**), functional MRI scans in (**Zille et al. & Gao et al.**), to images of mRNA expression levels across the brain in (**Kuncheva et al.**). The omics data examined in these studies were also diverse, including DNA genotyping data in (**Zille et al. & Gao et al.**), mRNA expression profiles in (**Wang et al. & Kuncheva et al.**), and proteomic expression profiles in (**Yan et al.**).

These studies were performed to better understand the brain transcriptomic patterns in healthy controls (**Kuncheva et al.**), the brain imaging genomic or imaging proteomic patterns in Alzheimer's disease (**Yan et al. & Gao et al.**) or schizophrenia (**Zille et al.**), and biological pathways from gene transcription, tissue morphology to survival outcome in lung cancer

(Wang et al.). As to the bioinformatics strategies, a variety of machine learning methods were employed or newly developed in these studies, including network analysis and clustering models used in (Wang et al., Gao et al. & Kuncheva et al.), regression models used in (Wang et al.), an adaptive association test used in (Gao et al.), an integrative regression and canonical correlation analysis model used in (Zille et al.), and an outcome-regularized sparse canonical correlation analysis model used in (Yan et al.).

While Gao et al. studied functional brain network as an innovative imaging phenotype, Kuncheva et al. aimed to identify gene clusters using whole brain spatial expression networks. The remaining three studies (Wang et al., Yan et al. & Zille et al.) shared a common theme to examine the relationship among three levels (i.e., omics features, imaging phenotypes, and clinical outcomes). This suggests a promising future direction to integrate imaging genomics with systems biology, which attempts to model complex and interactive multilevel biological systems using multimodal imaging and multidimensional omics data sets.

3. Acknowledgements

We would like to thank all the authors for their high-quality submissions and excellent presentations. We would like to thank all the reviewers for taking their time and effort to evaluate the papers and provide valuable feedbacks. We would like to thank Dr. Paul Thompson of University of Southern California for giving an outstanding invited lecture. We would like to thank Dr. Kun Huang of Ohio State University for sharing his valuable experience on how to organize a successful PSB session. We would like to thank the PSB 2017 chairs and Tiffany Murray of Stanford University for their great help and support.

References

1. A. R. Hariri and D. R. Weinberger. Imaging genomics. *Br Med Bull*, 65:259–70, 2003.
2. V. S. Mattay and T. E. Goldberg. Imaging genetic influences in human brain function. *Curr Opin Neurobiol*, 14(2):239–47, 2004.
3. A. R. Hariri, E. M. Drabant, and D. R. Weinberger. Imaging genetics: perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biol Psychiatry*, 59(10):888–97, 2006.
4. D. C. Glahn, T. Paus, and P. M. Thompson. Imaging genomics: mapping the influence of genetics on brain structure and function. *Hum Brain Mapp*, 28(6):461–3, 2007.
5. P. M. Thompson, N. G. Martin, and M. J. Wright. Imaging genomics. *Curr Opin Neurol*, 23(4):368–73, 2010.
6. L. Shen, P. M. Thompson, S. G. Potkin, L. Bertram, L. A. Farrer, T. M. Foroud, R. C. Green, X. Hu, M. J. Huentelman, S. Kim, J. S. Kauwe, Q. Li, E. Liu, F. Macciardi, J. H. Moore, L. Munsie, K. Nho, V. K. Ramanan, S. L. Risacher, D. J. Stone, S. Swaminathan, A. W. Toga, M. W. Weiner, A. J. Saykin, and Alzheimer’s Disease Neuroimaging Initiative. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav*, 8(2):183–207, 2014.
7. M. G. ElBanan, A. M. Amer, P. O. Zinn, and R. R. Colen. Imaging genomics of Glioblastoma: state of the art bridge between genomics and neuroradiology. *Neuroimaging Clin N Am*, 25(1):141–53, 2015.
8. W. B. Pope. Genomics of brain tumor imaging. *Neuroimaging Clin N Am*, 25(1):105–19, 2015.

9. A. J. Saykin, L. Shen, X. Yao, S. Kim, K. Nho, S. L. Risacher, V. K. Ramanan, T. M. Foroud, K. M. Faber, N. Sarwar, L. M. Munsie, X. Hu, H. D. Soares, S. G. Potkin, P. M. Thompson, J. S. Kauwe, R. Kaddurah-Daouk, R. C. Green, A. W. Toga, M. W. Weiner, and Alzheimer's Disease Neuroimaging Initiative. Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement*, 11(7):792–814, 2015.
10. L. Shen, S. Kim, S. L. Risacher, K. Nho, S. Swaminathan, J. D. West, T. Foroud, N. Pankratz, J. H. Moore, C. D. Sloan, M. J. Huentelman, D. W. Craig, B. M. Dechaire, S. G. Potkin, Jr. Jack, C. R., M. W. Weiner, A. J. Saykin, and Alzheimer's Disease Neuroimaging Initiative. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage*, 53(3):1051–63, 2010.
11. J. L. Stein, X. Hua, S. Lee, A. J. Ho, A. D. Leow, A. W. Toga, A. J. Saykin, L. Shen, T. Foroud, N. Pankratz, M. J. Huentelman, D. W. Craig, J. D. Gerber, A. N. Allen, J. J. Corneveaux, B. M. Dechaire, S. G. Potkin, M. W. Weiner, P. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Voxelwise genome-wide association study (vGWAS). *Neuroimage*, 53(3):1160–74, 2010.
12. D. P. Hibar, J. L. Stein, O. Kohannim, N. Jahanshad, A. J. Saykin, L. Shen, S. Kim, N. Pankratz, T. Foroud, M. J. Huentelman, S. G. Potkin, Jr. Jack, C. R., M. W. Weiner, A. W. Toga, P. M. Thompson, and Alzheimer's Disease Neuroimaging Initiative. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. *Neuroimage*, 56(4):1875–91, 2011.
13. M. Vounou, E. Janousova, R. Wolz, J. L. Stein, P. M. Thompson, D. Rueckert, G. Montana, and Alzheimer's Disease Neuroimaging Initiative. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. *Neuroimage*, 60(1):700–16, 2012.
14. J. Yan, L. Du, S. Kim, S. L. Risacher, H. Huang, J. H. Moore, A. J. Saykin, L. Shen, and Alzheimer's Disease Neuroimaging Initiative. Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics*, 30(17):i564–71, 2014.
15. J. L. Stein, S. E. Medland, A. A. Vasquez, D. P. Hibar, R. E. Senstad, A. M. Winkler, R. Toro, K. Appel, R. Bartecek, O. Bergmann, M. Bernard, A. A. Brown, D. M. Cannon, M. M. Chakravarty, A. Christoforou, M. Domin, O. Grimm, M. Hollinshead, A. J. Holmes, G. Homuth, J. J. Hottenga, C. Langan, L. M. Lopez, N. K. Hansell, K. S. Hwang, S. Kim, G. Laje, P. H. Lee, X. Liu, E. Loth, A. Lourdasamy, M. Mattingsdal, S. Mohnke, S. M. Maniega, K. Nho, A. C. Nugent, C. O'Brien, M. Pampmeyer, B. Putz, A. Ramasamy, J. Rasmussen, M. Rijpkema, S. L. Risacher, J. C. Roddey, E. J. Rose, M. Ryten, L. Shen, E. Sprooten, E. Strengman, A. Teumer, D. Trabzuni, J. Turner, K. van Eijk, T. G. van Erp, M. J. van Tol, K. Wittfeld, C. Wolf, S. Woudstra, A. Aleman, S. Alhusaini, L. Almasy, E. B. Binder, D. G. Brohawn, R. M. Cantor, M. A. Carless, A. Corvin, M. Czisch, J. E. Curran, G. Davies, M. A. de Almeida, N. Delanty, C. Depondt, R. Duggirala, T. D. Dyer, S. Erk, J. Fagerness, P. T. Fox, N. B. Freimer, M. Gill, H. H. Goring, D. J. Hagler, D. Hoehn, F. Holsboer, M. Hoogman, N. Hosten, N. Jahanshad, M. P. Johnson, D. Kasperaviciute, Jr. Kent, J. W., P. Kochunov, J. L. Lancaster, S. M. Lawrie, D. C. Liewald, R. Mandl, M. Matarin, M. Mattheisen, E. Meisenzahl, I. Melle, E. K. Moses, T. W. Muhleisen, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*, 44(5):552–61, 2012.
16. D. P. Hibar, J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivieres, N. Jahanshad, R. Toro, K. Wittfeld, L. Abramovic, M. Andersson, B. S. Aribisala, N. J. Armstrong, M. Bernard, M. M. Bohlken, M. P. Boks, J. Bralten, A. A. Brown, M. M. Chakravarty, Q. Chen, C. R. Ching, G. Cuellar-Partida, A. den Braber, S. Giddaluru, A. L. Goldman, O. Grimm, T. Guadalupe, J. Hass, G. Woldehawariat, A. J. Holmes, M. Hoogman, D. Janowitz, T. Jia, S. Kim, M. Klein, B. Kraemer, P. H. Lee, L. M. Olde Loohuis, M. Luciano, C. Macare, K. A. Mather, M. Mattheisen, Y. Milaneschi, K. Nho, M. Pampmeyer, A. Ramasamy, S. L. Risacher, R. Roiz-Santianez, E. J.

- Rose, A. Salami, P. G. Samann, L. Schmaal, A. J. Schork, J. Shin, L. T. Strike, A. Teumer, M. M. van Donkelaar, K. R. van Eijk, R. K. Walters, L. T. Westlye, C. D. Whelan, A. M. Winkler, M. P. Zwiers, S. Alhusaini, L. Athanasiu, S. Ehrlich, M. M. Hakobjan, C. B. Hartberg, U. K. Haukvik, A. J. Heister, D. Hoehn, D. Kasperaviciute, D. C. Liewald, L. M. Lopez, R. R. Makkinje, M. Matarin, M. A. Naber, D. R. McKay, M. Needham, A. C. Nugent, B. Putz, N. A. Royle, L. Shen, E. Sprooten, D. Trabzuni, S. S. van der Marel, K. J. van Hulzen, E. Walton, C. Wolf, L. Almasy, D. Ames, S. Arepalli, A. A. Assareh, M. E. Bastin, H. Brodaty, K. B. Bulayeva, M. A. Carless, S. Cichon, A. Corvin, J. E. Curran, M. Czisch, et al. Common genetic variants influence human subcortical brain structures. *Nature*, 520(7546):224–9, 2015.
17. A. L. Zieselman, J. M. Fisher, T. Hu, P. C. Andrews, C. S. Greene, L. Shen, A. J. Saykin, and J. H. Moore. Computational genetics analysis of grey matter density in Alzheimer’s disease. *BioData Min*, 7:17, 2014.
 18. Xiaohui Yao, Jingwen Yan, Sungeun Kim, Kwangsik Nho, Shannon L. Risacher, Mark Inlow, Jason H. Moore, Andrew J. Saykin, and Li Shen. Two-dimensional enrichment analysis for mining high-level imaging genetic associations. *Brain Informatics*, pages 1–11, 2016.
 19. M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. Luthman, J. C. Morris, R. C. Petersen, A. J. Saykin, L. Shaw, L. Shen, A. Schwarz, A. W. Toga, J. Q. Trojanowski, and Alzheimer’s Disease Neuroimaging Initiative. 2014 update of the Alzheimer’s Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimers Dement*, 11(6):e1–120, 2015.
 20. R. J. Hodes and N. Buckholtz. Accelerating Medicines Partnership: Alzheimer’s Disease (AMP-AD) knowledge portal aids alzheimer’s drug discovery through open data sharing. *Expert Opin Ther Targets*, 20(4):389–91, 2016.
 21. Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). *Prog Neurobiol*, 95(4):629–35, 2011.
 22. J. McCain. The cancer genome atlas: new weapon in old war? *Biotechnol Healthc*, 3(2):46–51B, 2006.
 23. K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*, 26(6):1045–57, 2013.
 24. F. S. Collins and H. Varmus. A new initiative on precision medicine. *N Engl J Med*, 372(9):793–5, 2015.
 25. M. McCarthy. US to launch major brain research initiative. *BMJ*, 346:f2156, 2013.
 26. L. Ohno-Machado. NIH’s Big Data to Knowledge initiative and the advancement of biomedical informatics. *J Am Med Inform Assoc*, 21(2):193, 2014.