

PROCEEDINGS

Open Access



GAW20: methods and strategies for the new frontiers of epigenetics and pharmacogenomics

Nathan L. Tintle^{1*}, David W. Fardo², Mariza de Andrade³, Stella Aslibekyan⁴, Julia N. Bailey⁵, Justo Lorenzo Bermejo⁶, Rita M. Cantor⁷, Saurabh Ghosh⁸, Philip Melton⁹, Xuexia Wang¹⁰, Jean W. MacCluer¹¹ and Laura Almasy^{12,13}

From Genetic Analysis Workshop 20
San Diego, CA, USA. 4 - 8 March 2017

Abstract

GAW20 provided a platform for developing and evaluating statistical methods to analyze human lipid-related phenotypes, DNA methylation, and single-nucleotide markers in a study involving a pharmaceutical intervention. In this article, we present an overview of the data sets and the contributions analyzing these data. The data, donated by the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) investigators, included data from 188 families ($N = 1105$) which included genome-wide DNA methylation data before and after a 3-week treatment with fenofibrate, single-nucleotide polymorphisms, metabolic syndrome components before and after treatment, and a variety of covariates. The contributions from individual research groups were extensively discussed prior, during, and after the Workshop in groups based on discussion themes, before being submitted for publication.

Background

This supplement to *BMC Proceedings* contains the proceedings of the GAW20 that was held March 4–8, 2017, in San Diego, CA, USA. The GAWs were initiated in 1982 and have traditionally been held approximately biannually. They provide a discussion forum for developing and evaluating statistical methods aimed at deciphering the architecture of human complex diseases, mainly by identifying their genetic risk factors. Discussion and comparison of methods is facilitated by providing the same data sets to all researchers. These data sets are chosen by the GAW Advisory Committee, taking into consideration the suggestions and concerns of attendees. Discussion of future data sets begins the final day of the Workshop and remains open for at least a year. Data sets must be well characterized, address urgent needs for analysis tools in genetic epidemiology, and be available

upon request prior to the Workshop. After the GAW organizers release the data sets, researchers analyze the data and prepare a manuscript to submit to the Workshop. All coauthors of submitted manuscripts are eligible to attend the Workshop. Active participation in group discussions is required, as well as attendance at overall presentation and discussion meetings. Individuals who provided data or participate in the Workshop organization may also attend. More information about GAW, including upcoming Workshops, may be found at <http://www.gaworkshop.org>.

Genetic analysis workshop 20

GAW20 was the first GAW to explore the emerging field of epigenetic data, providing an opportunity to explore methodological questions of interest in epigenetics in the context of a family-based, longitudinal study that also included a pharmaceutical intervention. As with previous GAWs, analyses of these data by GAW20 participants largely focused on dealing with the high dimensionality of the single-nucleotide polymorphism marker

* Correspondence: nathan.tintle@dordt.edu

¹Department of Mathematics and Statistics, Dordt College, 498 4th Ave. NE, Sioux Center, IA 51250, USA

Full list of author information is available at the end of the article



data, accounting for the family structure and handling longitudinal data, with the new wrinkle of integrating DNA methylation data, all within the context of a clinical trial. These issues are natural considering the data set provided, which is described in detail in Aslibekyan et al. [1].

Although complete data set details are provided in Aslibekyan et al. [1], we provide a brief overview of the data set now as an introduction to this volume. Data from 188 families [$N = 1105$ individuals] participating in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study were the focus of analysis for GAW20. Data available on these 1105 individuals consisted of: (a) DNA methylation at 463,995 cytosine-phosphate-guanine (CpG) sites measured before and after a 3-week treatment with fenofibrate; (b) 906,000 single-nucleotide polymorphisms; (c) metabolic syndrome components ascertained before and after the drug intervention; and (d) relevant covariates. Methylation and genotype data were subject to a variety of standard filtering and quality control procedures. GAW20 participants had the option of focusing their methodological investigations on this “real” data or on an alternative version that was simulated (“simulated data”). Following a complex, but realistic, genetic model that hypothesized genetic modification of methylation on triglyceride levels at select loci, 200 replicates of simulated posttreatment methylation and triglyceride measurements were generated for each individual. Simulated data provided participants the opportunity to provide additional statistical validation and assessment of method performance.

The availability of the GAW20 data was announced by email in the fall of 2016 to roughly 3200 individuals on the GAW mailing list, resulting in 81 separate requests for data to participate in the Workshop. The number of GAW20 attendees in March 2017 was 80. Although individuals were allowed to present more analyses at the Workshop than had been described in their submitted papers, each group was still required to report the results of some analyses prior to the meeting to participate. Manuscripts were distributed among participants prior to the Workshop, and participants were assigned to discussion groups to facilitate discussion before and during the Workshop. Manuscripts from the other discussion groups were also available for download from the GAW20 online discussion forum or upon request prior to the Workshop. After the Workshop, 39 individual papers were accepted for publication and constitute this proceedings volume, with 12 papers accepted for publication in *BMC Genetics*.

Participants and contributions were from many countries, with the United States of America, Canada, and Germany providing the largest numbers of contributions. Additional contributing participants were from Australia, China, India, the United Kingdom, the

Netherlands, Norway, Poland, Spain and Taiwan. The contributions were subdivided into 7 discussion groups by topic, with 1 group split into 3 subgroups to facilitate more detailed and focused discussions. The themes were Causal Modeling (Group 1), Data Mining and Machine Learning (Group 2), Epigenetics–Complex Models (Group 3a), Epigenetics–Gene Searching (Group 3b), Epigenetics–Longitudinal Analysis (Group 3c), GWAS (Genome-wide Association Studies) (Group 4), Genotype-by-Methylation (Group 5), Repeated Measures (Group 6), and Genetics of Treatment Response (Group 7). The papers in this proceedings volume are presented according to these groupings, with Groups 2 and 3a merged because of the overlapping goals of the papers in these groups. However, group assignment was often not easy and topics in groups may overlap. The contributed papers are preceded by the data description by Aslibekyan et al. [1] and a description of the model used to generate the simulated data by Kraja et al. [2]. Each group was led by a moderator with previous GAW experience. The moderator encouraged and organized the discussion and presentations prior to, during, and after the Workshop. Discussions largely started before the Workshop and continued at the Workshop within group meetings. Each discussion group, directed by the group leader, was also in charge of preparing a presentation of the issues discussed in the group and the conclusions drawn. These presentations were made to all GAW attendees in plenary sessions. There were also 2 poster sessions for presenting individual contributions. The Workshop closed with plenary sessions on lessons learned and planning for future GAWs. After the Workshop, the group leader was typically in charge of editing group manuscripts, as well as writing the summary paper for the group. To avoid possible conflicts of interest, articles to which the group editor contributed were reassigned to other groups for the editing process. Summary papers and individual papers deemed to be of highest impact are published in a supplement to *BMC Genetics*, and all other individual contributions are found in these proceedings.

Overall, GAW20 uncovered many new challenges and unsolved problems with epigenetic and pharmacogenomics data, although many of these challenges mirror those identified in the analysis of GWAS and whole-genome sequence data. The discussions highlighted the need for methodological development in almost all considered areas.

Acknowledgements

Numerous individuals contribute to GAW by helping select Workshop topics, providing data sets, conducting simulations, distributing data to the participants, leading discussion groups, overseeing the writing of group summaries, reviewing manuscripts, and managing the Workshop as well as the publishing process afterwards.

We are grateful to the GOLDN study for allowing GAW20 participants to use the data set around which this Workshop was based. The GOLDN study is funded by National Institutes of Health (NIH) R01 HL091357 (Arnett), NIH R01 HL104135 (Arnett), and NIH K01 HL136700 (Aslibekyan). Publication charges are paid by NIH R01 GM031575. The GAW is supported by NIH grant R01 GM031575. The GAW20 discussion groups were led by Mariza de Andrade, Stella Aslibekyan, Julia Bailey, Justo Lorenzo Bermejo, Rita Cantor, Saurabh Ghosh, Philip Melton, Nathan Tintle, and Xuexia Wang. We are grateful to them for their work before, during, and after GAW20 in initiating, organizing, and overseeing pre-Workshop communication, group discussions, group presentations, and summary paper writing.

A total of 46 individuals assisted in peer review of the papers in this volume: Christopher Amos, Elizabeth Atkinson, Joan Bailey-Wilson, Sheila Barton, Elizabeth Blue, Anne-Laure Boulesteix, Shelley Bull, Gemma Cadby, Jenny Chang-Claude, Brandon Coombes, Heather Cordell, Robert Culverhouse, Adrienne Cupples, David Fardo, Christine Fischer, Nora Francheschini, Derek Gordon, Han Hao, Audrey Hendricks, Johannes Heise, Yijuan Hu, Anne Justice, Inke König, Johannes Martini, Kari North, Michael Nothnagel, Sara Pendergrass, Elizabeth Pugh, Steve Rich, Stephanie Santorico, André Scherag, Mary Sehl, Noha Sharafeldin, Kim Siegmund, Henner Simianer, Claire Simpson, Janet Sinsheimer, Eric Sobel, Hans Stassen, March Suchard, Jae-Hoon Sul, Maggie Haitian Wang, Ellen Wijmsman, Zheng Xu, Peng Zhang, Mark Kos, and Zhaogong Zhang. We are grateful to them for their constructive comments, criticisms, and feedback.

Beginning with GAW7 in 1991, Vanessa Olmo has been responsible for major aspects of Workshop organization. We are grateful to her for the many things she does that keep GAW running smoothly, which includes interacting with participants, organizers, editors, and publishers; coordinating data requests and data distribution; facilitating selection of Workshop sites and making local arrangements; maintaining the GAW web site and mailing list; and preparing many aspects of the GAW proceedings. Stella Aslibekyan, Michael Province, Devin Absher, and Donna Arnett participated in data set preparation. Aldi Kraja, Ping An, and Petra Lenzini worked on data simulation. Zenaida Mendoza created the graphics and layout of the pre-Workshop volume. Thomas Dyer and Mark Kos assisted with data distribution efforts. Hannah Lazarus assisted with pre-workshop organization and onsite meeting management. Sophie Colunga liaised with authors and managed the publication process. Malinda Mann typeset the articles for these proceedings.

The GAW Advisory Committee assists with planning for the GAWs, including selection of workshop sites and topics. At the time of GAW20, its members were: Laura Almasy (chair), Julia Bailey, Josee Dupuis, Corinne Engelman, David Fardo, Jeanine Houwing-Duistermaat, Inke Koenig, Jean MacCluer, Andrew Patterson, and Michael Province.

Since 1982, GAW has been funded by the National Institute of General Medical Sciences (NIGMS), through grant R01 GM31575 to Jean MacCluer and Laura Almasy. This grant also provided scholarship funds to assist graduate students and postdoctoral trainees attending GAW20. We would like to recognize Donna Krasnewich for her ongoing support and for her efforts as program director for the GAW grant at the time of GAW20. These proceedings, as well as the continued work of statistical genetic methods development through the collaborative format of the GAWs, would not be possible without her support or that of NIGMS.

We are particularly grateful to Jean MacCluer, without her there would be no GAW. As always, we wish to express our gratitude to the GAW participants, whose ongoing, enthusiastic support and vigorous scientific discussions are the very foundations of the Workshop.

Funding

Publication of this article was supported by NIH R01 GM031575.

Availability of data and materials

The data that support the findings of this study are available from the Genetic Analysis Workshop (GAW), but restrictions apply to the availability of these data, which were used under license for the current study. Qualified researchers may request these data directly from GAW.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institutes of Health.

About this supplement

This article has been published as part of BMC Proceedings Volume 12 Supplement 9, 2018: Genetic Analysis Workshop 20: envisioning the future of statistical genetics by exploring methods for epigenetic and pharmacogenomic data. The full contents of the supplement are available online at <https://bmcpoc.biomedcentral.com/articles/supplements/volume-12-supplement-9>.

Authors' contributions

NLT, DWF, MA, SA, JNB, JLB, RMC, SG, PEM, XW, JWM, and LA participated in Workshop organization and editing of the *GAW20 Proceedings*. NLT drafted the text of this manuscript with contributions from LA. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Mathematics and Statistics, Dordt College, 498 4th Ave. NE, Sioux Center, IA 51250, USA. ²Department of Biostatistics, University of Kentucky, 725 Rose St, Lexington, KY 40536, USA. ³Division of Biomedical Statistics, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, USA. ⁴Department of Epidemiology, University of Alabama at Birmingham, 1655 University Blvd., Birmingham, AL 35205, USA. ⁵Department of Epidemiology, Fielding School of Public Health, University of California, 650 Charles E. Young Dr. South, Los Angeles, CA 90095, USA. ⁶Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany. ⁷Department of Human Genetics, David Geffen School of Medicine at University of California, 650 Charles E Young Dr. South, Los Angeles, CA 90095, USA. ⁸Indian Statistical Institute, 203 B T Rd., Kolkata, West Bengal 700108, India. ⁹Curtin/UWA Centre for Genetic Origins of Health and Disease, School of Pharmacy and Biomedical Sciences, Curtin University and School of Biomedical Sciences, The University of Western Australia, 35 Stirling Hwy. (M409), Crawley, WA 6009, Australia. ¹⁰Department of Mathematics, University of North Texas, 1155 Union Circle #311430, Denton, TX 76201, USA. ¹¹Department of Genetics, Texas Biomedical Research Institute, 8715 W. Military Dr., San Antonio, TX 78227, USA. ¹²Department of Genetics, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Blvd., Philadelphia, PA 19104, USA. ¹³Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, 3401 Civic Center Blvd., Philadelphia, PA 19104, USA.

Published: 17 September 2018

References

- Aslibekyan S, Almasy L, Province MA, Absher DM, Arnett DK. Data for GAW20: genome-wide DNA sequence variation and epigenome-wide DNA methylation before and after fenofibrate treatment in a family study of metabolic phenotypes. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0114-0>.
- Kraja AT, An P, Lenzini P, Lin SJ, Williams C, Hicks JE, Daw EW, Province MA. Simulation of a medication and methylation effects on triglycerides in the Genetic Analysis Workshop 20. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0115-z>.