

Section of Statistical Genomics and Genetics

ASA SSGG Quarterly Newsletter: June 2021

<https://pixabay.com/illustrations/dna-genetic-material-helix-proteins-3539309/>

Get Connected!

Twitter: https://twitter.com/ASA_SSGG

LinkedIn:

<https://www.linkedin.com/groups/12501931/>

Homepage:

<https://community.amstat.org/sectiononstatisticsingenomicsandgenetics/home>

Upcoming Events

Please check our website for details.

<https://community.amstat.org/sectiononstatisticsingenomicsandgenetics/meetings/upcomingevents>

Dr. Dionne Swift will be presenting in the next webinar on Thursday, June 24 at 1:00pm ET. Dr. Swift works as a Principal Statistician in the data and modeling sciences department at Procter & Gamble, providing statistical leadership in collaborative research. Her current research focuses on high-dimensional data analysis, multi-omics integrative analysis, and Bayesian data analysis and causal inference with applications in genomics. More information about Dr. Swift can be found at <https://magazine.amstat.org/blog/2021/02/01/dionne-swift/>.

Announcements

- ASA Section on Statistics in Genomics and Genetics (SSGG) call for Invited Session Proposals for JSM 2022 (see page 7 for details).
- NEW! The Member Engagement Committee (MEC) is launching a new series on career development and mentoring. Join us in June for a discussion on Time Management,

Research Strategies, and Healthy Habits for graduate students. The session will feature tips and helpful strategies for a more efficient and productive Ph.D. experience. The exact date and time will be announced shortly.

- The MEC is developing a survey to understand members' professional needs and expectations from SSGG. Please look out for the survey in the next few weeks and take a few minutes to complete it. Your feedback is essential for us to serve you in the most optimal manner.

Contents

Get Connected!.....	1
Upcoming Events.....	1
Announcements.....	1
Awards and Recognitions.....	2
Member Interviews.....	2
Summary of Recent Webinars.....	5
Summary of Section Members.....	5
Sessions in JSM 2021.....	6
Call for Invited Session Proposals.....	7
Job Opportunities.....	7
Section Executive Committee.....	8
Contact Us.....	8

Awards and Recognitions

Two SSGG members have been named Fellows of the American Statistical Association. Congratulations Profs Archer and Vitek!



Kellie Archer, Professor and Chair at Division of Biostatistics, College of Public Health, The Ohio State University. Her citation reads, “For exemplary research associating high-throughput ‘omic data to clinical outcomes; for methodological

contributions for high-dimensional data analysis; and for outstanding educational contributions to the statistical analysis of genomic data.” Dr. Archer’s primary research area has been in the development of statistical methods and computational algorithms for analyzing high-dimensional datasets, which frequently arise in studies that use high-throughput genomic assays. Her current research is focused on developing methods and software when the response is either ordinal, discrete, or a time-to-event response where subjects may also be cured, and the genomic data are collected cross-sectionally or longitudinally.



Olga Vitek, Professor in the Khoury College of Computer Sciences at Northeastern University. Dr. Vitek is President of the Boston Chapter of the American Statistical Association. Her citation reads, “For outstanding contributions of

statistical methodology and open-source software to the bioinformatics and proteomics research communities; for furthering statistics education among experimental scientists; and for service to the profession”. Research in Dr. Vitek’s lab explores

synergies between statistical science and machine learning, as applied to quantitative large-scale mass spectrometry-based investigations, to understand the functioning of living organisms. This includes statistical experimental design, detecting analyte’s signals in large and complex outputs produced by mass spectrometers, and causal and counterfactual inference of regulatory events among the analytes. The lab develops methods, open-source software and educational materials, that are broadly used in academia and industry. The lab is the lead organizer of the educational event May Institute on Computation and Statistics for Mass Spectrometry and Proteomics. This work has been recognized with the 2019 Chan Zuckerberg Essential Open Source Software Award, and with the 2021 Gilbert S. Omenn Computational Proteomics Award of the US Human Proteome Organization.

Member Interviews

We interviewed our section’s Distinguished Student Paper Competition awardees. They will be presenting their work at JSM 2021.



Yujie Jiang

Department of Statistics at Rice University.

Paper Title:
CliP: fast subclonal architecture reconstruction

for cancer cells from genomic DNA sequencing data

Q1. Please describe something that interests you about your current research or work.

A1. *My current project is on developing methods and solutions for the accurate measurement of high-throughput genomic data. Specifically, it is about developing a fast and accurate subclonal architecture reconstruction method. What interests me the most is, this method I am developing has the potential to provide biological insight in tumor evolution and advance precision cancer treatment. Also, I enjoy the process of contributing to this field*

from a frequentist's perspective, where most current subclonal reconstruction methods are from a Bayesian approach based on the Dirichlet Process.

Q2. Tell us how you got where you are today.

A2. During my undergraduate and first year of doctorate study, my research was mainly focusing on tensor & matrix decomposition. Realizing how my skillsets could be of help in the bioinformatics area, and since I had always enjoyed working on large-scale data analysis, I decided to join Dr. Wenyi Wang's lab last year. CliP is the first paper I worked on in the lab and is built on the effort of all co-authors. I hope to thank my advisor Dr. Wenyi Wang as well as the whole lab for supporting me from all aspects - it is because of their talent and resources that I am able to tackle the challenges throughout this process. Also, this project is inspired by the Cell paper "Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes", which is a joint work of hundreds of researchers under the entire PCAWG consortium, and I am very lucky that I have this opportunity to "stand on the shoulders of giants". I am still on the way and I look forward to making more progress in the future.

Q3. What do you enjoy doing outside of work?

A3. I enjoy doing sports (basketball, jogging, weightlifting, etc.) and playing video games. It is always a pleasure to hang out with friends, either hiking, watching a game, or simply having dinner together

Q4. What is your advice for others following in your footsteps? (junior faculty or students?)

A4. I think everyone is special and talented in their own different ways so I am not sure if my advice is applicable for everyone. One piece of advice I would like to share is, it is important to stay curious. Read more papers, keep an eye on cutting-edge techniques, and not be confined to the specific areas we are working on. To me, every time I suffer from setbacks, my curiosity could drive me to think from other perspectives, which always keeps me away from boredom and frustration.

.....

Jian Hu

Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania.



Paper title: "Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network"

Q1. Please describe something that interests you about your current research or work.

A1. I am very passionate about statistical genomics, which is a rapidly innovating field and with the surge of state-of-the-art techniques. This field has a strong demand for statistical tools to analyze high-volume data. Currently, one of the ongoing projects is about developing machine learning tools for spatial transcriptomics (ST) data to detect spatial domains and spatial variable genes.

Q2. Tell us how you got where you are today.

A2. I was a typical student in molecular biology and double majored in statistics using hypothesis testing and other traditional methods to analyze data. However, I realized that I dealt with high-dimensional data all the time and the model performance is not always superior. I started to train myself with machine learning techniques by taking courses from the department of computer science at Penn in the first few years of my PhD study. With transferring the knowledge between biostatistics and machine learning, I found I can innovatively develop a well-performed algorithm to predict cell types in single-cell RNA-seq analysis, and my biology background allows me to interpret the results.

Q3. What do you enjoy doing outside of work?

A3. I love many sports, including badminton, soccer and billiards. During the Covid time, I started to enjoy spending hours and hours in the kitchen and cooking.

Q4. What is your advice for others following in your footsteps? (junior faculty or students?)

A4. Statistical genomics is an interdisciplinary field where it asks everybody to jump out of the box that

you usually lived in. Learning and reading "out of the box" research methods and publications would inspire you somehow.

.....



Zhenxing Guo

Department of Biostatistics,
Emory University.

Paper title: Detecting m6A methylation regions from Methylated RNA Immunoprecipitation Sequencing

Q1. Please describe something that interests you about your current research or work.

A1. My current work mostly focus on developing statistical models for RNA methylation sequencing data analysis. Studies already report that changes in the degree of RNA methylations tend to enhance the translation of oncogenes, or inhibiting cancer suppressor genes. Therefore, studying the dynamics of RNA modifications can shed light on the status of disease and cancer, and thus provide useful biomarkers for disease diagnosis. These connections between RNA modifications and disease development make me excited about my current research.

Q2. Tell us how you got where you are today.

A2. Firstly, I have been being surrounded by a lot of motivated people. They are so clear about their career goals and working so hard. They build role models for me to follow. Secondly, I'm so lucky to have advisors that provide me with detailed and directed guidance and encouragement during my research. They are so supportive and responsive so that I can switch to them for help at any time. Thirdly, as a student, I work hard and think hard. Lastly but importantly, I'm being strongly and generously supported by my family. Without their support, It almost impossible to have what I got now.

Q3. What do you enjoy doing outside of work?

A3. During my spare time, I like hiking, cooking and

watching movies.

Q4. What is your advice for others following in your footsteps? (junior faculty or students?)

A4. Here is my advice according to my own experiences. (1) Read papers. It is a great way to learn and think. (2) Stay passionate about your work and work hard. (3) Stop being afraid to ask for help when bottlenecks occur.

.....

Amanda Brucker

Department of Statistics, North Carolina State University

Paper title: Association Test Using Copy Number Profile Curves (CONCUR) Enhances Power in Rare Copy Number Variant Analysis

Q1. Please describe something that interests you about your current research or work.

A1. I'm working on a project phenotyping adverse events using electronic health records and claims data. This involves lots of digging in the dirt of EHR data and stumbling across useful insights that help us decide on the best analytical approach. I also support multiple pulmonologists in their research related to COVID-19, which is very fulfilling and fascinating.

Q2. Tell us how you got where you are today.

A2. I attended a very small liberal arts college with no statistics department for my undergraduate degree but was able to participate in a Research Experience for Undergrads (REU) in applied math at the University of Nebraska Lincoln. This helped me get into a Summer Institute for Training in Biostatistics (SIBS) program at the University of Pittsburgh and a 1-year fellowship at the National Cancer Institute. I entered the PhD program in statistics at NC State University with support from a NHLBI training grant that allowed me to shadow biostatisticians at the Duke Clinical Research Institute (DCRI). I ended up working at the Duke Clinical Research Institute for 4 years while I was pursuing my degree, and now I work as a biostatistician in the Duke Department of Biostatistics and Bioinformatics! This path would not have been possible without the support of some key

advocates and mentors in my life and the support of publicly funded research institutions like the NIH and NSF.

Q3. What do you enjoy doing outside of work?

A3. I love hiking around the state parks of the Triangle area and the Tobacco Trail in Durham. I also love gardening with flowers and vegetable beds to attract birds and pollinators. I'm a plant nerd and always enjoy visiting the NC Botanical Garden, Duke Gardens, and JC Raulston Arboretum. Besides that, I enjoy playing old PC games and I listen to a lot of true crime, history, and current events podcasts.

Q4. What is your advice for others following in your footsteps? (junior faculty or students?)

A4. My advice is more material than philosophical: make use of the resources available to you when you're a student. 3 resources that had a significant impact on my graduate school experience were professor office hours (early PhD years), the gym and student health center (middle PhD years), and a dissertation completion grant (the last mile of my PhD marathon).

understanding of how genetic factors influence susceptibility to disease in admixed populations.

In March 2021, **Dr. Hongzhe Li**, presented several statistical methods in microbiome data science and explained their strong empirical performance with theoretical insights. In particular, he talked about how the microbiome affects metabolism, and the role of statistics in microbiome data science. He also introduced the role of shotgun metagenomics for understanding growth dynamics. To further explain the role of statistics in microbiome data science, he gave three examples: 1) data on the phylogenetic tree, 2) bacterial growth rate estimation and 3) Deep Learning for Biosynthetic Gene Cluster (BGC) prediction.

In April 2021, **Dr. Jin Zhou** talked about the within subject variability and diabetes complications. As a motivation for the talk, she investigated whether genetic variants contribute to glucose level and blood pressure variation, which may contribute to diabetes progression and complications. Her group developed a new method called, Method of moment, which is significantly faster than the existing method of MLE, to estimate a within-subject variance.

Summary of Recent Webinars

We summarize three recent webinars hosted by SSGG below. All our previous webinars were recorded and are available for section members at the Events tab in our section website (log in required).

In February 2021, **Dr. Genevieve Wojcik** spoke about bias in genetic epidemiology studies for admixed populations. Most previous genetic studies are overwhelmingly focused on populations of European descent. However, she addressed that a disproportionate level of chronic disease burden is found in minority populations. She described the bias in terms of four challenges in diverse multi-ethnic studies: 1) representation, 2) trait mapping, 3) generalizability and 4) translation. In the context of these challenges, it is important to gain a better

Summary of Section Members

As an extension of the section member analysis in the March 2021 newsletter, we further analyzed the employment sector of our section members.

Employment Sector

As of February 2021, SSGG has a relatively higher number of members from the Academic sector (not student) with 48%, compared to 35.6% in the overall ASA membership. Consequently, SSGG has a relatively smaller number of members from Business and Industry (14%) compared to the overall ASA membership (22.1%). For SSGG, the percentage of Academic (not student) members had been decreasing from 2010 until 2017, but has been increasing since 2017. The percentage of Business and Industry members has been around 10 – 15% in SGG, which is much lower than the

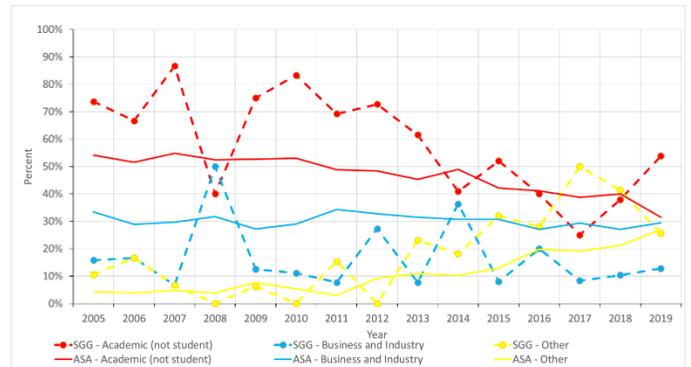
overall ASA membership (~30%) in the last decade. However, the percentage of members in the employment sector defined as “Other” has been increasing since 2015 in both SSGG and overall ASA membership (**Figure 1**)

Summary and Conclusions

From our current and previous analyses, we can draw a few conclusions regarding the membership in SSGG and ASA overall:

1. Both SSGG and ASA overall have a relatively high percentage of members who identify as White and Asian. We believe more diversity is necessary, and it is important to work with ASA to develop strategies to increase membership in SSGG and ASA for statisticians who identify as Black, Hispanic, Indigenous and other under-represented groups.
2. Regarding gender distribution, the section has a better balance (male 48.0%, female 42.0%) compared to ASA overall (male 53.4%, female 30.5%). However, more than 10% of members selected “Decline to Answer”, and we recommend ASA including an option for non-binary responses.
3. The percentage of members from the Business and Industry sectors in SSGG is significantly lower than the overall ASA membership. More people from Business and Industry should be encouraged to join SSGG.
4. The percentage of the employment sector in the “Others” category has been increasing in both SSGG and ASA overall. However, without more information, it is difficult to understand what professions are represented in this group. Thus, a more detailed categorization of the “Others” group should be provided in the employment sector category.

Figure 1. Percentage of SGG and ASA membership for each employment sector by year.



The percentage of members in the employment sectors labeled as Federal/National Government, Private Consultant/Self Employed, and State, Provincial, Local Government are not shown in the figure since they are relatively small.

Sessions in JSM 2021

Our section is sponsoring several exciting invited and topic contributed sessions at this year’s JSM. Please join us for these sessions!

Invited Sessions:

- Novel Methods for Microbiome Data Analysis, Sun, 8/8/2021, 1:30 PM - 3:20 PM
- Statistical Advances for Multi-Omics Data of Complex Diseases, Mon, 8/9/2021, 10:00 AM - 11:50 AM
- Recent Development in Computational Biology and Bioinformatics, Tue, 8/10/2021, 10:00 AM - 11:50 AM

Topic Contributed Sessions:

- Statistical Methods for Multi-Omic Data Analysis, Mon, 8/9/2021, 1:30 PM - 3:20 PM
- Statistical Methods for Studying Spatial Transcriptomics, Tissue Heterogeneity and Pleiotropy, Wed, 8/11/2021, 3:30 PM - 5:20 PM
- Novel Statistical Methods for Microbiome Data Analysis, Thu, 8/12/2021, 4:00 PM - 5:50 PM

In addition, we also sponsor many **Contributed Speed Sessions**, which span topics such as

statistical methods for data from single cell technologies, epigenetic/epigenomic data analysis, microbiome data analysis, and advancements in the analysis of genome-wide association studies. Please look at the JSM 2021 official website for further information.

- Sun, 8/8/2021, 3:30 PM - 5:20 PM
- Tue, 8/10/2021, 1:30 PM - 3:20 PM
- Wed, 8/11/2021, 10:00 AM - 11:50 AM
- Thu, 8/12/2021, 2:00 PM - 3:50 PM

Call for Invited Session Proposals

Although the JSM 2021 will be starting soon, this is an important time to start planning for next JSM in 2022, which will take place in Washington, District of Columbia, August 6-11, 2022.

We support your submission!

The Section of Statistics in Genomics and Genetics (SSGG) is a community of individuals interested in Statistical Genomics and Genetics within the American Statistical Association. The goals of SSGG is to foster research, education and influence of statisticians on genomics and genetics and associated applications. We are now calling for invited session proposals. Invited paper sessions consist of 3–6 speakers and discussants reporting new discoveries or advances in a topic on statistical genomics and genetics; invited panels include 3–6 panelists providing commentary, discussion, and engaging debate on a particular topic of interest for SSGG; and invited posters consist of posters presented during the Opening Mixer. We are soliciting proposals broadly related to statistical genomics and genetics, computational biology and relevant topics.

To submit a session proposal to be considered for 2022 JSM, please read the instructions and guidelines at <https://ww2.amstat.org/meetings/jsm/2022/>.

Be sure to check SSGG as your first choice of sponsor. Submissions are open now, and

close **September 8, 2021**. Decisions will be made early October 2021.

We hope to see you in Washington, DC in 2022!

Job Opportunities

- [Posted](#) 05.07.2021. Director of Biostatistics, Illumina. Details [here](#).
- [Posted](#) 05.07.2021. Senior Staff Biostatistician, Illumina. Details [here](#).
- [Posted](#) 05.06.2021. Postdoctoral Associate, Department of Biostatistics, Boston University School of Public Health. Statistical approaches for analysis of human genetics and genomics data with a focus on complex disease. Contact Dr. Gina Peloso (gpeloso@bu.edu).
- [Posted](#) 05.04.2021. Tenure-Track/Tenured Faculty Position, Division of Biostatistics, College of Public Health, The Ohio State University. Details and application link [here](#).
- [Posted](#) 04.29.2021. Staff Fellow Positions, FDA. Artificial intelligence and machine learning for medical imaging, precision medicine, and more. [LinkedIn job ad](#)
- [Posted](#) 03.26.2021. Statistical Genetic Analyst, Vanderbilt University. Contact Dr. Ran Tao (rtao@vumc.org).
- [Posted](#) 03.23.2021. Postdoctoral Research Fellow, Fred Hutchinson Cancer Research Center. Statistical Methods for Genomic/Microbiome Data. Contact Dr. Michael Wu (mcwu@fredhutch.org).
- [Posted](#) 03.18.2021. Postdoctoral Research Fellow, Johns Hopkins University. Genetic determinants of type 2 diabetes and its biomarkers. Contact Dr. Debashree Ray (dray@jhu.edu).
- [Posted](#) 03.08.2021. Postdoctoral Associate, Boston University School of Public Health.

Statistical genetics and genomics. Contact Dr. Ching-Ti Liu (ctliu@bu.edu).

- [Posted](#) 03.04.2021. Postdoctoral Fellow, Johns Hopkins University. Novel statistical methodologies for high dimensional omics studies, focusing on metagenomics and single cell sequencing studies. Contact Dr. Ni Zhao (nzhao10@jhu.edu).

Section Executive Committee

ASA SSGG Executive Committee		
Position	Name	Contact
Past-Chair	Saonli Basu	saonli@umn.edu
Chair	Katerina Kechris	katerina.kechris@cuanschutz.edu
Chair-Elect	Michael Wu	mcwu@fhcrc.org
Section Representative	Iuliana Ionita-Laza	li2135@columbia.edu
Secretary	Yingying Wei	ywei@cuhk.edu.hk
Treasurer	Li-Xuan Qin	qinl@mskcc.org
Program Chair	Sündüz Keleş	keles@biostat.wisc.edu
Program Chair-Elect	Lin Chen	lchen@health.bsd.uchicago.edu
Communications Officer	Ching-Ti Liu	ctliu@bu.edu
Member Engagement Committee Chair	Shili Lin	shili@stat.osu.edu

Contact Us

The newsletter was edited by Audrey Hendricks (audrey.hendricks@ucdenver.edu), Jongyun Jung (jongyun.jung@unlv.edu), Ching-Ti Liu (ctliu@bu.edu), Eric Lock (elock@umn.edu), and Anna Plantinga (amp9@williams.edu).

If you have something to share with our section or would like to contribute to our newsletters, please contact Ching-Ti Liu (ctliu@bu.edu).