INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY

Newsletter – *January 2019 Post-Election Edition*

Contents

Letter from the President
Celia Greenwood

To all IGES members

It is an honour to be your president for 2019! I would like to extend an enthusiastic welcome to our President-Elect Peter Kraft, and to warmly thank Inke König for her wonderful job as President last year. I would also like to welcome our new Board Members Jaya Satopagan and Eleftheria Zeggini, and to say a sincere thank you to Julia Bailey and Xihong Lin for their three years on the Board of Directors.

We are a fairly small international society, so it may be worth taking some time to reflect on the benefits of membership in IGES, and also to think about possible areas for development. At the 2018 Annual Meeting in San Diego, we asked past IGES presidents for their thoughts on future directions for the society. Several attendees said that they liked the small size of the Society. It groups us together through our common interests in a domain spanning not only epidemiology and genetics, but also many other areas such as statistics, genomics, bioinformatics, machine learning, prediction modelling, etc. As our Past President, Inke, said in her 2018 Presidential address, we are united by the idea that design and reproducibility principles remain important even in the big data era. Also, the size of the society

facilitates opportunities for junior members to interact with more senior members through formal mentoring or informal discussions.

Nevertheless, it may be worth thinking about how to bring in some new members to IGES from around the world – to increase our international reach and diversity, and to thereby enrich the Society. I hope you will reach out to me or to your Board of Directors with any suggestions you may have. I also look forward to discussing your ideas with you at our upcoming 2019 meeting in Houston, Texas (October 12-14, at the Hilton of the Americas), to be held back to back with the American Society of Human Genetics. Our proposed 2020 meeting location in Hong Kong (July 1-3, 2020) is one way that we are reaching out.

IGES is always looking for people interested in serving on our committees, so please consider volunteering by contacting iges@geneticepi.org. Also, if you have not already done so, please renew your IGES membership for 2019.

I wish you all the best for the upcoming 2019 year.

Celia Greenwood

2019 IGES President

Treasurer's Corner
Mariza de Andrade

I am pleased as Treasurer to share with IGES members that we are now in good financial standing, and that we will continue to be vigilant in our negotiations and management of resources to maintain this good standing.

I would like to remind IGES members that the 2019 IGES membership is open; an email reminder with details was sent to the members. You can also go to www.geneticepi.org for information. Please encourage your colleagues to join as well!

As a reminder for this year, IGES is a charitable and educational organization, and any contribution other than membership and meeting registration fee is tax-deductable. This can be done along with your membership renewal process. Your continuous support of IGES is appreciated!

We are looking forward to seeing you all in IGES 2019 in Houston, TX, USA. The IGES meeting will be immediately before the ASHG that will also take place in Houston.

Mariza de Andrade

Treasurer

Editor's Corner, Genetic Epidemiology
Sanjay Shete

Dear IGES members, as an official journal for our society, **with a new improved impact factor of 2.544**, Genetic Epidemiology invites you to submit your work in the fields of statistical, epidemiological and population genetics. Genetic Epidemiology is interested in both the methodological and applied papers. Examples include: applied genetic epidemiology papers (e.g. meta-analyses of GWAS, Secondary analyses of GWAS data), gene and environment interactions, risk prediction models, DNA methylation and RNA seq data analysis. Other novel work is welcome!

The December 2018 issue can be found here:

https://onlinelibrary.wiley.com/toc/10982272/2018/42/8

Benefits of Publishing in Genetic Epidemiology: There is no publishing cost for authors (e.g. page charges, black-white figures). In addition, every year, journal selects few papers for "open access" (unrestricted online access) publishing at no cost to the authors. Please register on Wiley online library to receive email alerts for new content and saved searches. The website for registration is http://onlinelibrary.wiley.com/user-registration

AUTHOR SERVICES: Wiley has an updated list of services it offers our authors. For more information, please visit http://authorservices.wiley.com

We now offer Kudos (https://www.growkudos.com/), a social media service that provides authors with a free set of tools to explain and share their published work for greater usage and impact. Authors also receive access to a publication dashboard where they can view downloads, citations, and altmetrics for their articles.

This is your journal: make it reflect your work by submitting your papers to Genetic Epidemiology!

Thanks and I look forward to your active participation in the journal.

Sanjay Shete
Editor-in-Chief
sshete@mdanderson.org

IGES 2019 Membership Drive

Dear IGES members,

2019 IGES membership registration/renewal is now open! The early-bird membership fee is available before March 31, 2019.

Please observe the fact that membership registration and annual meeting registration have been separate since 2014. Therefore you are NOT automatically registered as an IGES member just with meeting attendance. A separate registration for IGES Society membership is necessary:

http://www.geneticepi.org/membership/

Rates for 2019 IGES membership:

- Regular member **US$130**, if paid by March 31, 2019 (early-bird)
- Regular member **US$155** if paid on or after April 1, 2019
- Student and Post-Doctoral Fellow* member with online access to Journal **US$68**, rate continues all year but please aim to pay by March 31, 2019
- Student and Post-Doctoral Fellow* member **without** online access to Journal **US$25**, rate continues all year but please aim to pay by March 31, 2019
- Low- or lower-middle- income country* member with online access to Journal **US$68**; without online access to Journal **US$25.**

* For eligibility see IGES website

Benefits of IGES membership:

- Receive complimentary on-line access to **Genetic Epidemiology** (the offical IGES journal)
- Post announcements and job adverts on the IGES website at ***no cost***
- Attend annual IGES scientific meetings at ***reduced cost***
- Interact with **scientific peers** worldwide
- Learn about the **latest methodological developments**, **software**, and **research findings** in genetic epidemiology
- Help the society **promote** the field of **genetic epidemiology** internationally

- Continuous regular membership of IGES for 5 years or more makes you eligible for the title of "**Fellow Member of IGES**". See http://www.geneticepi.org/membership/iges-fellow-members/

Your IGES dues help to support…

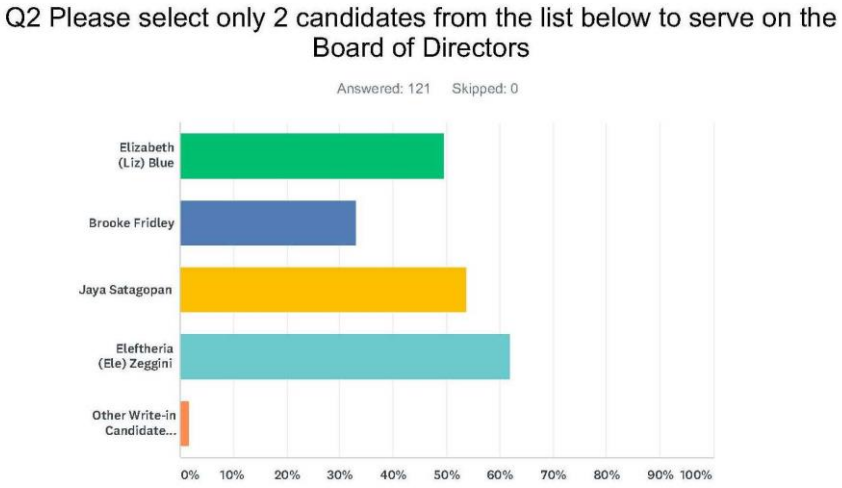
- **Travel expenses** to the annual IGES meeting for worthy students in financial need
- **Continuing education** and outreach to the scientific community about the discipline, analytical methods, and software used in genetic epidemiology through workshops and classes
- Subsidies for **student subscriptions** to the journal *Genetic Epidemiology*

IGES 2018 Elections for 2020 President and 2019 Members of the Board of Directors Election Results

Thank you to all the members who cast their votes in the recent ballot. Here are the results of the 2018 Elections.

We congratulate **Peter Kraft** on his election as IGES President in 2020. Peter will take on the role of President-Elect in 2019.

After an extremely close election, our congratulations also go to **Jaya Satagopan** and **Eleftheria Zeggini** on their election as members of the IGES Board of Directors to serve from January 1st 2019 to December 31st 2021.



The electoral turnout, as a percentage of all current members in 2018, was 32.7%.

Congratulations to the successful candidates and thank you to everyone for your candidacy and support of IGES.

[The ballot paper for these elections was sent out by SurveyMonkey. Note that all members may opt-out of emails from IGES sent via sources such as MemberClicks, Cvent and SurveyMonkey. However, please note that by opting out, you will no longer receive ANY mail at all from IGES, including important election announcements and ballot links, as well as the newsletters and our annual meeting announcements].

Report from ASHG/IGES/ISCB Joint Symposium: Working with Big Data in the Cloud—Research and Privacy
October 16, 2018

Summarized by Jessica Dennis, on behalf of the IGES Ethical Legal and Social Issues (ELSI) Committee

Symposium preamble:

Numerous large scale human genome research projects (e.g. UK Biobank, International Human Epigenome Consortium, 100,000 Genomes, AstraZeneca 2M Genomes, NIH All of Us) have been launched, producing considerable amounts of personal genome data for analysis. These make use of omics technologies including whole-genome and whole-exome sequencing, transcriptome sequencing, and epigenome sequencing.

The scale of these datasets requires use of cloud-computing resources to either store or analyze (or both) the vast amounts of data. Importantly, the shift to cloud-computing moves the personal data and analysis to an environment outside of personal/research lab control, thus necessitating an understanding of the legal and ethical obligations in maintaining data privacy, and experience in architecting and operating in cloud-compute environments to ensure human genome data is secure yet accessible by those with permissions.

This session presents an opportunity for the three communities (ASHG/IGES/ISCB) to come together to learn and discuss the legal and ethical frameworks governing personal genome data, particularly in the cloud-computing environment, and to share best practices for operating securely in this environment while still facilitating research. Practical applications to current big genome data projects will be presented and the session will close with a dynamic panel discussion with all speakers.

Moderators: Jaya Satagopan, IGES; and Jill Mesirov, ISCB

Schedule

1:00-1:05 pm	Introduction/Overview.
1:05-1:35 pm	Implications of the fourth (aka data driven) paradigm. Philip E. Bourne, Univ Virginia, Charlottesville.
1:35-2:05 pm	"Free trade" in the cloud? Bartha M. Knoppers, McGill Univ, Montreal, Canada.
2:05-2:30 pm	Infrastructure for analyzing and disseminating large-scale genetic data for type 2 diabetes and other complex diseases. Jason Flannick, Boston Children's Hosp, Broad Inst, Boston.
2:35-2:50 pm	Break
2:50-3:20 pm	Research using big data (PanCancer analysis of whole genomes). Christina Yung, Ontario Inst Cancer Res, Toronto, Canada.
3:20-4:00 pm	Panel discussion

Philip E. Bourne: Implications of the fourth (aka data driven) paradigm
Slides on SlideShare: https://www.slideshare.net/pebourne/implications-of-the-fourth-paradigm

Dr. Bourne introduced us to the fourth paradigm of research, which lies at the heart of all data science, not just data science in biomedicine. The four paradigms are: (1) A thousand years ago, we were focused on descriptions of natural phenomena; (2) In the last few hundred years, we advanced theoretical science (e.g., Newton's laws); (3) In the last few decades, computational science emerged, including simulations of complex phenomena; (4) Today: Data intensive science, with data generated from many sources.

Bourne next discussed the value of data science to current research. Digital data in the world are doubling every 2 years. How do we translate these data into meaningful improvements to the human condition? Data drive the virtuous data science cycle: methodologies transform discovery, which lead to new methodologies. Bourne argued that the scientific community therefore needs to re-think how we value contributions to this cycle. For example, the University of Virginia considers alternative metrics for promotion and tenure (e.g., contribution to datasets that are useable and highly cited, and less emphasis on published papers). In accordance with the importance of data science to current research, Bourne showed us that skills in data science are in demand now: data science offerings at research universities have skyrocketed, and many (>50%) are now offering MS degrees in data science.

What is the future of data science? Bourne drew an analogy with photography. When something becomes digitized, you move through this exponential "D" curve (Digitization, Deception, Disruption, Demonetization, Dematerialization, Democratization"). Photography has transitioned through this curve; could this also happen to biomedical research? Although the Advisory Board to the NIH Director was not terribly convinced, Bourne thought otherwise. The potential for "Disruption" could be profound. He predicted drivers of change would be:

1. Advances in machine learning. E.g., MIT investing $1 billion in an artificial intelligence institute. Expect lots of other announcements forthcoming about ramping up these initiatives at other institutions. Artificial intelligence methods rely on good quality data for training.
2. Advances in computing: E.g., GPUs originally designed for gaming, now used in movie industry. Architecture of chips is well suited to machine learning. Next generation of chips will be even more impressive (comments from a leading AI engineer at Google)
3. The private sector. Many developments are occurring in industry; academia has fallen behind, not very common in the past. Academic computer centers should be nervous! Cloud computing: E.g., Amazon: 20-30% of their business in these services. Profit margin 3x higher than internet business.

Bourne emphasized that embodied in all of data science is the notion of fairness and open access to content - every data science initiative has an ethical, legal, and social component. None of us know how to adequately address these issues. How do we balance between accessibility, privacy, etc.? Bourne concluded that technical challenges are the easy part - law and policy are our current biggest challenges.

Bartha Knoppers: "Free trade" in the cloud?
Slides accessible on ASHG meeting site

Dr. Knoppers talked about the ethical, legal, and social issues inherent to big data. In the first section of her talk, she presented examples of "International governance: a decade of data sharing (e.g., International Cancer Genome Consortium (ICGC), International Human Epigenome Consortium (IHEC), and The Global Alliance for Genomics and Health (GA4GH))". She showed us the remarkable growth and productivity of ICGC datasets and how the ICGC is now involving industry (e.g., ARGO: Accelerating Research in Genome Oncology), but this

poses legal challenges. Nonetheless, access to ARGO is pretty quick and easy with an average wait of 5 working days. Of the 350 applications approved so far, 91% were from academia. IHEC (a member of GA4GH) is slightly different since the data collected are epigenetic as opposed to genetic. Epigenetic data are more identifiable than genetic data because environmental exposures can be deduced, and whether the data fall under the Genetic Information Nondiscrimination ACT (GINA) is unclear. IHEC is however committed to providing free access to its epigenetic data through a data access committee. Clinically actionable results will be acted on at a local level, not internationally. The GA4GH framework is different again. The consortium has created mechanisms to support and foster data sharing across different projects and countries. Their legal instruments and guidelines don't start with the presumption that research is dangerous. Instead, their guidelines argue the inverse and take a human rights approach: everyone has a right to benefit from science and its applications. The framework asks of governments: what have you done so that I can benefit from human research? Framework took 13 legal experts to develop. Most laws to date haven't been developed with big data in mind. GA4GH is leading on this front.

In the second section of her talk, Knoppers discussed the "Impact of the EU Regulation – Legal Interventionism: GDPR and the next decade". GDPR is legislation that came into effect on May 25, 2018, and has the potential to affect projects using research data banks and big data. The law Introduces protections for data subjects with hefty penalties. It moves away from privacy and more on controlled access. As an example, see Y. Erlich *et al.*, "Identity inference of genomic data using long-range familial searches" *Science* 10.1126/science.aau4832 (2018). Law enforcement needs some rules about when they can access these data. A major challenge facing the open-access community, therefore, is if we can really re-identify, how do we protect privacy?

In the third section of her talk, Knoppers presented "Implementing the GDPR for translational medicine – Self-regulation – a code of conduct for health data". She emphasized that a code is necessary because international science collaborations have blossomed from 2009 to 2014. Rules are numerous and hard to interpret, and the public perceives a lack of oversight. She recommended an AJHG special issue on data sharing last week.

Jason Flannick: Infrastructure for analyzing and disseminating large-scale genetic data for type 2 diabetes and other complex diseases
Slides accessible on ASHG meeting site

Dr. Flannick's main message was that data storage and computation are no longer rate-limiting steps in research; we need to invent methods to disseminate and interpret high-dimensional data. As the scale of data continues to grow (e.g., the UK Biobank and All of Us research programs), the cloud will serve an important function not only with respect to data storage, but also in sharing and disseminating our findings (i.e., in preventing walled gardens). The AMP-T2D Knowledge Portal is a case study in open, big-data science. The project is a public-private partnership to advance the use of human genetics in designing new medicines, with an explicit aim to make data more accessible. Other examples of portals include the GA4GH and the NIH Data commons, which are in various stages of development. While each component of building a portal (data storage and access, computation, knowledge storage, and knowledge access) has its challenges, Flannick argued that knowledge storage and knowledge access are the key bottleneck. Computation alone is not enough to advance science: the generation of "hairball" relationship networks, or the application of "black magic" machine learning algorithms are not intuitive for scientists. Nonetheless, these tools are useful, and Flannick suggested we introduce these automated procedures gradually, akin to how technology has been gradually introduced into cars, leading us towards self-driving vehicles. In closing, Flannick argued that our ability to

analyze has outpaced our ability to interpret. We must think carefully about how to organize and represent these data.

Christina Yung: PanCancer analysis of whole genomes
Slides accessible on ASHG meeting site

Dr. Yung shared several examples of successful cloud computing, all of which were undertaken by the International Cancer Genome Consortium (ICGC). The consortium was launched in 2008 with the goal to generate and share 25,000 tumor genomes across 50 tumor types or subtypes by 2018. Data have been steadily accumulating, with a final release targeted for early 2019. Dr. Yung has been involved in the development of an ICGC data portal to query, visualize and analyze these data, which uses the Overture suite of software.

The PanCancer Analysis of Whole Genomes project (PCAWG) was launched by ICGC in 2014, with the goal to generate and share 2,800 tumor/normal whole genomes. Data are being generated across many sites, and data harmonization has been a big undertaking (e.g., in variant calling for WGS). Computing was another obstacle, which the project overcame by distributing across 14 compute clouds and high-performance clusters. Nonetheless, analyses took >2 years to complete (in part because cloud technologies were less mature in 2016). Now that data are in a useable form, important questions on tumor biology are being asked of the data. A coordinated set of publications (~60 manuscripts submitted) are currently moving through Nature & NPG journals (see https://www.biorxiv.org/search/pcawg). E.g., the data have led to the development of an "electronic pathologist" that predicts tumor type from WGS, which could aid pathologists in the clinic, help with automated QC on archival specimens, and classify based on circulating tumor DNA in liquid biopsy. PCAWG data is available for research, with dockerized workflows to ensure usability and reproducibility. These workflows bring the algorithms to the data (stored in the cloud), instead of vice-versa, since downloading sequencing data is unsustainable due to its size. The Overture suite of software again serves this role with PCAWG data, as with ICGC data. Next steps (launched in 2018) after ICGC & PCAWG: analyze tumor genomes from >100,000 cancer patients by 2028, and characterize genomic alterations in relation to high-quality clinical data including treatment and outcome.

Q&A:

- Michael Hoffman: How do we prevent people hiding behind privacy concerns to further their career?
    - Knoppers: The rare disease community leads in this respect. The community has gone a long way to respect patients' wishes that data must be shared.
    - Bourne: Data sharing is incomplete in part because mechanisms aren't in place to ensure compliance. We must push to close the loop and require documentation to show what data were deposited, where, and when. Governing bodies still struggle with this.
- Joan Bailey-Wilson (comment): Data sharing in computational genomics is commonplace. When you enroll patients, some may not want their data to be shared broadly. We need to allow those people to participate in our study. As researchers and physicians, we need to respect the participants' wishes.
- Audience member: Sometimes data are contracted with another organization to develop and return. How do we ensure compliance around where data are stored and computed upon? We may not be able to get back all the data we need in the current geopolitical climate.
    - Knoppers: Not easy, but ensure you have a plan in place that lays out how data will be destroyed when you ask it to be.
    - Jill Mesirov: What are the principal clinical applications / translation of approaches leveraging cloud computing?
    - Flannick: To help clinicians / experts interpret results of an analysis
    - Bourne: The phrase "in the clinic" relates to money when you say it to hospital administrators. What is the cost of cloud computing? Institutions can charge overhead on cloud resources but not on purchased computing equipment. Cloud services are pay-as-you-go vs. resources you put

in your own institution are paid for up front. What is more cost-effective for the hospital / institution? They might like cloud computing.
    - Yung: Hospitals don't want you to put the data elsewhere, so we need to develop new ways to analyze the data. E.g., we can distribute algorithms to different data centers and/or aggregate results.
    - Mesirov: The "Patients like me initiative" is a great example of the value of cloud computing in the clinic. The resource has lots of clinical and phenotypic data. Physicians struggling to diagnose a patient can mine the data to find patients with similar symptoms and to get treatment ideas. The resource has become very popular in California and is a great opportunity for cloud computing (i.e., this is an example of pattern matching for difficult clinical cases).
- Audience member: When the scale of datasets and analyses are so large, what do we do with false positive results? What are the clinical implications? False positives are always an issue in an automated approach.
    - Yung: Most of the data we've used to date has been in a research context so we haven't had to worry about the direct clinical implications of calling a false positive, although this is starting to become an issue (e.g., in calling a pathogenic variant). Bottom line: we need clinicians to guide us on how to make results useful in the clinic.
    - Mesirov: all results need a manual, secondary review.

IGES 2019 in Houston, USA
Save the date!

IGES 28th Annual Meeting

12-14 October 2019

Houston, TX, USA

Hilton of the Americas

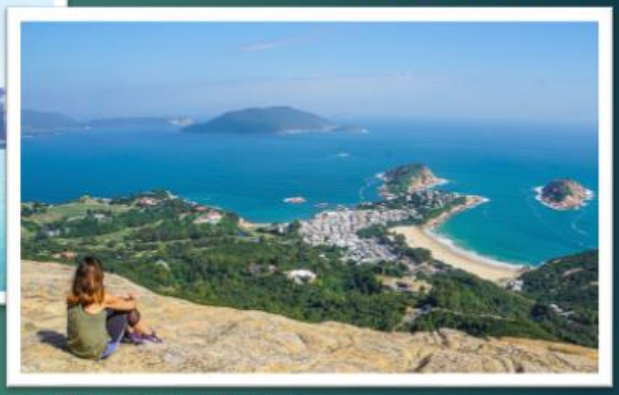
(before the ASHG Meeting)

IGES 2020 in Hong Kong
Save the date!

IGES 2020: Annual meeting of the INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY

July 1-3, 2020, Hong Kong

SAVE THE DATE!



At the Hyatt Sha Tin
Near the Chinese University of Hong Kong
20 minutes from downtown by train

Combine your IGES conference trip with:

- International Biometric Society meeting (Seoul, Korea, July 5-9) or
- 6th International Conference of Quantitative Genetics (Brisbane, Australia June 14-19)

2019 IGES Officials

The names of all the IGES officials are available on our **website:**

http://www.geneticepi.org/organization/

Current **officers** are: President: Celia Greenwood; Past-President: Inke König; President-Elect: Peter Kraft; Treasurer: Mariza de Andrade; Secretary: Heather Cordell; Editor-in-Chief, Genetic Epidemiology: Sanjay Shete.

Board members comprise the officers and the following 6 people: David Balding, Jeanine Houwing-Duistermaat, Liz Gillanders, Pak Sham, Ele Zeggini and Jaya Satagopan.

The **Education Committee** is co-chaired by Stephanie Santorico and Todd Edwards.

The **ELSI Committee** is chaired by Daniel Shriner.

This year's **Publications Committee** is chaired by Peter Kraft (ex-officio).

The **Scientific Program Committee** for 2018 is chaired by Eleanor Wheeler.

The **Young Investigators' Committee** is chaired by James Cook.

The **Communications Committee** is chaired by Heather Cordell (ex-officio).

Please refer to the above website for the current respective committee members.

The **Wiley/Genetic Epidemiology Liaison Committee** is chaired by Mike Province. Members are Sanjay Shete, Angelo Canty, and Alexander Wilson.

The **IGES webmaster** is Sarah Gagliano. The **IGES Facebook contact** is Elizabeth Blue, the **Twitter contact** is Priya Duggal and the **LinkedIn contact** is Han Chen.

Membership and conference administration is organized by Vanessa Olmo.

IGES Web Site: http://www.geneticepi.org/
IGES Facebook page: https://www.facebook.com/geneticepi?ref=hl
IGES Twitter page: https://twitter.com/genepisociety
IGES LinkedIn page: https://www.linkedin.com/groups/12061041/

IGES Facebook page exclusively for Young Investigators:
https://www.facebook.com/pages/International-Genetic-Epidemiology-Society-Iges-Next-Generation/174416209303988?ref=hl

The January 2019 Post-Election Newsletter was edited, proofread, and formatted by Heather Cordell