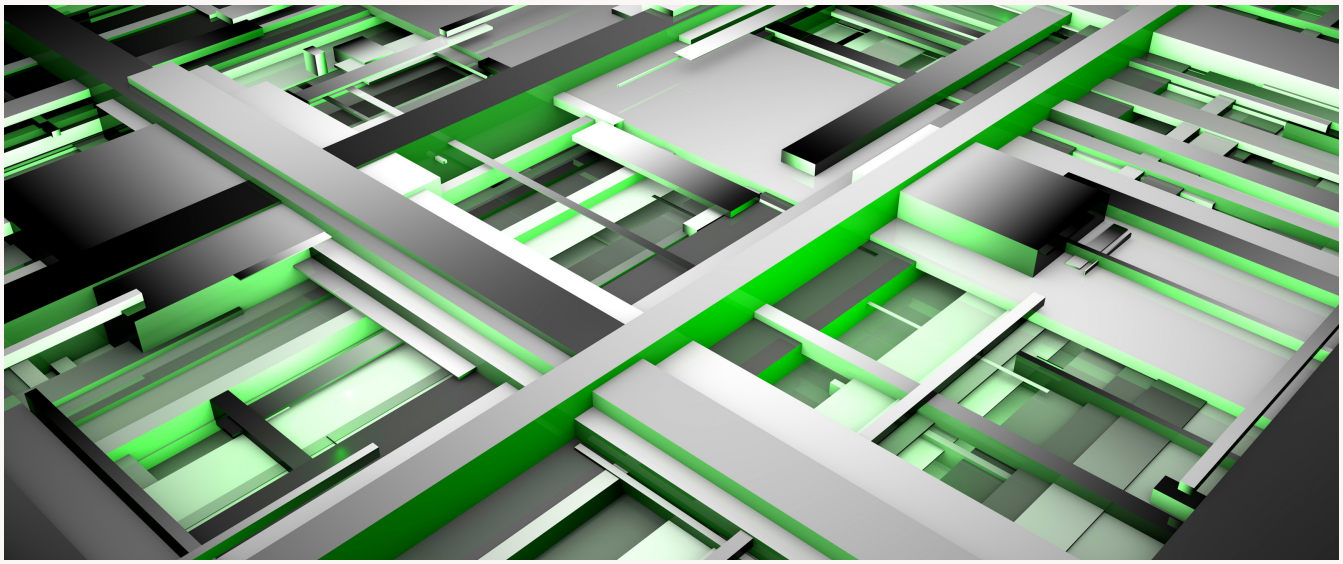
SEPTEMBER 2021 | ISSUE 3

ISCBNEWSLETTER



ISMB/ECCB 2021

We came, we adapted, we learned, and we left 2,100+ connections richer from nearly 80 countries.

ISCB jumped to the forefront of scientific virtual conference interfaces by improving on the successes and those lessons learned from 2020 to create what was the most robust and interactive virtual scientific conference within the field.

ISMB/ECCB 2021 continued to adapt to a constantly changing environment while bringing together what makes ISMB/ECCB so unique, the perfect cocktail of cutting edge science, knowledge building, tutorials, community based and networking opportunities, and so much more for this much anticipated computational biology enriched online experience.

Spanning 6 jam-packed days of pure excellence, ISMB/ECCB 2021 showcased the backbone of the conference, ISCB's Communities of Special Interest (COSIs), enabling intensified community involvement and bolstering its reputation as a conference of strong science and technology. Each of the 21 COSIs ran their own session (track) as part of the conference. All attendees could COSI track hop or pick one each session, ensuring everyone had the opportunity to attend any of these sessions in an either live or on-demand format.

in this issue

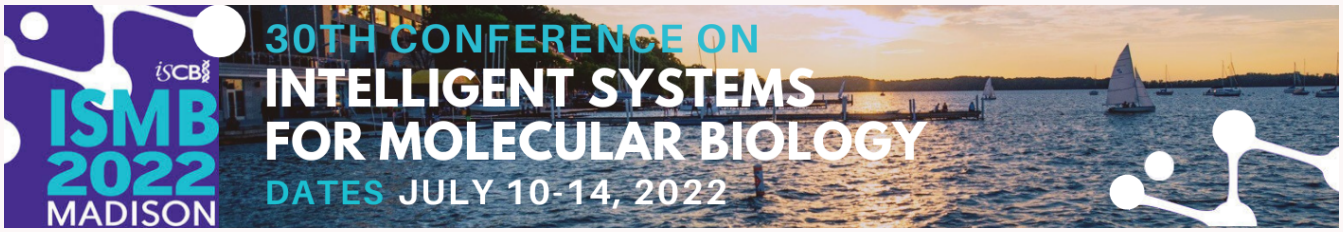
ISCB
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Throughout the six day conference, attendees had the opportunity to hear and interact live with speakers delivering 6 renowned keynote speakers, 3 of which were ISCB Award Winners and immerse themselves into a plethora of sessions that had a track and interest line for everyone. In an age of such dynamic issues and a program offering 800+ talks and posters, there were days that ran the gambit from Population scale analysis of human sequence data to Teaching the Instagram generation. No two days or sessions were ever the same. From the immensely popular COVID-19 track, EDI Panel Discussion Takes Center Stage at NIH ODSS Track, Public Affairs & Policy Committee - Science Communication and Science Journalism, to Technology Talks put on by many of our exhibitors, the level of scientific excellence was unmatched.

As is the goal each year, ISMB/ECCB brought together scientists from computer science, molecular biology, mathematics, statistics and related fields, and provided an intense forum for disseminating the latest developments in bioinformatics/computational biology in a virtual environment.

Thank you to not only our attendees but also to our exhibitors and sponsors for their support. The virtual exhibit hall and technology track demos were key features for attendees to continue learning, connecting, and networking. These are unprecedented times and it is with your support, ISCB is able to bring you this ultimate scientific experience.

Now, let us close the book on virtual only and look towards a brighter future in Madison, Wisconsin for ISMB 2022 where we cannot wait to welcome you back IN-PERSON or virtual. We are so pleased to offer a hybrid experience. Your comfort level, your choice of experience.



Renew Your Membership by 12/31/21 and be entered into a raffle to win a COMPLIMENTARY ISMB 2022 Registration!

Discount to ISMB and ECCB
Discount on OUP
Bioinformatics & ISCB Bioinformatics Advance Charges
Discount on publishing in OUP/ISCB
Bioinformatics Advances
Complimentary registration to ISCBacademy

Raffle closes 12/31 - *end 2021 stronger than it started with a plethora of opportunities and renew your membership today!*

Raffle Ticket

RENEW YOUR MEMBERSHIP & ENTER FOR YOUR CHANCE TO WIN:

100% OFF
ISMB 2022 REGISTRATION

Open Until
31 DECEMBER 2021

https://www.iscb.org/iscb-member-registration

Accessible 3D protein models to accelerate scientific discovery

DeepMind and EMBL-EBI to make millions of protein structure predictions freely available to the scientific community

Disruptive scientific breakthroughs raise more questions than they answer. They open new research avenues and can inspire entirely new fields of study. Just as the Human Genome moment marked the beginning of a revolution in genomics, so too AlphaFold might usher in a new era in biology.

By Prof. Dame Janet Thornton, Director Emeritus of EMBL-EBI

Photo Credit: EMBL-EBI

AlphaFold uses artificial intelligence to predict 3D protein structures. At the end of 2020, the CASP communityrecognised it as the first AI system to reach a level of accuracy similar to experimental models. In response, the scientific community called for DeepMind, whose scientists designed the AlphaFold system, to make the data and the computer code openly available.

The virtuous cycle of open data

DeepMind has now risen to the challenge. In collaboration with EMBL-EBI, it has made the AlphaFold protein predictions, source code and methodology freely and, crucially, openly available to the global scientific community through the AlphaFold database. The initial release contains more than 350,000 protein structures, from human and other species of biological interest, and this will expand to millions of proteins in the coming months.

Building on decades of expertise in making the world's biological data available, EMBL's European Bioinformatics Institute (EMBL-EBI) is working with DeepMind to ensure the predictions are Findable, Accessible, Interoperable and Reproducible (FAIR) so that researchers everywhere can make the most of them.

AlphaFold was trained using data from public resources – including UniProt, PDB and MGnify, which are co-hosted at EMBL-EBI – so it's very fitting that its predictions are now openly available to all. This is a perfect example of the virtuous cycle of open data. By sharing data, the community can drive discovery faster than any one individual. Open data benefits all: public and private, experimental and computational, basic and applied research.

A wealth of opportunities

This ability to predict protein structure with unprecedented accuracy will underpin a revolution in biology as it allows us to understand better how all living things work. AlphaFold has many applications relevant to human health, agriculture and climate change.

By providing high-quality 3D structures for almost all human proteins, AlphaFold also frees structural biologists to focus their work on the more exciting questions of how proteins interact and function – something that AlphaFold doesn't currently predict.

Enzymes, which are also proteins, are nature's catalysts, but they are very difficult to design in a lab. Protein structure predictions can help scientists to design new enzymes, with new functions, such as processing waste or degrading plastics. Accurate protein structure predictions can also pave the way to improving crops so that they can handle climate change.

The possibilities for applications related to human health are endless, for example tackling some of the most serious diseases by predicting the structures of the proteins involved, characterising how they interact, and understanding how they cause disease. New proteins could be designed for novel vaccines or biological therapies to modulate diseases, and new candidate drugs can be identified more effectively.

Experimental researchers will be able to accelerate their structural studies to focus on complex biological systems, where experimental structural data at very high resolution are difficult to obtain.

A note of caution

While it's true that AlphaFold is, so far, the gold standard for protein prediction, there are limitations to the method and the database, and these are important to note.

Almost all proteins function by interacting with other proteins, nucleic acids (DNA or RNA) or small molecules. AlphaFold doesn't currently predict such complexes.

Proteins are also dynamic systems, with disordered regions that adapt their structure to their environment. Their dynamics and folding 'from scratch' have yet to be elucidated.

There are certain protein regions where AlphaFold produces only a low-confidence prediction (often for disordered regions). The AI system provides a confidence score as a helpful guide. Furthermore, AlphaFold has not been trained for predicting the effect of mutations, which can be critical in understanding why some individuals are susceptible to certain diseases. So like any method, AlphaFold will have its limitations that will inspire new and exciting avenues of research.

AI as a tool for science

AlphaFold has illustrated the power of AI to improve 3D protein structure predictions. It complements existing methods and reveals new insights, but does not replace experimental methods to determine structures. This work serves as an exemplar of what is possible - and it is clear that AI will find many such applications in broader scientific research.

The power of AI underlies the AlphaFold predictions, based on data gathered by scientists all over the world during the last 50 years. Making these models available will undoubtedly galvanise both the experimental and theoretical protein structure researchers to apply this new knowledge to their own areas of research and to open up new areas of interest. This contributes to our knowledge and understanding of living systems, with all the opportunities for humanity this will unlock.

Read more

Explore the AlphaFold database.
Read more about the scientific implications of the AlphaFold database launch.

Source papers

Tunyasuvunakool, K., et al. (2021). Highly accurate protein structure prediction for the human proteome.Nature. Published online 22 07; DOI: 10.1038/s41586-021-03819-2
Jumper, J., et al (2021). Highly accurate protein structure prediction with AlphaFold. Nature. Published online 15 07. DOI: 10.1038/s41586-021-03819-2

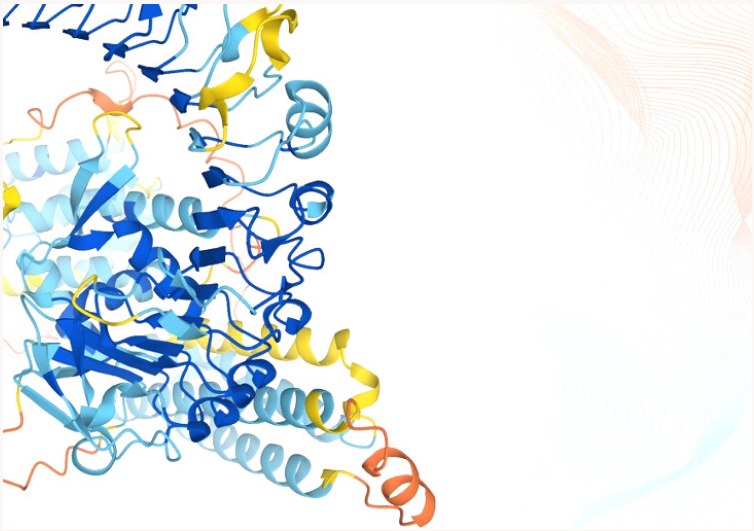


Photo Credit: Karen Arnott

ISCB Issues Inaugural Equity, Diversity, and Inclusion Report

In late 2017, a small group of passionate ISCB members challenged the ISCB leadership to better address equity, diversity, and inclusion within the professional society. Over the course of the next four years, this passionate group led strategic discussion with the ISCB Board of Directors to set a path for the Society to make equity, diversity, and inclusion part of its fabric.

The ISCB Equity, Diversity, and Inclusion (EDI) Committee has prepared ISCB's first Equity, Diversity and Inclusion Report (access report here) as well as ISCB first Equity, Diversity, and Inclusion Strategic plan (access strategic plan).

ISCB Equity, Diversity, and Inclusion Strategic Plan

Developed in 2020, the equity, diversity and inclusion strategic plan focuses on the following main areas:

- **Increasing social accountability for change in the ISCB society**
- **Obtaining data and developing measures to assess progress**
- **Voluntary training: The "ISCB Awareness toolkit"**
- **Recruitment initiative**
- **Mentoring**

To date, the committee has successfully guided the Society and implemented three of the main focus areas through programs, reports, statements, and resource development. Equity, diversity, and inclusion is top of mind for the organization. ISCB started small with a focus on increasing gender balance within the conference; organizers and steering committees now are focusing on diversity beyond gender.

ISCB Equity, Diversity, and Inclusion Annual Report

A key priority of the strategic map was the collection and measuring of data. Data collection is voluntary via the diversity survey within the ISCB membership profile or when registering for ISCB events. The use of the data is strictly used for benchmarking.

The committee recognized that the data presented in this report may not yet accurately capture diversity within ISCB or the greater computational community (response rate on the survey was over 81% on gender and 56% on ethnicity), but felt comfortable with the level of response to make recommendations to improve diversity within ISCB. The committee hopes that the ISCB members will continue to complete the diversity report when renewing their membership or registering for conference to increase the accuracy of the data.

A key priority of the strategic map was the collection and measuring of data. Data collection is voluntary via the diversity survey within the ISCB membership profile or when registering for ISCB events. The use of the data is strictly used for benchmarking. The committee recognized that the data presented in this report may not yet accurately capture diversity within ISCB or the greater computational community (response rate on the survey was over 81% on gender and 56% on ethnicity), but felt comfortable with the level of response to make recommendations to improve diversity within ISCB. The committee hopes that the ISCB members will continue to complete the diversity report when renewing their membership or registering for conference to increase the accuracy of the data.

ISCB invites you to review the report in its entirety. The major findings and recommendations included:

1. A deeper look at the award process is necessary. The report indicates a bias during the selection process that is most prominent at the nomination stage. As ISCB awards are nomination based, the only way to improve this bias is to address it at the nomination stage.

2. Include completion of the ISCB awareness toolkit in membership profile survey and offer incentives for members to do so (conference ribbons, banners), especially for members that serve on ISCB committees

3. Set a goal to move forward with offering as many conferences as financially feasible as hybrid events - in-person and virtual attendance.

4. Collect demographic info for abstract submitters, talk selections, invited talks, etc. for all ISCB-associated conferences. Our initial assessment supports the existence of bias in recognition in our community. Continuing to collect more and better data will help create better strategies on how to address it.

5. Encourage empirical research into equity, diversity and inclusion within computational biology science.

The ISCB leadership has already started to take actions on fulfilling the recommendations from the report. We continue urge the greater computational biology to consider completing the diversity survey when becoming or renewing your membership and when registering for conferences. We also strongly encourage the committee to review the Awareness Tool Kit and other resources that are available on the ISCB Equity, Diversity and Inclusion webpage (https://www.iscb.org/iscb-edi). Change takes time but if we all work together we can make progress more quickly.

ISCB
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Equity, Diversity, and Inclusion 2020-2021 Annual Report

Vol. 1

2021 ISCB Rocky Mountain Bioinformatics Conference

Registration Open

https://www.iscb.org/rocky2021

The eighteenth conference is a meeting of the International Society for Computational Biology (ISCB) held each year in Snowmass, Colorado. The Rocky series began nineteen years ago as a regional conference, and has grown into an international program with a spotlight on regional development in the computational biosciences.

The conference includes short "flash" presentations (10-minute talks), poster presentations and keynote presentations on current projects including significant works-in-progress involving the application of advanced computational methods to significant problems in biology or medicine.

The presenters of the Rocky conference are early and late career research scientists representing a broad spectrum of universities, industrial enterprises, government laboratories, and medical libraries from around the world.

The conference is a chance to get to know your colleagues near and far, seek collaborative opportunities, and find synergies that can drive our field forward.

In-Person December 2 - 4, 2021
Viceroy Hotel - Snowmass/Aspen, Colorado

EDI Panel Discussion Takes Center Stage at NIH ODSS Track

Track Highlights from ISMB/ECCB 2021

Dr. Susan Gregurick, Director of NIH ODSS, organized and co-chaired with Dr. Karol Watson, Professor of Medicine in the Division of Cardiology at the David Geffen School of Medicine at UCLA, a vibrant panel discussion on "Diversity in Data Science Training and Research."

An array of data science experts from underrepresented groups, coming from academic institutions (e.g., Hispanic-Serving Institutions, HSIs; Historically Black Colleges & Universities, HBCUs), legal industry, and technology industry tackled critical questions focusing on issues of equality, diversity, and inclusion (EDI) in multiple dimensions of data science training and research:

How do we define, measure and operationalize EDI? EDI concepts can mean different things in different settings, and there is no “one size fits all” approach. What must we do to gain support for EDI beyond procedural compliance or symbolic initiatives? What mechanisms do we apply to increase the diversity and/or cultural competence of those involved in data science training? What are the appropriate metrics for evaluating our progress?

The panelists not only raised questions on fundamental issues relating to EDI in data science training and research, but offered thoughtful solutions to these complex problems, defining a path forward to overcome the gaps and disparities in data science training. Their overall heartening messages and lessons learned in supporting and mentoring young EDI investigators was an inspiration to all.

ISCB Committees Getting to Work at ISMB/ECCB 2021

Public Affairs & Policy Committee - Science Communication and Science Journalism

The COVID-19 pandemic has brought science into the limelight of public perception in an unprecedented fashion. Scientists have become more present in the media, accelerated transmission of new results has been effected by the public starting to pay attention to unreviewed research reports, journalists have been faced with the requirement to bring such emerging results to the public on short notice. Today's panel moderated by Thomas Lengauer, ISCB's Immediate Past President, focused on challenges facing both scientists and members of the media. Panelists, Kai Kupferschmidt, Independent Science Journalist, Emma Hodcroft, University of Bern, John Moult, Institute for Bioscience and Biotechnology Research (IBBR), and Freya Robb, Science Media Centre, joined to discuss issues on how to best effect communication to the public on science results and also and especially on the science process.

Looking to get involved with ISCB? Check out ISCB Committee page: https://www.iscb.org/iscb-committees and reach out to admin@iscb.org today!

Green ISCB: ISCB Groves

Join us as we work towards healing our environment

The International Society for Computational Biology (ISCB), in partnership with All Things Small and Green and Trees for Life, has set up its first grove of native trees, the ISCB 2020 grove, to offset carbon emissions caused by ISCB activities such as those emitted during the conferences, workshop or panel meetings that it organizes. With this grove, ISCB members are also offered a simple and convenient way to offset their own carbon emissions such as those incurred during traveling to ISCB conferences or workshops.

COSI Session Highlights from ISMB/ECCB 2021

ISMB/ECCB 2021 was made up of 21 COSI sessions. These sessions are the backbone of the conference - don't miss next years line-up in Madison!

HiTSeq

Our first Hitseq session on Sunday 25th featured new methods for the analysis of genomes and transcriptomes using long reads. The adaptive sampling during nanopore sequencing by Weilguny allows optimal usage of sequencing effort and we discovered how Bionano long reads can be used to study the coupling of promotor and enhancers methylation. The statistics of kmers was for the first time discussed, new graph-based methods to assembly and compact genomes were presented and we learnt about the structure of the centromere with Olga Kunyavskaya. Our highlight was our keynote Angela Brooks who presented her lab's effort to characterize isoform changes in cancer using Nanopore and the LRGASP challenge for evaluation of long-read RNA sequencing methods.

CompMS
Computational Mass Spectrometry

The CompMS COSI wrapped up a successful two days on Tuesday. The first day focused on computational metabolomics, and started off with a keynote by Dr. Theodore Alexandrov on how computational mass spectrometry enables spatial and single-cell metabolomics. On the second day Dr. Vadim Demichev shared his views on recent advances enabling high-throughput proteomics during the opening keynote, and during the closing keynote Dr. Ed Huttlin presented the BioPlex project for biological discovery of the 'social network' within a human cell.

Additionally, the program featured 15 oral presentations during which speakers discussed a wide range of computational mass spectrometry metabolomics and proteomics advances. Topics reached from developments in single-cell metabolomics, over machine-learning applications in proteomics, all the way to exciting new developments for screening interactome data.

biovis

BioVis started a two day sessions with a keynote by Seán O'Donoghue about inter-disciplinary practices in biological visualization. Seán shared his experiences using mixed practices in approaching several scientific challenges for instance, in multi-omics and genetic variation analysis, and shared reflections on establishing visualization as a powerful technique. The keynote was followed by an invited talk delivered by Liz Marai on their RuleBender paper that received our newly introduced Test of Time Award, which is given out to an influential paper published at BioVis at least 10 years ago. The following two sessions focused on visualization in the field of genomics as well as cellular and omics data. Besides short talks for abstracts, there was one ISMB proceedings talk and an invited highlight talk given by Jen Rogers.
The second day started with talks focusing on the integration of visualization and machine learning techniques. While some of the talks explored explainability issues within deep learning techniques for drug repurposing and analysis of high-dimensional data, others looked into how hierarchical visualization and association-focused techniques could support the analysis of transcriptome data and single cell RNA-seq datasets. The next set of talks, featuring another ISMB proceedings and an invited highlight talk by Helena Jambor, introduced new techniques to visualize imaging data using some innovative techniques such as visual exemplars, and novel glyph based representations that serve as pictorial summaries, as well as empirical research outlining best practices for producing effective images in publications. The closing session featured a keynote talk by Jessica Hullman who shared her research and insights on uncertainty visualizations, and how models of inference can transform our understanding of how people reason with data visualizations.

Spread over three days, and attracting approximately 100 attendees in each session, RegSys presented cutting-edge research in regulatory & systems genomics. We enjoyed exciting keynote talks from Céline Vallot, Fabian Theis, Anaïs Baudot, Olga Troyanskaya, Camille Berthelot, and Juanma Vacquerizas, along with four proceedings papers and seventeen contributed talks.

Day 1 focused on advances in characterizing cell type-specific regulatory networks using single cell assays. The keynotes, in particular, demonstrated the power of applying single cell technologies to characterize epigenomic features of cancer drug resistance, and regulatory relationships in large-scale perturbation screens. Day 1 closed with several interesting talks that featured different ways to characterize cooperative interactions between transcription factors.

Day 2 broadened the track themes. Inspirational keynotes integrated multi-modal data types using increasingly powerful computational techniques, yielding insights into the mechanisms underlying human diseases. Day 2 contributed talks spanned themes such as profiling cell-specific splicing events, determining the impact of viral integration events, and analyzing nucleosome features on cell-free DNA.

Day 3 turned to the theme of comparative regulatory genomics, with a keynote and contributed talk characterizing how cell-specific enhancers have evolved across mammals. Day 3 also focused on advances in profiling the three-dimensional organization of the genome. Talks demonstrated methods for increasing the resolution of chromatin interactions, while a keynote cast doubt on whether genome organization influences gene expression.

A steady theme throughout all days and sessions was the growing adoption of neural network techniques in regulatory genomics, with interesting methodological advances demonstrated in several application domains. We had a very enjoyable virtual program overall, and we encourage the RegSys community to continue our conversations on the RegSys slack channels: https://join.slack.com/t/regsyscommunity/shared_invite/zt-tprp1q5g-pNcDNiFUXyJQacdfY1nGWg

TM COSI featured two outstanding keynotes, one by Dr. Pysalo (University of Turku) highlighting the evolution of transformers in the biomedical field and opportunities for collaboration. The second keynote was a joint presentation by Dr. Ball (FDA) and Dr. Hirschman (MITRE), reporting on the quest for text mining methods to modernize post-market drug safety assessment at FDA, the needs and challenges, followed by a description of results from the evaluation of methods for automated extraction of adverse drug events, and experiments with humans-in-the-loop. The session selected talks covered applications of text mining methods (both traditional machine learning and Deep learning) across various fields, including -omics (metagenomics, annotations of -omics samples), protein folding, clinical discharge letters, biocuration, and science communication. The COSI ended with the text mining spotlight section with two talks. First, Dr. Lu described efforts at NCBI to bring to the community freely available tools and methods that have been developed in light of their own needs, with a focus on LitSuggest, a web-based system for biomedical literature recommendation and curation, and on TeamTat an annotation tool for team collaboration. Finally, Dr. Poon from Microsoft Health Futures provided an overview of PubMedBERT, and its advantages over other BERT-based models in the biomedical domain.

Evolution and Comparative Genomics (EvolCompGen) COSI ran a successful 3 day session starting off with 8 talks. The talks illustrated the broad use of phylogenetic trees and networks, such as calibrated species tree inference, individual cell lineage reconstruction, or admixture event inference. Consistent with the zeitgeist, we also witnessed creative uses of machine learning approaches for comparative genomics analyses—to speed up orthology prediction, to detect copy number variation, and to infer co-evolving genes from phylogenetic profiling. The second day at the Evolution and Comparative Genomics (EvolCompGen) COSI featured eight talks and several posters coming in from at least three continents.

They spanned a variety of subjects from the evolution of proteins, domains, gene content, tumor cells to evolutionary events such as horizontal gene transfer, endosymbiotic gene transfer, copy number alterations, tandem duplication of protein domains and DNA repair through non-homologous-end-joining. We also got to see more examples of machine-learning applied to evolutionary analysis and an exciting webapp for characterizing protein using molecular evolution. Finally, we ended the day with more discussion of domain evolution and linguistic methods applied to cancer. While we miss in-person interactions, we have had a good time connecting with COSI members on Cafe Connect and Slack! The final day of the EvolCompGen COSI, we featured seven enriching presentations on a variety of topics concerning the evolution of genomes, genes, or at the intermediate level of "syntenies" (chromosomal gene clusters). At the gene level, we discussed quantification of introgression (gene flow between species), convergent evolution, phylogenetic profiling, and the identification of new metabolic pathways. At the genome level, we had talks on whole-genome duplication, and more specifically an innovative method for allo-polyploidy versus auto-polyploidy prediction was highlighted. A new model to infer tumor phylogeny and call variants using single-cell data was also presented. The closing talk was on Super-Reconciliation, a novel method for inferring the evolution of syntenies through segmental duplications and HGTs. This presentation made the bridge with the panel discussion, where we discussed a variety of open questions on ancestral genome reconstruction, orthology prediction, and unifying methods for synteny evolution.

The capstone of this year's Evolution and Comparative Genomics track was a panel discussion focusing on the evolution of chromosomal gene clusters, ancestral genome reconstruction, and phylogenetic comparative methods. A spirited discussion ensued touching on tradeoffs between computational costs and accuracy and how the rapid growth of genomic data sets influences this tradeoff, the pros and cons of parsimony versus probabilistic models, hybrid approaches that attempt to assess uncertainty while preserving the speed of parsimony and how such methods can be used to infer convergent evolution, and, finally, whether convergent evolution introduces noise in phylogenetic inference. These discussions aroused so much interest and generated such a lively exchange amidst the panelists, moderators, and participants, that we continued brainstorming for another full hour in a new Zoom e-space after the session ended officially. We wrapped up our three-day COSI enriched with new ideas and new collaboration perspectives.

BOSC 2021 (https://www.open-bio.org/events/bosc-2021/), the 22nd annual Bioinformatics Open Source Conference, kicked off day 1 with opening remarks by BOSC Chair Nomi Harris, speaking from her home in California at 4:00am local time. The opening session also included an overview of the Open Bioinformatics Foundation by OBF President Peter Cock. The first BOSC 2021 keynote talk, "Significant heterogeneities: Ecology's emergence as open and synthetic science", was delivered live by Christie Bahlai. Ecology has not been a big topic at past BOSCs, but Christie's approach to open and inclusive science strongly resonated with our community. One attendee commented, "Christie Bahlai's keynote was inspiring and a breath of fresh air. +1 to calling out open science purists and the call for collective responsibility." Christie's keynote was followed by a session on Standards and Practices for Open Science. One speaker in that session, Dhrithi Deshpande, spoke about the disappointingly low percentage of scientific papers that provide a link to the code (only about 12% -- though up from only 1% in 2016!), and noted that articles that share code tend to get cited more.

The next session, Tools for Open Science, started off with a well-received talk by Thorin Tabor about GenePattern, a reproducible research platform built on top of Jupyter Notebook.
Day one's final BOSC session was a joint session with Function COSI, chaired by Iddo Friedberg, featuring a keynote talk by Lara Mangravite on "Open approaches to advance data-intensive biomedicine." 210 participants attended this joint session. Lara noted that clinical applications need access to high-quality data, but broad accessibility of human clinical data is difficult due to privacy issues, and access to analysis capabilities is distributed inequitably and tends to leave out those in the global south. Lara discussed some possible mitigations for those challenges.
Birds of a Feather, which bring together participants to chat informally about a shared interest, are always a popular feature of ISMB. In the BoF "Next steps for computational reproducibility toward fully executable papers", there were discussions around 'What is reproducibility?" and how do we incentivize new PIs to make reproducible work? Some cloud frameworks like Terra allow for reproducibility and lower the barrier for those who don't have computational infrastructure in house.

On the other hand, what about cloud costs? And is the reason why more papers aren't reproducible the lack of skilled software engineers on the team?
The day ended with a hangout in the BOSC roundtable (which quickly ran into the 16-person maximum) that evolved into a discussion of the OBF and new ways for it to serve the open source bioinformatics community.
The final day BOSC 2021 closed with an enlightening talk by the last of our three keynotes, Thomas Hervé Mboa Nkoudou speaking about "Contribution of the maker movement to biotechnology in Africa: An open science perspective" (in French with English subtitles).

Virtual CAMDA 2021 took off to a full house, with the well over 100 delegates likely grateful that they didn't have to cram into a real room, while reducing their carbon footprint by 50 tons of carbon dioxide.
Opening the session, Professor Francesca Ciccarelli of the Francis Crick Institute highlighted advances in predicting cancer driver genes not just at the cohort level but for individual patients, and their cumulative effects in age. This was followed by a variety of contributions by CAMDA delegates from Cambridge, Isarel, Kyiv, and Stanford, who exploited an impressive variety of modern AI language models to identifying scientific papers relevant to drug induced liver injury, as benchmarked by a CAMDA challenge organized jointly with the US FDA with support of IARAI Vienna.
The session was rounded off by reports on testing the robustness and reproducibility of computational genomics tools, an in silico evaluation of SARS-CoV-2 primer performance, and an intriguing study of mechanistic models of COVID-19 infection and antiviral interventions. Join us on Thursday for keynotes by Director Weida Tong (NCTR/FDA) about AI myths and their implications on regulatory science, and Director Nikos Kyrpides (DOE Joint Genome Inst) cover Microbiome Data Science from the Earth Microbiome to the Global Virome.

The 2nd day of virtual CAMDA 2021 started with an exceptional talk by Weida Tong, director of the Division of Bioinformatics and Biostatistics at the NCTR of the US FDA. Dr Tong discussed five common myths about AI and their implications for regulatory science, and provided a broad overview of how the FDA uses AI/ML techniques in support of the prediction of Drug Induced Liver Injury (DILI).

This was followed by further presentations employing a variety of AI language models to identifying scientific papers relevant to drug induced liver injury, as benchmarked by a CAMDA challenge organized jointly with the US FDA with support of IARAI Vienna.
The seond half of the day was devoted to the Metagenomic Phage Forensics of Anti-Microbial Resistance challenge based on recently published MetaSUB Consortium data, which was explored by the CAMDA community this year for the first time, yielding novel complex relationships between phages and AMR that indicated a new for further investigations. The rich world of modern meta genomic data was then brilliantly characterised in our keynote by Nikos Kyrpides from the DOE Joint Genome Institute, ranging from the Earth microbiome to the Global virome.

CAMDA's final day ended with an introduction to Disease Maps for modelling COVID-19 by Maria Peña Chilet of the FPS in Seville, followed by contributed talks of the matching data analysis challenge. The final sessions hosted ISMB/ECCB proceedings presentations on using AI/ML in cancer studies as well as AI-driven Cloud Laboratories. The day wrapped with the CAMDA Cafe where the CAMDA community discussed the Grand challenges of our times to tackle in the coming years, and finished with the awards ceremony for the traditional CAMDA Trophy for the best presentation.

The iRNA COSI 2021 kick-off was given by Manuel Irimia who provided a fascinating keynote on the evolution of alternative splicing networks and the extent of their tissue conservation. Today's contributed talks focused on mRNA stability, circular RNA evolution and prediction, the study of transcriptomics in health-related applications and tools for the analysis of 3D RNA structures. Our first day ended with entertaining flash talks / quiz and poster session.
The second day of the iRNA COSI 2021 was dominated by epitranscriptomics. The keynote was given by Kathi Zarnack from the Buchmann Institute for Molecular Life Sciences in Germany who discussed the detection/prediction of m6A RNA modifications using transcriptomics data and machine learning, identifying sites within consensus DRACH motifs but also others, providing a rich resource for further studies.

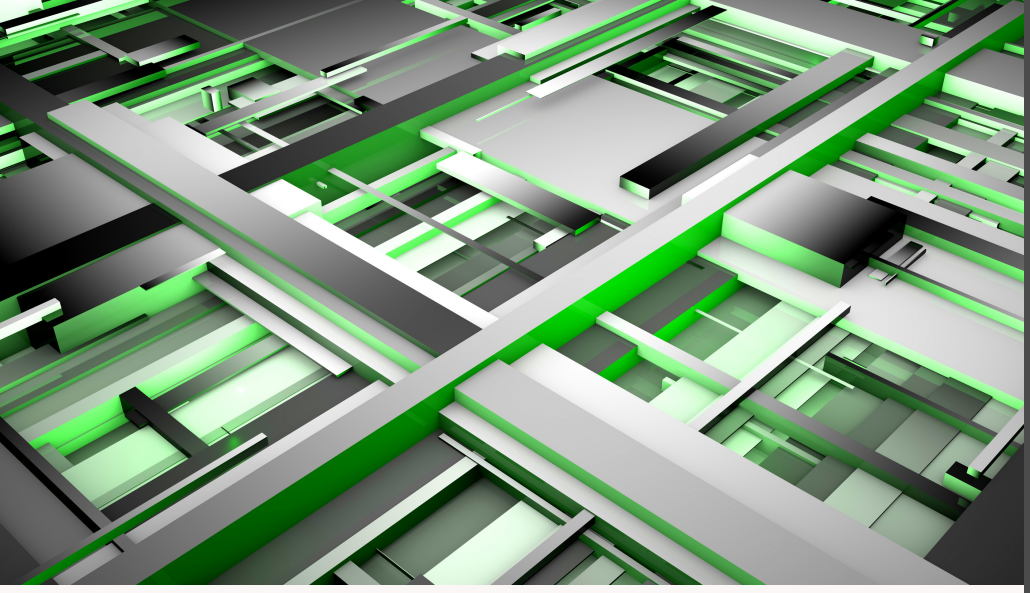
Two contributed talks on splicing kinetics and modeling of multivalent binding by RNA-binding proteins were followed by five talks on RNA modification detection from sequencing, raising important common issues and challenges which were addressed in our live panel discussion, to which participated Nicole Martinez, Shengdong Ke and Schraga Schwartz. The panel highlighted the fact that while a decade ago, the challenge in epitranscriptomics was to determine where the modifications are, with rapid and considerable increases in technology capacity brought on by innovations in the field, we are now switching gears and focusing on questions like the function of the modifications, how to address the interconnectedness of the process and the effects throughout the life cycle of the targeted RNA, and the use of systematic perturbations and their analysis using machine learning to understand what the modifications are really doing in the cell. Our panel was followed by our second poster session and our traditional social time with quiz during which we took time to catch up and learned much trivia about the Olympics including surprising facts.

The third and last day of the 2021 iRNA COSI had an emphasis on RNA structure/RNA interactions/RNA quantification using long reads and featured two fascinating keynotes. Yue Wan discussed mapping of RNA-RNA interactions with applications illustrated in the analysis of SARS-CoV-2 while Liang Huang talked about the prediction of RNA secondary structure, with interesting historical considerations and comparisons across fields, finishing with the important application of their linear tool to SARS-CoV-2. Contributed talks discussed the transcriptomics analysis of a lncRNA, a snoRNA and a miRNA network as well as important considerations in the analysis of long read and direct RNA sequencing. In addition, the community heard about two challenges started within the context of the RNA society meeting with calls for participation. Overall, the 2021 iRNA COSI was very successful with varied aspects of computational RNA research explored, a timely live panel discussion, fascinating keynotes and enthusiastic poster presentations.

During the first day of the VarI meeting we had two keynote talks. Alessandra Carbone (Sorbonne University) presented her recent work on the analysis of evolutionary information for detecting protein sites in viral proteomes important for protein-protein interaction and drug resistance.

Later, Ben Langmead (John Hopkins University) presented methods for reducing reference bias in the detection of human variations. The proposed approaches are based on new algorithms that allow to align sequencing reads to the reference genome of different populations. The VarI meeting also hosted a proceeding talk from Chirag Jain who presented a new approach for minimizing variation graph size that allows to reduce the computational complexity in the methods for reducing genome bias. In the first days the program included 6 selected presentations focusing on variant prioritization, sex dependent genotype association studies and pleiotropy. Finally, we had a poster session with 18 works presented.

VarI COSI's second and final day at ISMB ECCB 2021 started with a keynote presentation titled "Mutate everything" by Ben Lehner from the (Centre for Genomic Regulation, Barcelona). Ben presented recent work on mutational scanning of protein coding genes allowing to quantify mutations' effects on protein folding, binding and aggregation as well as to characterize protein free energy changes for the identification of allosteric sites. Mutational scanning was also a main topic of the round table discussion moderated by Yana Bromberg (Rutger University) with Douglas Fowler (University of Washington), Daniel Gilchrist (NHGRI) and Predrag Radivojac (Northeastern University), where current challenges on variant annotation and interpretation were addressed. Adding to the round table, a total of 7 selected talks were presented with a major focus on clinical interpretation of human variants from sequencing studies, covering a broad range of aspects such as the structural and mutational features of pathogenic variants, the interpretation of Copy Number Variants, and the characterization of gain-of-function variants. The session included a presentation from Variantyx, VarI COSI's main sponsor, where Alexander Kaplun presented recent computational developments for variant calling in non-uniquely mappable regions from short-read WGS data.



Upcoming ISCBacademy Webinars

An online webinar series including the ISCB COSI, COVID webinars, Indigenous Voices and practical tutorials. We aim to inspire, connect, and communicate the science while providing a hands-on experience accessing and using newly developed bioinformatics tools while ensuring best practices for rigour and reproducibility.

November 19, 2021 at 11:00 UTC - *Elements of Style in Reproducible Workflow Creation and Maintenance: A Hands-on Tutorial* by Anne Deslattes Mays and Christina Chatzipantsiou - Hosted by ISCB

November 23, 2021 at 11:00 AM EST *Robust single-cell discovery of RNA targets of RNA-binding proteins and ribosomes* by Kristopher Brannan, University of California at San Diego - Hosted by iRNA and RNA Society

December 9, 2021 at 11:00 AM EST *DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of enhancers* by Bernardo Almeida, Research Institute of Molecular Pathology - Hosted by MLCSB

https://www.iscb.org/iscbacademy-upcoming

ISCB
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY