GSRS22 Bioinformatics Session on Oct 20th (Thursday)

Theme: Regulatory Data Science for Food, Drug, and Beyond

Venue: Lavendar room at Orchard Hotel Singapore (https://gcrsr.net/2022-gsrs/);

Time/Date: 8:00 am-5:00 pm, Oct 20, 2022 (the GSRS22 conference is between Oct 19-21)

8:00 am – 8:40 am Session 0: Welcome and Session Overview

Co-Chairs: William Slikker (Former GCRSR Chair, USA) and Weida Tong (NCTR/FDA, USA)

- AI4TOX – An FDA Artificial Intelligence (AI) program for toxicology**,** Weida Tong (NCTR/FDA, USA)

8:40 am – 10:20 am Session 1: Food and Drug Safety

Co-Chairs: George Kass (EFSA, Italy) and William Slikker (Former GCRSR co-chair, USA)

1. Data science for food safety: how to integrate new streams of data in the risk assessment process? George Kass (EFSA, Italy)
2. Text Analytics for Food Safety Monitoring, Benjamin Er (National Centre for Food Science, SFA, Singapore)
3. Use of Machine-Learning and Artificial Intelligence for Drug Toxicology, Peter Newham (Clinical Pharmacology and Safety Sciences, AstraZeneca, UK)
4. Data science in early derisking of drug targets and discovery chemistry, Ruth Roberts (University of Birmingham, UK)
5. Panel Discussion and Q/A

10:20 – 10:40 – Break

10:40 am – 2:40 pm Session 2: Standards and Best Practice

Co-Chairs: Tim Gant (UK Health Security Agency, UK) and Maurice Whelan (JRC, European Commission)

1. Open, FAIR and reproducible science, Susanna Sansone (Oxford University, UK)
2. Taking omics from concept to application in chemicals regulation, Tim Gant, (UK Health Security Agency, UK)
3. Reference materials and datasets for improved reproducibility, Yuanting Zheng (Fudan University, Shanghai, China)
4. Bridging scientific data with evidence needs for regulatory safety assessment, Maurice Whelan (JRC, European Commission)

12:00 – 2:00 – Lunch

5. Two decades of effort in developing best practices for genomics and its contribution to regulatory science, Leming Shi (Fudan University, Shanghai, China)
6. Panel Discussion and Q/A

2:40 pm – 4:40 pm Session 3: Innovative Application

Co-Chairs: Michael Renaudin (Swissmedic, Switzerland) and Lam Kwok Yan (Nanyang Technological University, Singapore)

1. Tricia - Leveraging NLP to enhance risk assessment of incoming incident reports, Alexander Horst (Swissmedic, Switzerland)

2. Medicrawl – Crawling online marketplaces for illegal products, Nicolas Perez Gonzalez (Swissmedic, Switzerland)

3:20 pm – 3:40 – Break

3. AI empowers the assessment for hepatotoxicity potential and clinical endpoints in drug development, Wenjun Bao (SAS, USA)
4. Assessing allergenicity risk of proteins with AllerCatPro 2.0, Sebastian Maurer-Stroh, (Bioinformatics Institute, Agency for Science, Technology & Research, Singapore)
5. Panel Discussion and Q/A

4:40 pm – 5:00 – Closing Remarks and/or Final Q/A (discussion)

William Slikker Jr., Ph.D.
Former director of National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. William Slikker, Jr. was the director of FDA's National Center for Toxicological Research (NCTR) before his retirement. He received his Ph.D. in pharmacology and toxicology from the University of California at Davis. Dr. Slikker holds adjunct professorships in the Department of Pediatrics, as well as the Department of Pharmacology and Toxicology at the University of Arkansas for Medical Sciences. He has held committee chairmanships or elected offices in several scientific societies including the Teratology Society (serving as president) and the American Society for Pharmacology and Experimental Therapeutics (chair, Developmental Pharmacology Section and member, Program Committee). Dr. Slikker is also the co-founder and past president of the MidSouth Computational Biology and Bioinformatics Society. He is currently associate editor for NeuroToxicology and associate editor for the "Environmental Health" section of Experimental Biology and Medicine. He is the past president of The Academy of Toxicological Sciences and the Society of Toxicology. He is a recipient of the 2014 George H. Scott Memorial Award from The Toxicology Forum and was invited to present the Warkany Lecture at the 2015 annual meeting of the Teratology Society. In early 2019, the Academy of Toxicological Sciences selected Dr. Slikker to receive the prestigious Mildred S. Christian Career Achievement Award. The Society for Birth Defects Research and Prevention selected Dr. Slikker to be the recipient of the 2022 Edward W. Carney Distinguished Service Award.

Dr. Slikker has authored or co-authored over 380 publications in the areas of transplancentalpharmacokinetics, developmental neurotoxicology, neuroprotection, systems biology, and risk assessment. Dr. Slikker's recent research has highlighted the concern for thousands of infants and toddlers who undergo longer-duration general anesthesia. He has performed research with his team and published over 25 peer-reviewed papers outlining the issue of brain-cell death and cognitive-function deficits in animal models that may result from several hours of anesthesia at a critical time of development. He has also, with the use of in vitro and in vivo techniques in rodents and nonhuman primates, defined possible mechanisms of toxicity and protective pathways to prevent the detrimental effects of general anesthesia. Through these and related scientific contributions, he has identified and characterized a host of minimally invasive biomarkers of neurotoxicity including the use of preclinical imaging (MRI, MicroPET/CT), genomic and lipidomic analysis, and modeling approaches to characterize and quantify adult and developmental neurotoxicity. He has also served on several national/international advisory panels for ILSI, HESI, CIIT, EPA, NIEHS, NAS, NIH and WHO.

Weida Tong, Ph.D.
Director, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Weida Tong is Director of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). He has served science advisory board for several multi-institutional projects in Europe and USA. He also holds adjunct appointment at several universities. In addition, he is the founder and board chairperson of newly established international MAQC Society. His division at FDA is to develop bioinformatic methodologies and standards to support FDA research and regulation and to advance regulatory science and personalized medicine. The most visible projects from his group are (1) conducting the Microarray and Sequencing Quality Control (MAQC/SEQC) consortium to develop standard analysis protocols and quality control metrics for emerging technologies to support regulatory science and precision medicine; (2) development of liver toxicity knowledge base (LTKB) for drug safety; (4) *in silico* drug repositioning for the enhanced treatment of rare diseases; and (4) development of various tools such as ArrayTrack™ suite to support FDA review and research on pharmacogenomics. In addition, his group also specializes in molecular modeling and QSARs with specific interest in estrogen, androgen, and endocrine disruptor. Dr. Tong has published more than 250 papers and book chapters.

AI4TOX – an FDA Artificial Intelligence (AI) program for toxicology

AI4TOX is a program that aims to apply the most advanced AI methods to develop new tools to support FDA regulatory science and strengthen the safety review of FDA-regulated products. The program consists of 4 initiatives:

- **AnimalGAN**: To predict animal toxicology data for untested chemicals through learning models that leverage existing animal data in support of the 3Rs approach (replace, reduce, and/or refine animal studies) in the safety evaluation of FDA-regulated products.
- **SafetAI**: To develop novel deep learning methods for toxicological endpoints that are critical to the safety review of drug candidates before entering clinical trials.
- **BERTox**: To develop the most advanced AI-powered Natural Language Processing (NLP) to facilitate analysis of FDA documents and public literature for improved efficiency and accuracy of information retrieval and toxicity assessment.
- **PathologAI**: To develop an effective and accurate framework for analysis of histopathological data from animal studies to facilitate digital pathology in preclinical application.

The products from these research endeavors will undergo evaluation with a focus on their potential suitability to support regulatory decision-making at FDA.

George Kass, Ph.D.
Chief Scientist Office
European Food Safety Authority (EFSA)
Parma, Italy

Dr. George Kass was trained as a biochemist. He received his PhD in biochemical toxicology from the Karolinska Institute in Stockholm in 1990. After a post-doc at the Swiss Federal Institute of Technology in Zurich he returned to the Karolinska Institute as Assistant Professor. In 1994 he moved to the University of Surrey in the UK where he became Professor of Toxicology. He moved to the European Food Safety Authority in 2009, where is Lead Expert in toxicology. He has published over 140 papers in the field of toxicology and chemical risk assessment.

Data science for food safety: how to integrate new streams of data in the risk assessment process?

Over the past decade, digital technologies and data have transformed the economy and society, affecting all sectors of activity and our daily lives. Not surprisingly, data have been at the heart of EFSA's strategy and activities because of the urgent need to evolve in the way data are collected and reported, and how data models and IT infrastructure are used in the sharing of data between data providers and EFSA. In addition, regulators, including EFSA, are preparing themselves for new data streams from the new approach methodologies (NAMs) and from animal-derived OMICs data through ongoing 3R efforts aimed at changing the way chemical risk assessment is conducted. Furthermore, several EU projects and initiatives are providing a wealth of environmental and human biomonitoring data that should also be integrated in the risk assessment processes linked to food and feed safety. Making data openly available and accessible is a key principle of EFSA, as currently done with open data sets through the Knowledge Junction on Zenodo, EFSA's curated, open repository for the exchange of evidence and supporting materials used in food and feed safety risk assessments. EFSA is also looking into additional novel information streams, crowdsourcing, real-time monitoring systems throughout the food chain, 'Internet of Things', combining standards to improve data exchange capability and much more to ensure EFSA can create a growing pool of large, complex scientific data sets accessible with minimal manual intervention.

Benjamin ER, MSc
Acting Specialist Team Lead (Statistics & Modelling)
Research & Exposure Science Department (RESD)
National Center for Food Science (NCFS)
Singapore Food Agency (SFA)

Mr Benjamin Er is the Acting Specialist Team Lead of the Statistics & Modelling team at SFA. He oversees data science projects undertaken by the team to provide insights on emerging risks to food safety in Singapore. These include looking at the risk factors contributing to gastroenteritis outbreaks locally, analysing trends in global food recall notifications and using text analytics to make sense of consumer feedback and web data for food safety monitoring. Prior to joining SFA, Benjamin worked as a Statistician at the Singapore Ministry of Health where he was involved in estimating the burden of diseases in Singapore, which helped inform policy makers and the public on the state of Singapore's population health. Benjamin holds a Master of Science in Medical Statistics from the London School of Hygiene and Tropical Medicine, and a Bachelor of Science in Computational Biology from the National University of Singapore.

Text Analytics for Food Safety Monitoring

A vast amount of data is collected by SFA daily for the monitoring and assessment of food safety risk from events occurring locally and globally. Many of such data are in free-text format which makes it difficult to quickly make sense of in a systematic way. In this session, we will look at two use cases of textual data that is collected, namely news articles on disease outbreaks scraped from the Internet, and public feedback on food poisoning received via feedback channels, and how text analytics techniques are used to analyse them. We will introduce the techniques used to process and classify these data into sensible categories before critical information is extracted from them. Such initiatives have empowered SFA to make quick assessment of risk from a myriad of information sources, and to anticipate and pre-emptively manage any emerging threats to food safety in Singapore expediently.

Peter Newham, Ph.D.
Vice-President, Safety Sciences,
Clinical Pharmacology and Safety Sciences,
R&D, AstraZeneca
Cambridge, UK

Dr. Peter Newham is the Vice-President of Safety Sciences in Clinical Pharmacology and Safety Sciences, R&D, AstraZeneca. In this role, he is accountable for the non-clinical safety dimension of the entire AstraZeneca R&D portfolio. The Safety Sciences organisation comprises departments in Cambridge UK, Gothenburg Sweden, Gaithersburg US and Waltham US delivering discovery phase "drug –hunting" safety, safety innovation, investigative toxicology, non-clinical regulatory safety, genetic toxicology, reproductive toxicology and safety pharmacology for a broad and diverse range of drug modalities. Key safety innovation programs include development of advanced cell models, micro-physiological systems coupled to quantitative systems pharmacology models to enable translational risk assessment of emerging drugs. Applications of multi-omics, including spatially resolved metabolomics and transcriptomics are being developed to establish higher resolution understanding of drug effects at the cellular and tissue level. With increasing volumes of data being generated, machine-learning approaches are utilized and are being developed to accelerate safety hazard identification and improve design and selection across various drug modalities.

Use of Machine-Learning and Artificial Intelligence for Drug Toxicology

The field of preclinical toxicology has traditionally been considered an essentially descriptive discipline where drug-related effects on biological systems were recorded with limited insight into the molecular mechanisms that drive the observed effects. More recently, the use of deep phenotyping of drug biological effects is permitting the selection of alternative safer compounds, understanding of pre-clinical species relevance and its translatability to human, prediction of safety events, ways to mitigate unavoidable side effects in the clinic, or development of safety biomarkers. Investigative (or mechanistic) toxicology has been gradually gaining momentum and is now a must-have practice in the pharmaceutical industry. The challenges to successful application are driving investigative toxicology to call upon multiple science fields and technologies, from the well-established to barely emerging and to constantly innovate.

In this presentation will present how this discipline is acting on the ever-evolving pharmaceutical portfolio and how various approaches are impacting the drug development paradigm and especially the use of AI/ML approaches. We will highlight recent promising approaches happening in the preclinical arena where utilization of in silico tools has been shown to augment decision-making in the selection of the most promising molecules, even at the stage of initial compound design, across a variety of disciplines. AL/ML tools applied to advances in cell biology that improve DMPK and safety endpoints, will be presented, as well as highlighting approaches being leveraged to influence design of alternative, newer drug modalities. Ultimately, it is hoped that these efforts will allow us to accelerate drug discovery from years to months helping to more successfully identify the most promising medicines for patients at lower costs.

Ruth Roberts, PhD, ATS, FBTS, ERT, FRSB, FRCPath
Director and Cofounder, ApconiX, Alderley Park UK
Chair and Director of Drug Discovery, University of Birmingham, UK

Dr. Ruth Roberts is Chair and Director of Drug Discovery at Birmingham University, UK and is Cofounder of ApconiX, an integrated toxicology and ion channel company. Before that Ruth was Global Head of Regulatory Safety at AstraZeneca (2004-2014) and Director of Toxicology for Aventis in Paris, France (2002-2004). Ruth is current Chair of the HESI board of Trustees, has served on SOT council and is past president of EUROTOX, the British Toxicology Society (BTS) and of the Academy of Toxicological Sciences (ATS). Ruth was the recipient of the SOT Achievement award in 2002, the EUROTOX Bo Holmstedt Award in 2009, the SOT Founders award in 2018 and is the recipient of the 2022 ATS Millie Award, given for outstanding achievement. ApconiX recently received the 2022 Queen's Award for Enterprise. With more than 150 publications in peer-reviewed journals, Ruth is committed to developing and implementing science-led approaches to drug discovery and development.

Data Science in Early Derisking of Drug Targets and Discovery Chemistry

Early derisking of drug targets and chemistry is essential to provide drug projects with the best chance of success. Target safety assessments (TSAs) use target biology, gene and protein expression data, genetic information from humans and animals and competitor compound intelligence to understand the potential safety risks associated with modulating a drug target (1). However, there is a vast amount of information, updated on a daily basis that must be considered for each TSA.

We have developed a data science-based approach that allows acquisition of relevant evidence for an optimal TSA. This is built on expert-led conventional and artificial intelligence-based mining of literature and other bioinformatics databases. Potential safety risks are identified according to an evidence framework, adjusted to the degree of target novelty. Expert knowledge is necessary to interpret the evidence and to take account of the nuances of drug safety, the modality and the intended patient population for each TSA within each project.

Alongside understanding the potential risks associated with inhibiting or activating a drug target, it is key to evaluate the different lead candidates emerging from discovery chemistry to understand their potential for toxicity. This is frequently assessed in early 'Mini Tox' studies in the rodent and in the maximum tolerated dose/dose range finding studies (MTD/DRF) studies carried out prior to selecting one drug candidate to go forward to GLP toxicology testing. However, there is a constant drive to move away from animal testing. We have developed a deep generative adversarial network (GAN)-based framework capable of deriving new animal results from existing animal studies without additional experiments. Using pre-existing rat liver toxicogenomic (TGx) data from the Open Toxicogenomics Project-Genomics-Assisted Toxicity Evaluation System (Open TG-GATES), we generated Tox-GAN transcriptomic profiles with high similarity (0.997 6 0.002 in intensity and 0.740 6 0.082 in fold change) to the corresponding real gene expression profiles, proving its utility in gaining a molecular

understanding of underlying toxicological mechanisms and gene expression-based biomarker development. To the best of our knowledge, the proposed Tox-GAN model is novel in its ability to generate in vivo transcriptomic profiles for different treatment conditions from chemical structures and holds great promise for generating high-quality toxicogenomic profiles without animal experimentation.

Overall, both Tox-GAN and TSAs take full advantage of the most recent developments in data science and can be used within drug projects to identify and mitigate risks, helping with informed decision making and resource management. These approaches should be used in the earliest stages of a drug project to guide decisions such as target selection, discovery chemistry options, in vitro assay choice and end points for investigative in vivo studies.

1. Roberts, RA (2018) Understanding drug targets: there's no such thing as bad news. Drug Discovery Today, 23, 1925-1928. https://doi.org/10.1016/j.drudis.2018.05.028
2. Xi Chen, Ruth Roberts, Weida Tong, Zhichao Liu, Tox-GAN: An Artificial Intelligence Approach Alternative to Animal Studies—A Case Study With Toxicogenomics, *Toxicological Sciences*, Volume 186, Issue 2, April 2022, Pages 242–259, https://doi.org/10.1093/toxsci/kfab157

Timothy W Gant.
Head of the Toxicology Department
UK Health Security Agency
Harwell Science Campus
Oxfordshire, UK.

Professor Tim Gant is Head of the Department of Toxicology at UKHSA and Visiting Professor at Imperial College London School of Public Health. He trained in Toxicology and Pharmacology at the School of Pharmacy, University College London (undergraduate and postgraduate). He is a European Registered Toxicologist, Fellow of the British Society of Toxicology and Member of the American Society of Toxicology. Post PhD he undertook postdoctoral periods with the National Cancer Institute, Bethesda USA (1988 to 1993) and Medical Research Council UK (1993 to 2002). With the Medical Research Council he was appointed to tenure (2002) and group leader (2002 to 2011). He moved to a position within the Health Protection Agency (HPA) in 2011 to pursue interests in the translation of science to public health, research, and training in government. The HPA subsequentially transitioned to Public Health England (PHE) in 2013 and to the UK Health Security Agency (UKHSA) on October 1, 2021. Within UKHSA he chairs the Postgraduate training sub-committee for UKHSA which oversees support for postgraduate training within UKHSA and is co-lead for the Affiliated Research Centre of PHE which is incorporated within the Open University of the UK. He currently directly supervises one PhD student (microplastics). He is Co-Editor of Toxicology Letters the journal of the European Society of Toxicology. He currently serves (chair for 3 years) on the Emerging Issues Committee of the Health and Environmental Sciences Institute, Washington DC, and on the Scientific Advisory Board of ECETOC. For six years chaired the scientific sub-committee of the British Toxicological Society. He co-leads a program grant between Imperial College and UKHSA funded by the National Institutes of Health Research called Environmental Exposures and Health (https://eeh.hpru.nihr.ac.uk/).

Taking omics from concept to application in chemicals regulation

Genomics in various guises has been part of the research toolkit since the late 1990s. The first ‘omics methods consisted of nylon-based dot blots that made use of available cDNA clones. These clones were laboriously collected and sequenced using dideoxy sequencing (Sanger) with gel resolution that could sequence about 400bp in a day or two; when working well. The advent of 96 well capillary sequencing led to an increase in capacity and much larger clone collections that were held in bacterial libraries. Making use of this resource for parallel hybridizations required a new format. Enter the glass microarray the bedrock of all ‘omics. Further technological developments led to increases in microarray capacity to the whole genome. More recently high throughput sequencing, with a capacity that could not have even been envisaged in the days of dideoxy sequencing, has replaced microarrays. In parallel technological developments in Nuclear Magnetic Resonance and Mass Spectral Analysis led to similar advances in protein and metabolite analysis. Right from the start there were predications about how these technologies would revolutionize toxicology allowing greater insight into pathological changes, understanding of mechanisms, and in particular points of departure. There were though issues including recording of metadata, transformation of data and finally interpretation. The recording of metadata was addressed early by agreements with journal editors about the minimum requirements required for publication and submission. Good practice in data generation for microarrays and sequencing has been addressed by the exemplary work of the MicroArray/Sequencing

Quality Control (MAQC/SEQC). Data transformation has proven more of an issue and inhibited the application of ‘omics methods in regulatory toxicology because different outcomes can be achieved from the same data depending on how mathematical transformations are applied to the raw data. These are difficult to critique because of the size of the datasets. In 2014 a CEFIC/LRI project produced a dataset using microarrays known as the Combined Low Dose Exposures to Anti-Androgenic Substances (EMSG56) set. Post laboratory work the issue was how to process that data which led to the realization that there were no internationally accepted standard procedures for data processing or recording. This led to a series of scoping meetings under the auspices of the CEFIC/LRI program that examined methods for data transformation and set a path ahead. This path led to a seminal meeting in 2016 which established a work program to look at both transcriptomics and metabolomics (proteomics was felt to be too complex) and derive frameworks for reporting data and data transformation. Additionally, a standard method for bioinformatic transformation of the data to the point of differential expression and against which other mathematical transformation methods can be compared was derived. This program was presented to the EAGMST group of the OECD in June 2017 and formally included on the workplan in December 2018 led by Health Canada, US-EPA and the University of Birmingham, UK.. To date this has led to draft reporting frameworks (OECD ‘omics reporting frameworks) for transcriptomics and metabolomics and an optimal data analysis framework. Some of output has been published, or is in the process of being published, and the first draft of the OECD Guidance Document was presented to EAMST in June 2022. In this presentation some scientific history of this work will be reviewed, the outcomes of the work to date and what remains to be done.

Maurice Whelan, Ph.D.
Deputy Director, Health, Consumers and Reference Materials
Head of Unit, Chemical Safety and Alternative Methods
European Commission, Joint Research Centre (JRC)
Ispra, Italy

Prof. Maurice Whelan is Deputy Director of the Directorate for Health, Consumers and Reference Materials of the European Commission's Joint Research Centre (JRC) and head of its Chemical Safety and Alternative Methods Unit, based in Ispra, Italy. He also heads the JRC's EU Reference Laboratory for alternatives to animal testing (EURL ECVAM). Maurice is the EU co-chair of the OECD Advisory Group on Molecular Screening and Toxicogenomics that is responsible for the OECD programme on Adverse Outcome Pathways; a member of the Steering Committee of the European Partnership for Alternative Approaches to Animal Testing (EPAA); and chair of the Regulatory Advisory Board of the European Organ-on-Chip Society (EUROoCS). His publications include over 200 scientific papers and a book on the validation of alternative methods for toxicity testing. He has held a number of external appointments including the 2017-2018 Francqui Chair for alternative methods at the Vrije Universiteit Brussel (VUB, Belgium) and is currently visiting Professor of Bioengineering at the University of Liverpool (UK).

Bridging scientific data with evidence needs for regulatory safety assessment

Although regulatory frameworks for food, drug and chemicals vary across the world, they share a common aim of protecting human health and the environment while supporting innovation, competitiveness and trade. To stay relevant and effective however, regulatory processes need to adapt and evolve to respond to several drivers, including emerging risks, changing socio-political priorities, and advances in scientific knowledge and methods. The latter represents significant opportunity for safety assessment practice, particularly in relation to phasing in mechanistic based approaches in toxicology and the state-of-the-art computational, in vitro and 'big data' tools that underpin them. A number of challenges need to be addressed however, including: how to optimally combine and interpret atypical scientific data streams to deliver the information needed for regulatory decision-making; how to strike the right balance between flexibility on one hand, to embrace the rapid evolution of the scientific toolbox, and formality on the other hand, motivated by expectations for validation and standardisation to support implementation; and how to establish the scientific credibility of new approaches to build confidence and gain acceptance. In a broader context too, uptake of new approaches for regulatory applications needs robust trust practices that allow for critical interrogation of scientific methods and data by different actors in the decision-making chain. Understandability is a key requirement in this respect, where the goal is to make it easier to compare elements, see similarities and differences, and to contextualise scientific data in an appropriate framing. In this respect, modern scientific knowledge management practices that are tailored to bridge scientific data with evidence needs in regulation have an important role to play. The Adverse Outcome Pathway framework and related OECD programme is a prominent example.

Professor Susanna-Assunta Sansone, PhD
Professor of Data Readiness, Engineering Science;
Associate Director, Oxford e-Research Centre;
Academic Lead for Research Practice;
University of Oxford, Oxford, UK

Prof. Susanna-Assunta Sansone is a Full Professor in Data Readiness at the University of Oxford, an Associate Director and Principal Investigator at the Oxford e-Research Centre, and the newly appointed University-wide Academic Lead for Research Practice. She is also a Consultant for Springer Nature, and Founding Honorary Academic Editor of *the Scientific Data* journal. An author of the FAIR Principles, she is a recognized as a world thought leader on data sharing and reuse. Susanna leads on the R&D Data Readiness Group, in the Oxford's Department of Engineering Science, and focuses on data sharing, reproducibility and the evolution of scholarly publishing. She has 22 years of experience in these areas. Before joining Oxford in 2010 she was a Team Coordinator at the EBI for 9 years, and previously a Senior Scientist in one of the spin-out companies of Imperial College, where she got her PhD in Molecular Biology. In the Life Science she is the driver behind many flagship resources in the European research data infrastructure, such as FAIRsharing and the FAIR Cookbook (co-developed with pharmas), which are also resources of the ELIXIR Interoperability Platform, where she is an ExCo.

Open, FAIR and reproducible science

Data science enables us to make sense of the massive amounts of data being produced each and every day, impacting on all disciplines and offering an unparalleled economic potential. The reuse of other people's data is providing useful insights for new research questions and products, and driving new scientific discoveries. Realizing its potential will allow for unprecedented advances in scientific discovery and application. However, the growing scale and diversity of the scholarly digital asset (including datasets, algorithms, software and scholarly articles) is placing strain on the mechanisms we currently have for peer review and quality control of the information that is shared.

The FAIR Principles (for Findable, Accessible, Interoperable, and Reusable data) have succeeded to unit stakeholders world-wide behind a common concept: good data management under common standards. FAIR has become a fundamental enabler for digital transformation in the public and in the private sectors. And FAIR it is no longer optional. However, the FAIR Principles are aspirational, and putting FAIR into practice is work in progress; it "takes a village"!

I will provide an overview of two community-driven resources, and how they contribute to the necessary guidelines to turn FAIR into reality. FAIRsharing (https://fairsharing.org), which focuses on promotive the value and use of community standards and terminologies, key elements of FAIR; and the FAIR Cookbook (https://faircookbook.elixir-europe.org), which leverages the expertise of data professional in academia and industry to deliver recipes that address all operational aspects of making and keeping data FAIR. Both resources are part of ELIXIR (https://elixir-europe.org), which unites Europe's leading life science organisations to coordinate, integrate and sustain bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research.

Yuanting Zheng, Ph.D.
Associate Professor
School of Life Sciences
Fudan University
Shanghai, China

Dr. Yuanting Zheng is an associate professor at the School of Life Sciences of Fudan University. She is the executive secretary of the China chapter of the international Massive Analysis and Quality Control (MAQC) Society. Her research focuses on reference materials for multiomics quality control, aiming to translate multiomics technologies into clinical companion diagnosis. Dr. Zheng has been spearheading the Chinese Quartet Project (chinese-quartet.org) for quality control and data integration of multiomics profiling since 2016 by developing, characterizing, and disseminating the reference materials of DNA, RNA, proteins, and metabolites derived from the same immortalized B-lymphoblastoid cell lines of a Quartet family of father, mother, and identical twins. Dr. Zheng's research is funded by several grants from the National Key R&D Project of China, the National Natural Science Foundation of China, and the National High Technology Research and Development Program of China. She has published over 50 peer-reviewed papers in genomics and clinical pharmacology with >2000 citations by SCI journals.

Multiomic reference materials and datasets for improved reproducibility

Multiomic profiling is an unprecedentedly powerful tool to characterize the same samples with complementary features orchestrating the genome, epigenome, transcriptome, proteome, and metabolome. It promises to demonstrate advantages over any single-omics type in identifying clinically important sample subtypes of inherently subtle differences, as well as in interconnecting intricate molecular features of perturbed diseases or actionable treatment targets. However, there is a lack of multiomic reference materials with ground truth to demonstrate wet-lab proficiency and data reliability of multiomic profiling. To this end, we established the first suites of publicly available multiomic reference materials of matched DNA, RNA, proteins, and metabolites derived from a family quartet of parents and monozygotic twin daughters (chinese-quartet.org). We then systematically evaluated the wet-lab proficiency in multiomic data generation and the performance of various data integration methods based on built-in truths in terms of both sample classes defined by family membership and multiomic feature characteristics obeying the central dogma. We demonstrated that, for any given omics type, the “ratio”-based profiling data, i.e., by scaling the commonly used absolute feature values of study samples relative to those of concurrently measured reference sample(s) on a feature-by-feature basis, were highly reproducible across platforms, labs, and batches, thus fundamentally improved the performance in horizontal (within omics) and vertical (cross-omics) data integration. In addition, wet-lab proficiency in data generation, as objectively assessed by the performance metrics of differentiating subtle quartet sample differences, varied substantially and further determined the integration performance, regardless of omics type. Importantly, bioinformatic methods exhibited appreciable differences in performance of integrating various datasets. Our findings urge a paradigm shift from “absolute” to “ratio"-based multiomic profiling for improved reproducibility. Our approach, when widely adopted, can fundamentally advance the integration of diverse multiomic datasets from various scenarios by making them inherently of high quality and reproducibility through improved lab proficiency and ratio-based profiling with universal reference materials.

Leming Shi, Ph.D.
Professor
School of Life Sciences, Human Phenome Institute, and Shanghai Cancer Center
Fudan University
Shanghai, China

Dr. Leming Shi is a professor at the School of Life Sciences, Human Phenome Institute, and Shanghai Cancer Center of Fudan University. Dr. Shi's research aims to improve the success rate of drug discovery and development and to promote precision medicine by generating and integrating high-quality multiomic data. During his tenure at the US FDA, Dr. Shi conceived and led the MicroArray and Sequencing Quality Control (MAQC/SEQC) consortium together with Dr. Weida Tong for quality control and standardization of transcriptomic and genomic data, publishing four special issues in *Nature Biotechnology*. These efforts led to the launch in 2017 of the International MAQC Society (www.maqcsociety.org) to enhance the reproducibility of high-throughput technologies. Dr. Shi served as its first president and is its Chief Science Officer and a member of the board of directors. Dr. Shi was a co-founder of Chipscreen Biosciences and co-developed the chemogenomics-based drug discovery platform, resulting in the marketing approvals of one novel HDAC inhibitor (Chidamide) for treating cancers in China and Japan, and one novel PPARα/γ/δ pan-agonist (Chiglitazar) for treating type 2 diabetes in China. More recently, Dr. Shi and his collaborators developed a multiomics-based molecular subtyping approach for precision treatment of triple-negative breast cancers, significantly improving patient survival. Dr. Shi has published over 200 peer-reviewed papers with >15,000 citations by SCI journals and an *h*-index of 60, and (co-)led the development of four ISO and CLSI standards and guidance on data quality in genomics and transcriptomics.

Two Decades' Effort on Developing Best Practice in Genomics and Its Contribution to Regulatory Science

The field of drug discovery and development has been facing two major challenges. First, the success rate of clinical trials of drug candidates is as low as 10%, which means that although there are many candidates in clinical trials, only few of them can eventually be approved by regulatory agencies for marketing. Secondly, the clinical efficacy of drugs is as low as 50%, which means that for drugs approved for marketing, many patients cannot benefit from using them. In short, good drug candidates, those that can eventually be approved for marketing, fail to stand out in all stages of new drug research and development. And at the same time, "good" patients, those who can benefit from drug treatment, cannot be accurately identified before taking a drug. This situation is not sustainable for the pharmaceutical industry, the patients, and regulatory agencies. The presenter reasoned that the widespread irreproducibility of preclinical research results may be an important factor contributing to the low success rate in clinical trials and low efficacy in drug treatment. There is high expectation on the use of omics technologies to improve the effectiveness of drug discovery and development as well as precision treatment tailored to a patient's omics characteristics. Precision medicine innovation depends on high-quality big data. Only by ensuring the high quality and standardization of data generation, analysis and interpretation can reliable biomarkers of disease prevention, diagnosis and treatment be discovered and validated, thus improving the success rate of new drug development and the clinical efficacy of drug treatment. This presentation will highlight the experiences of the MicroArray and Sequencing Quality Control (MAQC/SEQC) consortium, an international effort involving hundreds of scientists from academia, industry, and government in developing best practice in genomics to help make precision medicine a reality and contribute to regulatory science.

Michael Renaudin
Lead Swissmedic 4.0
Swissmedic, Swiss agency for therapeutic products
Bern, Switzerland

Michael Renaudin studied social anthropology and later sociology at the University of Bern. After working in IT, project management and adult education (in the private and public sectors), he studied human resources. He has worked at Swissmedic for 12 years. He was first responsible for organisational learning and development, and is currently head of Swissmedic 4.0, Swissmedic's innovation lab. The aim is to quickly develop solutions in the field of AI by developing prototypes and MVPs in order to exploit the technological potential of these solutions on the one hand and to promote the digital skills of employees on the other. The holistic approach of Swissmedic 4.0 also includes the dimensions of people and organisation. In his role, he is responsible for leading the digital initiative, as well as for the areas of strategy, innovation and information/communication.

Kwok Yan LAM, Ph.D.
Associate Vice President (Strategy and Partnerships), President's Office
Professor, School of Computer Science and Engineering
National Technological University (NTU)
Singapore

Prof. Lam is the Associate Vice President (Strategy and Partnerships) and Professor in the School of Computer Science and Engineering at the Nanyang Technological University (NTU), Singapore. He is currently also the Executive Director of the Strategic Centre for Research in Privacy-Preserving Technologies and Systems (SCRiPTS), and Director of NTU's SPIRIT Smart Nation Research Centre. From August 2020, Professor Lam is also on part-time secondment to the INTERPOL as a Consultant at Cyber and New Technology Innovation. He served as the Director of the Nanyang Technopreneurship Center 2019-2022, and as Program Chair (Secure Community) of the Graduate College at NTU 2017-2019. Professor Lam has been a Professor of the Tsinghua University, PR China (2002-2010) and a faculty member of the National University of Singapore and the University of London since 1990. He was a visiting scientist at the Isaac Newton Institute of the Cambridge University and a visiting professor at the European Institute for Systems Security. In 2018, Professor Lam founded TAU Express Pte Ltd, an NTU start-up which specializes in AI and Data Analytics technologies for Smart Cities applications. TAU is a spin-off of the Intelligent Case Retrieval System project, a collaboration between NTU and the Singapore Supreme Court. In 1997, he founded PrivyLink International Ltd, a spin-off company of the National University of Singapore, specializing in e-security technologies for homeland security and financial systems. In 2012, he co-founded Soda Pte Ltd which won the Most Innovative Start Up Award at the RSA 2015 Conference. In 1998, he received the Singapore Foundation Award from the Japanese Chamber of Commerce and Industry in recognition of his R&D achievement in Information Security in Singapore. Prof Lam received his B.Sc. (First Class Honours) from the University of London in 1987 and his Ph.D. from the University of Cambridge in 1990. His research interests include Distributed Systems, IoT Security Infrastructure and Cyber-Physical System Security, Distributed Protocols for Blockchain, Biometric Cryptography, Homeland Security and Cybersecurity.

Alexander Horst
Digital Transformator @ Swissmedic.4.0
Swissmedic, Swiss agency for therapeutic products
Bern, Switzerand

Alexander Horst studied business information technology and was ever since interested in leveraging biomedical data as a basis for human decision-making. At Swissmedic, the Swiss agency for therapeutic products, he works within the Swissmedic 4.0 team where he and his colleagues work closely together with both internal and external Stakeholders to find opportunities within the constantly changing digital age. They experiment with innovative approaches, such as involving machine learning or big data with the goal of making the artifacts and knowledge available throughout the organization.

Tricia - Leveraging NLP to enhance risk assessment of incoming incident reports

As part of the market surveillance activities, Swissmedic experts systematically evaluate all incoming incident reports reported by manufacturers of medical devices related to incidents that have occurred in Switzerland. Firstly, the experts perform a risk-based triage of all incoming incidents to help the agency to prioritize its resources. Risk in this context is defined as the severity, the detectability as well as the probability of occurrence for each incident. By training machine-learning models for each subtask and providing the most relevant information for the expert's decision-making process, the current triage-workflow can be simplified. The expert's validation of the produced results helps to build trust in the models among experts (and thus the users of the models). Moreover, false predictions can be identified upon the expert's corrections and therefore corresponding enhancements of the models can be applied.

Nicolas Perez Gonzalez, Ph.D.
Data Scientist, Swissmedic 4.0
Swissmedic
Switzerland

Dr. Nicolas Perez Gonzalez is a Data Scientist at Swissmedic, the Swiss Agency for therapeutic products. After obtaining his Ph.D at The Johns Hopkins University, he moved to Switzerland to work on Clinical Machine Leaning in the University Hospital of Zurich and the University of Zurich where he developed a multimodal model for comprehensive analysis of images and reports in the hospital as well as optimization algorithms among other projects. He currently works with the Swissmedic 4.0 Team as the technical lead in the MediCrawl project at Swissmedic but has additional projects focused on medicament recognition from images, and document similarity. His work focuses in innovation and the implementation of new technologies within Swissmedic. These solutions are mostly oriented towards the improvement of processes at Swissmedic making the stakeholder's engagement of key for the success of these innovative solutions.

MediCrawl

Swissmedic is the National Authorisation and supervisory authority for drugs and medical products. It ensures high-quality, safe and effective medical products. As part of its mandate, Swissmedic regularly oversees online markets on the search for illegal medicinal products. This search is time consuming as well as resource intensive. In order to optimize this workflow, MediCrawl was developed in the innovation team within Swissmedic. MediCrawl is a platform developed in the Cloud that allows a centralized analysis of products of interest found online. These products are collected via crawling and are later presented to the Swissmedic specialists. Due to the large number of incoming products of interest, a series of filters have been developed including a Machine Learning based filter based on the text description of these. With a dataset of 8,000 cases, we trained a classifier based on BERT embeddings, which identifies the binary signal and labels if the case is relevant for Swissmedic. In development, our classifier identified relevant products with a precision of 0.85 and recall of 0.92. 3 months later, with new unseen data, precision dropped to 0.61 while the relevant score of recall remain high at 0.94. Initial results of MediCrawl are promising and the implementation in Swissmedic is underway. Further research will be focused on improving scores with new methodologies as well as the benefits of moving training into a MLOps pipeline.

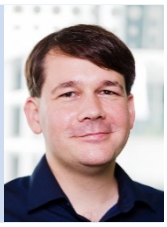
Wenjun Bao, Ph.D.
Chief Scientist and Director of Advanced Analytics R&D
JMP Statistic Discovery
SAS Institute Inc.
Cary, North Carolina, USA

Dr. Wenjun Bao is a Chief Scientist and Director of advanced analytics for JMP statistical Discovery, SAS Institute Inc. Before joining SAS, she was an Intramural Research Training Award (IRTA) Fellow at NIH (National Institutes of Health), a professor at Duke University, and a scientist at the US EPA (Environmental Protection Agency). She has rich experiences in clinical, bioinformatics, biochemistry, and molecular biology research. She has expertise in variety data analysis including AI/ML models in clinical trial and genomics data analysis and text mining with multiple publications in peer-reviewed journals. Dr. Bao has been a research grant review committee member for NIH since 2005 and a research adviser for scientists at universities and government agencies. Dr. Bao is a Board of Director for CDISC and an adjunct professor at Fudan University.

AI Empowers the Assessment for Hepatotoxicity Potential and Clinical Endpoints in Drug Discovery and Development

Artificial intelligence (AI) offers great potential for improving Drug Discovery and development (DDD) at different stages including Discovery, Pre-clinical, Clinical and Post-Clinical testing, along the development progression. However, navigating through AI is a challenge for non-computational scientists. We employ an easy-to-use AI tool requiring no coding skills to generate interpretable results through the deep learning XGBoost predictive model as well as through text mining for genomics, clinical trial, and text data from different stages of DDD. XGBoost can use genomics data from Next-Gen Sequencing to predict drug responses and predict clinical outcomes using integrated clinical trial data from CDISC domains, the global standard for clinical trial data format. Text documents such as drug labels or abstracts of published papers are part of monitoring the post-marketing safety of drug and therapeutic biologic products that are very important to the public health. Warnings and/or Precautions for 678 unique prescription drugs and 12k abstracts were converted to document term matrix (DTM) by text mining. Then Hepatotoxicity or drug-induced injury (DILI) potential was predicted by DTM through XGBoost. The XGBoost not only generates statistical metrics that are better than those from traditional statistical models, but it also ranks the important features from the DTM and creates a prediction formula to predict new drug's DILI potential. The statistical metrics give the sense how good the model results are. The rank of the important features offers insight of the biological meanings. The prediction formula extends the prediction for new data with similar attributers. AI tools such as text mining and XGBoost, used in combination, can enhance analysis and prediction for the different stages of drug discovery and development.

Sebastian Maurer-Stroh, Dr.rer.nat.
Executive Director, Bioinformatics Institute
Agency for Science, Technology & Research
Singapore



Dr. Sebastian Maurer-Stroh studied theoretical biochemistry at the University of Vienna and wrote his master and PhD thesis at the Institute of Molecular Pathology (IMP). After FEBS and Marie Curie fellowships at the VIB-SWITCH lab in Brussels, he has been leading the sequence analytics portfolio in the A*STAR Bioinformatics Institute (BII) since 2007 and Infectious Disease Programme since 2010. He is the Executive Director of BII since January 2021. His computational team is well known for successes at the public-private interface in Singapore from Precision Medicine to Consumer Product and Food Safety through protein allergenicity assessment and of course for his critical contributions to national and global viral pathogen surveillance through the GISAID data science initiative that has become the single most important source for virus outbreak data sharing and analysis in this pandemic powering public health responses globally.

Assessing allergenicity risk of proteins with AllerCatPro 2.0

Proteins in food and personal care products can pose a risk for an immediate immunoglobulin E (IgE)-mediated allergic response. Bioinformatic tools can assist to predict and investigate the allergenic potential of proteins. This presentation will describe AllerCatPro 2.0, a web server that can be used to predict protein allergenicity potential with better accuracy than other computational methods and new features that help assessors making informed decisions. We predict the similarity between input proteins using both their amino acid sequences and predicted 3D structures towards the most comprehensive datasets of reliable proteins associated with allergenicity. These datasets currently include 4979 protein allergens, 162 low allergenic proteins, and 165 autoimmune allergens with manual expert curation from the databases of WHO/International Union of Immunological Societies (IUIS), Comprehensive Protein Allergen Resource (COMPARE), Food Allergy Research and Resource Program (FARRP), UniProtKB and Allergome. Various examples of profilins, autoimmune allergens, low allergenic proteins, very large proteins, and nucleotide input sequences showcase the utility of for predicting protein allergenicity potential. The web server is freely accessible at https://allercatpro.bii.a-star.edu.sg.