PRISME Forum

Pharmaceutical R & D Information Systems Management Executives

PRISME Forum

TECHNICAL MEETING

Opportunities in Translational Science Biopharma R&D including Information Management, Analysis and Visualization

PRISME Forum Chair:

Olivier Gien, *Sanofi*

Technical Meeting Chair:

Martin Leach, *Alexion*

Nov 18 - 19, 2015

Plainsboro, New Jersey, USA

Host: Bristol-Myers Squibb

Download and stay up-to-date with our

PRISME Forum 2015 Technical Meeting App:

http://my.yapp.us/PRISMETECH

Meeting Venue

All sessions will be hosted by Bristol-Myers Squibb at its Plainsboro Campus located at 777 Scudders Mill Road, Plainsboro Township, NJ 08536.

Hotel

The Westin Princeton at Forrestal Village - 201 Village Boulevard, Princeton, New Jersey 08540, United States.

Contacts

Program Coordinator: +44.77.68.173.518/jcmwise@prismeforum.org

Secretariat: office: +1.224.938.9523/logistics@prismeforum.org

On site, SMS: +1.312.622.1234

PRISME Forum Technical Meeting Advisory Committee

Olivier Gien
Chair, PRISME Forum
Global Head, Clinical IT, *Sanofi*

Anastasia Christianson
Head of Translational R&D IT, *Bristol-Myers Squibb*

Lars Greiffenberg
Director, R&D IT and Translational Informatics, *Abbvie*

Martin Leach
VP R&D IT, *Alexion*

Jasmin Saric
Head of IT Translational Science & Epidemiology, *Boehringer-Ingelheim*

Sandor Szalma
Head, Translational Informatics & External Innovation, *Janssen*

John Wise
Programme Coordinator, *PRISME Forum*

PRISME Forum Host

The PRISME Forum Technical Meeting Advisory Committee would like to thank Bristol-Myers Squibb for hosting the fall 2015 meeting.

PRISME Forum Statement of Compliance

"All meetings, working groups and communications will be open to all Members and any records thereof will be non-confidential and available for inspection by any Member. The Members acknowledge that discussing any commercially sensitive topics, including costs, volumes, inventories, sales level methods, channels of distribution, access to future products, markets, current or future prices, profitability, ***contract pricing or trading terms*** is prohibited. The Members of PRISME will strictly comply with all laws relevant to their activities, including US state and federal anti-trust laws and European competition laws."

THEME: Opportunities for Translational Science in Biopharma R&D Including Information Management, Analysis and Visualization

BACKGROUND

Translational Science is a discipline often described as the rapid iteration of learnings from bench to bedside. Those learnings encompasses activities such as disease understanding, target identification and credentialing, biomarker research and patient stratification. Translational Science is seen as one of the key levers in pharma's search for this R&D productivity goal and self-evidently, Translational Science information management, analysis and visualization is a key supporting discipline. This PRISME Forum Technical Meeting will set out the current situation in TS (translational science) information management, analysis and visualization and then explore where improvements are needed and how these improvements might be effected. Furthermore, priorities will be assigned to these improvements bearing in mind scientific need and technical feasibility. Additionally, the meeting will review how the PRISME Forum might stimulate new project ideas within this domain for an IMI-2 or indeed other consortium-based initiatives. A PRISME Forum Technical Meeting Advisory Committee (TMAC) has been set up to provide guidance on agenda topics and appropriately qualified speakers.

FOCUS

The agenda for this PRISME Forum Fall 2015 Technical Meeting will focus on relevant topics including:

- Cases demonstrating how TS (translational science) and its associated technologies have enhanced bio-pharma R&D
- New TS technologies - empowering new insights and services – thereby lowering the barrier to entry to bio-pharma R&D
- Future directions for TS including the exploitation of "big data"
- The integration of TS data across multiple platforms including traditional enterprise systems
- Experiences and examples from suppliers of TS technologies that either directly - or by analogy - will provide opportunities for R&D and R&D IT.
- The need for better access to better ontology for better description of translational science phenomena

MEETING OUTCOMES

PRISME Forum members will:

- Enhance their understanding of translational science and the opportunities for R&D IT to support and develop that discipline
- Take back to their parent organizations tactical learnings that can be deployed quickly to support translational science and
- As a stretch goal identify a common, pre-competitive obstacle that can best be palliated or cured by a cross-company, inter-disciplinary project team working under that umbrella of a grant awarding organization such as the IMI.

PROGRAM

All sessions will be hosted by Bristol-Myers Squibb at its Plainsboro Campus located at 777 Scudders Mill Road, Plainsboro Township, NJ 08536.

WEDNESDAY, November 18, 2015		
18:45	*Gather in hotel lobby for departure to PRISME Forum Group Reception (Business + Technical)*	
THURSDAY, November 19, 2015		
8:00	*Shuttle transfer between hotel and meeting venue*	
8:30	*Check-in; poster installation*	
8:45	**Welcome Notes**	**Olivier Gien,** Chair, *PRISME Forum;* Global Head, Clinical IT, *Sanofi*
8:50	**Introduction**	**Martin Leach,** Technical Meeting Chair, *PRISME Forum*; VP, R&D IT, *Alexion*
9:00	**SESSION I: PERSPECTIVES**	**Chair: Martin Leach,** Technical Meeting Chair, *PRISME Forum*; VP, R&D IT, *Alexion*
9:00	**Translational Science and its Importance in Decision Making**	**Mike Burgess**, Senior Vice President, Exploratory Clinical and Translational Research, *Bristol-Myers Squibb*
9:30	**Converting Available Data into Useful Information in the *Omics* Era**	**George Davey-Smith**, Professor, *University of Bristol*
10:00	**Big Data and Translational Research at Google-Scale**	**Vik Bajaj**, Chief Scientist, *Google Life Sciences*
10:30	*Coffee Break*	
11:00	**SESSION II: LIGHTNING TALKS**	**Chair: Martin Leach** , Technical Meeting Chair, *PRISME Forum*; VP, R&D IT, *Alexion*
11:00	**Genotype-Phenotype Data Driving Medical Discovery and Patient Management**	**Joe Donahue**, Senior Vice President, *BioReference Laboratories*
11:10	**Stratifying Tumor Genomic Alteration Landscapes Through Comprehensive Analytics**	**Eric Neumann**, Vice President of Knowledge Informatics, *Foundation Medicine*
11:20	**Using Machine Learning and Expert Human Guidance to Automate Clinical Data Integration to CDISC Standards**	**Timothy Danford**, Field Engineer, *Tamr*
11:30	***MODERATED PANEL DISCUSSION – Q&A***	
12:00	**SESSION III A: POSTERS**	**Chair: Lars Greiffenberg,** Director, R&D IT and Translational Informatics, *Abbvie*
12:00	Introductions	**Lars Greiffenberg,** Director, R&D IT and Translational Informatics, *Abbvie*
12:05	**Poster Rotations** *(Three 15 minute rotations)*	
P1	**Power of Reverse Engineering and Forward Simulation Platform for Driving Precision Medicine**	**Iya Khalil**, Executive Vice President and Co-Founder, *GNS Healthcare*
P2	**Data Integration Challenges and Opportunities in Translational R&D**	**Albert Wang**, Associate Director, Exploratory Clinical & Translational Research IT, *Bristol-Myers Squibb*
P3	**Translational Science: Current Information Challenges and Solutions**	**Phil Scordis**, Director, Translational Bioinformatics, *UCB*
P4	**Harness Patient Preference from Social Media**	**Mark Wolff**, Consultant, *SAS Institute*
P5	**Translational R&D Analytics: Powering Data-Driven Advances in Immunotherapy**	**Kaushal Desai**, Associate Director, Translational R&D Analytics and Decision-Support, *Bristol-Myers Squibb*
P6	**TranSMART & eTRIKS – Current Status**	**Jay Bergeron**, Director, Translational and Bioinformatics, *Pfizer*
12:50	*Lunch*	

13:30	**SESSION III B: POSTERS**	**Chair: Lars Greiffenberg,** Director, R&D IT and Translational Informatics, *Abbvie*
	Poster Session (Remaining three 15 minute rotations)	
14:15	**SESSION IV: PLENARY PRESENTATIONS**	**Chair: Anastasia Christianson**, Head of Translational R&D IT, *Bristol-Myers Squibb*
14:15	Linked Data at Work in Translational Science	**Hans Constandt**, CEO and Co-Founder, *ontoforce*
15:15	*Coffee Break* *Preparations for* **SESSION VI: Bringing It All Together** *Table Captains: Jay Bergeron, Lars Greiffenberg, Jasmin Saric, and Sandor Szalma will meet to frame the structure of the session. There will be a slide to provide clear instructions for the round-table discussion groups.*	
15:45	**SESSION V: KEYNOTE PRESENTATION**	**Chair: Martin Leach,** Technical Meeting Chair, *PRISME Forum*; VP, R&D IT, *Alexion*
15:45	*A Different Industry Perspective on Multiple Input Data Analytics*	**Tara Grabowsky**, Chief Medical Officer, *Vencore* **Oodaye Shukla**, Chief Data Scientist, *Vencore*
16:15	**SESSION VI: BRINGING IT ALL TOGETHER**	**Chair: Sandor Szalma,** Head, Translational Informatics & External Innovation, *Janssen Research & Development LLC*
16:15	***Objective*** *Create a preliminary draft of a topic proposal for review by the IMI DKM SGG or another relevant project organizing body.* ***Breakouts*** *The audience will be divided in four groups. The table captains listed above will coordinate the group's review, consideration and drafting advice for the next steps (on the basis of the day's proceedings and the instructions displayed on the big screen).*	
17:00	**Plenary: Table Captains' Feedback**	
17:15	**Plenary: Discussion of readout; Determine next steps**	
17:30	**Awards**	**Chair: Martin Leach,** Technical Meeting Chair, *PRISME Forum*; VP R&D IT, *Alexion*
17:45	*Networking Reception*	
19:30	*Return to Hotel*	

BIOS AND ABSTRACTS

PRISME Forum Chair: Olivier Gien

Global Head, Clinical IT, *Sanofi*

Olivier Gien, PhD, was elected as the Chairman of the PRISME Forum at the November 2014 PRISME Forum Business Meeting.

Dr. Gien is the Global Head of Clinical IT at Sanofi. He is a Chemical Engineer by training and holds a PhD in Organic Chemistry. His PhD work focused on leveraging Artificial Intelligence technologies and retrosynthetic analysis to build a system helping chemists in the design of synthetic routes.

Dr. Gien started his career in the Exploratory Unit of Sanofi's Hungarian affiliate in Budapest then took charge of Information Systems for Industrial Chemical development at Sanofi's Sisteron site. He led then Global Discovery Research Information Systems at Sanofi-Synthelabo, then Sanofi-Aventis in Montpellier, before taking on his new role in Paris area in 2010.

PRISME Forum Technical Meeting Chair and Sessions I/II/V Chair: Martin Leach

VP, R&D IT, *Alexion*

Martin Leach, PhD is Vice President IT for R&D at Alexion Pharmaceuticals. Prior to Alexion, Dr. Leach led the R&D IT group and created the first Data Sciences group then subsequently led the Global Data Office at Biogen. Prior to Biogen, Dr. Leach was the CIO of the Broad Institute of MIT & Harvard and before that he led a team of several hundred IT professionals and contractors at Merck & Co. providing information technology and research computing to Discovery & Pre-clinical Sciences at Merck Research Laboratories.

Dr. Leach has worked with several other pharmaceutical, biotechnology and life sciences organizations as part of Booz Allen Hamilton, where for 2 years, he was a leader in their IT strategy practice. Prior to Booz Allen, Dr. Leach had the CIO role, leading the IT and Informatics at CuraGen Corporation and worked with 454 Corporation to create their bioinformatics function. He obtained his PhD in pharmacology from Boston University School of Medicine. He is a member of Silicon Valley advisory groups on technology for Sierra Ventures and Canaan Partners. Recent awards include 2011 CIO Magazine award for "Ones to Watch" and in 2012 he was voted FierceBiotech's Top 10 Biotech Techies.

Session IIIA/B Chair: Lars Greiffenberg

Director, R&D IT and Translational Informatics, *Abbvie*

Lars is the director of R&D, IT and Translational Informatics at AbbVie. Prior to this, he was Global Head Translational Medicine Solution Center at Sanofi R&D-IS. In that role, Lars was responsible for implementing a global core technical platform to support all business units in their execution of individual translational medicine/informatics approaches in Sanofi. For this goal, he had worked closely with leading scientists and bioinformaticians but also with best in class third party software companies.

Prior to this position, Lars served as Sanofi's Global Head Biology Solution center (covering all kind of biological solutions from screening towards genomics applications); Aventis' program manager for large global implementations for in-vitro and phenotypic screening solutions (600+ users) and, project manager for the implementation of global biological screening IT solutions (high throughput screening).

Lars is a molecular biologist with a PhD in Microbiology from the University of Wuerzburg.

Session IV Chair: Anastasia Christianson

Head of Translational R&D IT, *Bristol-Myers Squibb*

Anastasia Christianson is Head of Translational R&D IT at Bristol-Myers Squibb (BMS) where she is responsible for strategy and delivery of all Information and Technology needs for Translational Research and Development at BMS. Prior to BMS, Anastasia spent 20 year at AstraZeneca Pharmaceuticals in various roles across Discovery and Clinical Development ranging from leading drug projects to establishing Genomics and Biomedical Informatics, and leading IT organizations and global programs across Discovery and Development.

Outside BMS, Anastasia is adjunct professor at the University of Delaware's (UD) Center for Bioinformatics & Computational Biology (CBCB), contributed to the development of Bioinformatics and Biomedical Informatics Master's degree and certification programs at UD, and chaired the external advisory committee that established a Quantitative Biology degree program (also at UD). Anastasia is a member of PhRMA's (Pharmaceutical Research and Manufacturers of America) Translational Medicine Advisory Committee and sits on the editorial board of Inderscience's International Journal of Data Mining and Bioinformatics.

Anastasia received her Ph.D. in Biological Chemistry from the University of Pennsylvania followed by postdoctoral training in Cellular and Developmental Biology at Harvard University. Her passion is leveraging data for evidence-based decision making.

Session VI Chair: Sándor Szalma

Head, Translational Informatics and External Innovation, R&D IT, *Janssen R&D LLC*

Sándor Szalma is head of Translational Informatics and External Innovation, R&D IT in Janssen Research & Development, LLC. He is responsible for implementing the translational informatics capability and establishing and fostering large scale open innovation collaborations to support discovery, translational, clinical and safety science including public-private partnerships and various pre-competitive collaborations such as Innovative Medicines Initiative consortia. His current interests span researching, developing and applying bio- and medical informatics methodologies for enhancing drug discovery and development, biomarkers and translational medicine and utilizing emerging technologies such as social and mobile computing and remote sensors for discovering digital biomarkers, improving clinical development, pharmacovigilance and healthcare. Sándor is member of the board of the Pistoia Alliance, member of the Translational Medicine Advisory Committee of the PhRMA Foundation and co-leading the Data & Knowledge Management Strategic Governance Group of EFPIA.

Sándor's past positions included president of MeTa Informatics, general manager of QuantumBio and senior director of Computational Biology and Bioinformatics at Accelrys, Inc. He was co-founder of Acheuron Pharmaceuticals, Inc. He lectured at UCSD Extension and now is adjunct professor at Rutgers University in the Computational Biology and Molecular Biophysics program. He is the author of more than 40 scientific publications and book chapters and two patents. He received his doctoral degree in chemistry from A. Szent-Györgyi Medical University in Szeged, Hungary.

SESSION I
PERSPECTIVES

Session Chair: Martin Leach

VP, R&D IT, *Alexion*

Mike Burgess

Senior VP, Exploratory Clinical and Translational Research, *Bristol-Myers Squibb*

As Head of Exploratory Clinical and Translational Research (ECTR), Mike is responsible for the strategy and execution of all early development activities (Phases I and II), clinical biomarkers, clinical pharmacology and translational medicine across all global therapeutic areas (Cardiovascular, Immunology and Fibrosis, Genetically Defined Disease, Immuno-oncology Oncology, Virology). Mike ensures that the ECTR organization achieves industry-leading efficiency in progressing the pipeline, and effectiveness in transitioning assets through Proof of Concept. He leads the ECTR Clinical Biomarkers strategy, develops the clinical pharmacology profiling plan for all therapeutic areas, and plays a key leadership role as we continue to build the company's Translational R&D capabilities. Under Mike's leadership, we will enhance the company's Full Development Portfolio with Phase III compounds, build disease mechanism of action understanding and earlier evaluation of pharmacodiagnostic possibilities and ensure our long-term sustainability.

Mike joined BMS in early 2013 from Hoffman La-Roche where he was most recently interim head of the company's Pharma Research and Early Development organization. Prior to that he led the Hoffman La-Roche Oncology Discovery and Translational Area, one of five disease focused organizations. At the same time he served as Head of Global Large Molecule Research which is responsible for all antibody and protein engineering from the discovery of molecules through to the definition of the GMP process for first in man studies. In these positions Mike had responsibility for defining and executing strategy, ensuring scientific and translational excellence across sites globally and progressing large and small molecules into Development.

Mike is a physician scientist with a Ph. in molecular biology and has 10 years specialist clinical experience in pediatrics and pediatric oncology. Mike has successfully led the global development and registration of a wide variety of anticancer agents and established multiple collaborations across industry and academia to design and implement global translation medicine strategies.

Mike received his medical training and PhD from the University of Bristol, U.K., and is a member of the Royal College of Physicians and the Royal College of Pediatrics and Child Health.

Translational Science: the Importance in Biopharma R&D

Translational science or medicine in its simplest form can be thought of as a series of decisions which are essential for the transition of a molecule to a medicine and for the dialogue between physician and patient. The importance of these decisions is borne out from the outset of development under the "quick kill" paradigm. Under this paradigm molecules are terminated at the earliest opportunity thus preserving resources for those molecules which have a higher probability of making it through the next stage of development. The ultimate product of this process is the decision to ensure that right patient receives the right drug at the right dose for their disease. This ability to terminate early has been identified as the single most important factor governing the success of a pharma company.

Like all decisions there is a high dependence on quality data. This data is often complex in structure and diverse in source. The importance of systems and people with the appropriate skills to proactively interrogate this data and drive hypotheses is fundamental to translational medicine. There is now a very clear relationship between translational medicine and the success of a drug under development. This has been described as part of the framework described by Pfizer's 3 pillars. These concepts will be discussed in greater detail.

George Davey-Smith

Professor, *University of Bristol*

George Davey Smith is professor of clinical epidemiology at the University of Bristol, honorary professor of public health at the University of Glasgow and visiting professor at the London School of Hygiene and Tropical Medicine. He is Scientific Director of the Avon Longitudinal Study of Parents and Children (ALSPAC) and Director of the MRC Integrative Epidemiology Unit (IEU). His major research interest relate to the use of genetic epidemiology for informing understanding of the causal influence of environmentally modifiable risk factors and how social inequalities in health are generated by exposures acting over the entire life course. George has also worked on HIV\AIDS prevention in Nicaragua and India and on issues around the history of epidemiology, meta-analysis, lay epidemiology and epidemiological methodology. He is coeditor of the International Journal of Epidemiology.

Converting Available Data into Useful Information in the Omics Era

A vast amount of publicly-accessible data are becoming available, from genome-wide association study findings through large-scale projects such as ENCODE, the NIH Epigenetics roadmap, and GTEx. Mendelian randomization is one approach through which these data can be combined to produce evidence on the potential effects of pharmacotherapuetic manipulation of particular factors. Mendelian randomization is an increasingly widely used approach utilizing genetic variants as causal anchors in the analysis of observational data. The approach circumvents many of the problems of conventional observational data analysis, including confounding, a variety of biases, and reverse causation. The approach can be applied to many stages of therapeutic target validation, including establishing the causal effects of a pharmacologically modifiable biomarker, predicting the full phenotypic range of on-and-off-target pharmacotherapeutic effects, (including long-term effects) and prioritizing potential agents for formal randomized controlled trial evaluation. Applications of such approaches to optimal utilization of available data will be briefly presented.

Vik Bajaj

Chief Scientist, *Google Life Sciences*

Dr. Vikram (Vik) Bajaj is the Chief Scientist of Google Life Sciences, where he manages its translational and basic life sciences research programs. He is also an Associate Professor (consulting) at the Stanford School of Medicine, an Affiliate Scientist of the Lawrence Berkeley National Laboratory and the University of California, Berkeley, for which he serves on the advisory board of the College of Chemistry. His broad research interests lie at the interface of the physical sciences, engineering, and the life sciences, including experimental and computational systems biology, nanotechnology, and spectroscopy for materials science and molecular imaging and other applications. He was previously a principal investigator at UC Berkeley/LBNL.

Big Data and Translational Research at Google-Scale

SESSION II
LIGHTNING TALKS

Session Chair: Martin Leach

VP, R&D IT, *Alexion*

Joe Donahue

Senior Vice President, *BioReference Laboratories*

Joe Donahue is currently SVP Commercial Collaborations and Innovation at BioReference Laboratories, a global leader in diagnostic and genetic testing, and holds board seats with Torreya Insights, a transactions enabling company to the life sciences sector; GenoSpace, a spin-off of Dana-Farber Cancer Institute developing an information ecosystem to support genomic medicine; and GenePraxis, a healthcare IT company making genomic and genetic information accessible and actionable to clinicians and patients.

Prior to joining BioReference, Joe was a SVP at Thomson Reuters where he was responsible for global life sciences sales, partnerships and 'big data' software development. Joe's life sciences experience also includes executive level positions at MDL Information Systems, Spotfire, LION bioscience AG and InforSense. He holds degrees in chemistry and computer science from Villanova University in Villanova, Pennsylvania.

Genotype-Phenotype Data Driving Medical Discovery and Patient Management

There are few researchers from pharmaceutical companies or academic research organizations that would challenge that drug discovery has become – more than ever before - a data and informatics driven effort. In the 1970's and 80's combinatorial chemistry gave us huge libraries of molecules, which required High Throughput Screening technologies and helped us amass huge databases of compound-target interactions. Now, with the Human Genome sequenced with increasing regularity, the amount of omics data makes the efforts of the 1980s seem small in comparison.

Most recently, we have embarked on a mission to accumulate as much patient data as possible to drive our understanding of disease and accelerate the drug discovery and approval process. We're buying datasets that we don't quite know what to do with yet, whether the data are obtained by spitting into a tube that we mail away, via social media posts or by investigating whether we are descendants of the Vikings. Unfortunately, not all data has equal value for the task at hand. In a clinical setting, it's not the quantity of data but the accuracy and clinical association that is of paramount importance.

In this talk, we'll discuss how patient data, medical and clinical experts and analytics are coming together to make game changing differences in disease understanding and medical research.

Eric Neumann

Vice President of Knowledge Informatics, *Foundation Medicine*

Dr. Eric Neumann is Vice President of Knowledge Discovery and Technology Innovations at Foundation Medicine, and is responsible for organizing FMI's knowledge assets and supporting knowledge discovery initiatives. He is a recognized world expert in semantic information and data modeling for the life sciences, and has worked on many information initiatives for the pharmaceutical and life sciences, including the W3C Semantic Web Healthcare and Life Science Interest Group (HCLSIG), BioPAX, and OMG Life Sciences.

Eric has co-founded several companies, including Selventa, Clinical Semantics, and Pangenx. He has also been a consultant for several pharmaceutical companies to enhance their management of scientific and pharmacological information. Eric has also been the Global Head of Knowledge Management for Scientific and Medical Affairs within Sanofi-Aventis, overseeing the reuse of knowledge and insights across a multinational company. Prior to that, he was VP of Informatics at Beyond Genomics, and systems biology company.

Eric is a graduate from MIT and holds a PhD in neurobiology, developmental genetics, and pharmacology from Case Western Reserve University.

Stratifying Tumor Genomic Alteration Landscapes through Comprehensive Analytics

Many different tumors have been heavily studied and sequenced, leading to scores of findings about altered genes. This explosion in knowledge has not been matched yet with sufficient clinical success, but the role genetic drivers is growing steadily.

We describe a novel molecular classification system that captures the combinatorial nature of relationships between alterations in these diseases. We use this classification to mine for enrichment of variants of unknown significance, and demonstrate a method for segregating unknown variants with functional importance from passengers and SNPs.

Timothy Danford

Software Engineer, *Tamr*

Timothy Danford is a computer scientist working on advanced automation approaches to Big Data Variety in the pharmaceutical and healthcare industries. Previously, he was a software architect, engineer and founding team member for Genome Bridge LLC, a Broad Institute subsidiary organized to develop cloud-based SaaS genomic analysis pipelines.

He has experience in developing data management services, applications and ontologies for bioinformatics and genomics systems at Novartis and Massachusetts General Hospital. As a PhD Student in Computer Science at MIT CSAIL, he focused on computational functional genomics. He is a contributor to ADAM, an open-source project for bioinformatics on Spark.

Using Machine Learning and Expert Human Guidance to Automate Clinical Data Integration to CDISC Standards

Despite a large investment in data standards both at the industry level (CDISC) and within clinical trial sponsor organizations, clinical data integration is still an extremely manual and costly effort. In addition to FDA submission requirements, many study sponsors have a need to integrate data across trials and programs, which often proves even more challenging. Most efforts to provide some level of automation depend on locking down source data formats, which is a challenge as new types of source datasets, such as biomarker data, are being generated.

This talk will present a novel approach to the mapping and transformation of clinical data to standards such as SDTM, which applies the combination of machine learning with expert human guidance to automate up to 90% of the work. Machine learning algorithms are used to map source datasets in different formats from systems such as EDC to the SDTM schema, first working at the variable level and then at the value level. Where mappings cannot be determined by the machine, the organization's own domain experts are leveraged to provide input, which further trains the model, allowing automation to improve over time. The end result is a system that provides the speed and scalability of an automated system with the required accuracy of human curation.

This talk will describe this hybrid approach which was developed working with researchers from MIT and leading healthcare companies - and discuss some early results in clinical data integration to CDISC standards.

A sample case study will show how the machine-driven, human-guided approach could vastly reduce the time and complexity of converting clinical data to CDISC standards for FDA submission, while improving the consistency and availability of integrated data for downstream analysis and use. We will offer observations of how this approach can also help start (or resurrect) clinical data warehousing initiatives, work currently in process at Tamr.

SESSION IIIA/B
POSTERS and DEMONSTRATIONS

Session Chair: Lars Greiffenberg

Director R&D IT and Translational Informatics, *Abbvie*

Iya Khalil

Executive Vice President and Co-Founder, GNS Healthcare

Dr. Khalil is a technology entrepreneur and physicist with a vision of transforming medicine into a discipline that is quantitative, predictive, and patient-centric via big data analytic approaches. She co-founded two big data companies, Via Science and GNS Healthcare, and is the co-inventor of the proprietary computational engine that underpins both entities.

She trained in theoretical physics at Cornell University, and has more than 11 years of experience in "big data" analytics for healthcare, medicine, and the life sciences. She has led several key foundational collaborations with providers, pharmaceutical companies, foundations, and government agencies. Dr. Khalil's expertise spans applications in drug discovery, drug development all the way to treatment algorithms that can be applied at the point of care. She is a frequent speaker at industry events and conferences, has appeared in several industry journals, published several articles in the field, and was recognized by President Obama at a White House dinner as a leading entrepreneur in genomic medicine.

More recently, she was named to the PharmaVOICE 100 list of the most inspiring people in the life-sciences industry. She was recognized for her ability to build bridges across the life-science and healthcare industries, bringing people together to harness the power of predictive modeling to change the lives of patients.

P1: Power of Reverse Engineering and Forward Simulation Platform for Driving Precision Medicine

Every day, vast amounts of healthcare data are collected from real world medical visits on patient treatment regimens and subsequent outcomes, clinical trials, as well as in vitro research. This big data raw material provides a rich asset to investigate for understanding therapeutic effectiveness and patient care. Datasets range from genomics and other omic data to clinical to medical and pharmacy claims to electronic medical records to registries and beyond, and analyses can discover and predict optimal drug combinations, biomarkers, drivers of disease, novel interventions, mechanisms of action, disease models, portfolio optimization, and personalized care/treatment algorithms.

Key to leveraging this data and uncovering which treatments and interventions specifically improve a patient outcomes are powerful analytic approaches. Utilizing causal mathematics and machine learning to create in silico disease networks directly from data has been a successful approach to identify predictive and causal mechanistic associations. Simulations of resultant models unlock the knowledge within complex data, enabling personalized, actionable predictions and precision targeting of interventions. We have leveraged this framework and encoded in a software platform (REFS™) that has enabled applications from target discovery, to patient stratification to the ability to predict drug-drug combination effects from single drug treatments.

Successful case studies demonstrating success of combining rich data assets with causal modeling and machine learning will be presented. These successes demonstrate the power of precision medicine to identify and predict patient outcomes will significantly increase the ability of healthcare leaders and professionals to make better decisions to improve patient care.

Albert Wang

Associate Director, Exploratory Clinical & Translational Research IT, *Bristol-Myers Squibb*

Albert (Al) Wang is an Associate Director in the Translational R&D IT group at Bristol-Myers Squibb, focusing on Data & Systems Integration. Al has been at Bristol-Myers Squibb for 15 years, designing and building technology solutions for a variety of R&D functions, including genomics, proteomics, biomarker discovery, clinical pharmacology & pharmacometrics, clinical biorepositories, and medical affairs. He has a Bachelor's degree in Bioengineering and a Master's Degree in Computational Biology, both from the University of Pennsylvania.

P2: Data Integration Challenges and Opportunities in Translational R&D

Data integration continues to be a significant impediment to translational research and development. In particular, the ability to flexibly combine diverse biomarker data with relevant patient-level information to produce analysis-ready data sets is a challenge that has not been fully solved by existing tools and approaches. This talk will describe the use cases around biomarker data in drug research & development, as well as an ongoing project to implement a platform that assists in these use cases.

Phil Scordis

Director, Translational Bioinformatics, *UCB*

With an early start in Biochemistry and Computer Science provided by degrees from Oxford and UCL, Bioinformatics has long been a passion that carried Dr. Phil Scordis through a PhD with UCL and Postdoc work in Manchester, and into BioPharma life in UCB Celltech, as a Bioinformatician and then lead for Bioinformatics. Phil Scordis took on the responsibility of leadership for Computational Research and diversification occurred quickly; which, while strongly rooted in Computational Biology, has shifted his focus from computational structural/systems biology into the semantic web, knowledge exploitation and enterprise social media domains.

In recent years the focus has broadened out into areas of Data Science and leadership, including a 2 year interim role leading the Global Discovery Research Informatics division and today a focus on building Translational Bioinformatics.

P3: Translational Science: Current Information Challenges and Solutions

The premise of Translational Research relies on building bridges between clinical concepts & data and basic research data & biological concepts (and vice versa) in order to bring insights into the complexity of human disease / treatment opportunities.

As this is inherently reliant on bringing data from multiple domains together to address these challenges, then how much better do we need to get at data integration?

Exemplified through a set of internal use cases we highlight some of the challenges of enabling translational research activities, and the translational informatician at the heart of it. The emphasis is on exposing some of the internal approaches but we bring into this the context of a workshop recently hosted by the EBI in which some of the observations were magnified.

The poster aims to highlight some of the opportunities of Translational Research and provoke discussion around the challenges of bringing technologies to play in the context of a rapidly evolving ecosystem of data, analytical tools and infrastructure.

Mark Wolff

Consultant, *SAS Institute*

Dr. Wolff has over 20 years of experience in the health and life science industries as a scientist and analyst working in the U.S. and Europe. Mark joined SAS in 2005 and is a Principal Industry Consultant within in the Health and Life Sciences Global Practice. Mark's areas of expertise at SAS include the development and application of advanced and predictive analytics in the life sciences and healthcare with a particular interest in patient safety and outcomes.

Currently his work focuses on methods and application of unstructured data and text analytics in support of safety/outcomes research, development of intelligent, decision support systems and data visualization. Mark holds a Bachelor of Science degree in Biology from Loyola College in Maryland and a Master of Science and Doctorate in Toxicology from North Carolina State University.

P4: Harness Patient Preference from Social Media

Massive amounts of biomedical data, from the "omic" disciplines to hospital beds, are being generated everyday. This "embarrassment of riches" shows no sign of moderation as biotelemetry from wearable devices will soon add to the present burden of volume and heterogeneity. Utility of these data in support of the discovery and development of novel therapeutics and their translation to the bed side will only grow in complexity. With the ever increasing ratio of computational power to cost and rapid developments in visualization and automated data analysis, there exists an unprecedented opportunity to apply innovative approaches to the wealth of previously analytically unavailable data.

Kaushal Desai

Associate Director, Translational R&D Analytics and Decision-Support, *Bristol-Myers Squibb*

Kaushal Desai is an Associate Director in R&D IT at Bristol-Myers Squibb (BMS). As an analytics lead in Translational R&D IT, Kaushal is responsible for leading the delivery of data-driven insights to support decision-making in early clinical and translational research.

Prior to BMS, Kaushal was a strategy lead at the Design and Interpretation Center of Excellence in Global Medicines Development at AstraZeneca. During his 14 years of research informatics experience in the pharma industry, Kaushal has led analytics and IT solution delivery initiatives across late discovery and clinical development.

Kaushal received his PhD in Information studies and dual master's degrees in Biomedical Engineering and Information systems from Drexel University.

P5: Translational R&D Analytics: Powering Data-Driven Advances in Immuno-therapy

Immunotherapy has emerged as one of the most promising therapeutic mechanisms with significant clinical benefit in multiple disease areas. Some of the very first immunotherapy trials have led to several interesting scientific questions and 'in silico' testable hypothesis that make secondary exploratory analysis a strategic imperative for pharmaceutical R&D IT groups.

This poster will outline recent work aimed at delivering data-driven insights to support immunotherapy research at Bristol-Myers Squibb. Specific case studies will show case fast-growing business demand for analytics in the exploratory clinical space and demonstrate how data-driven research is enabling the delivery of foundational science to enrich translational R&D programs.

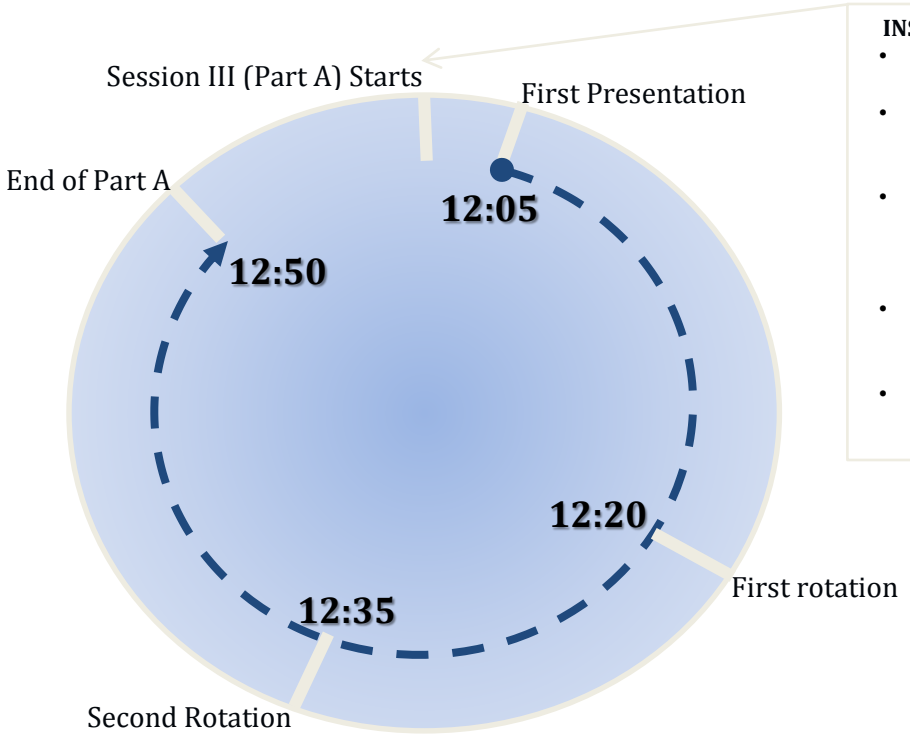
Jay Bergeron

Director, Translational and Bioinformatics, *Pfizer*

Jay Bergeron is Director, Translational and Bioinformatics, at Pfizer and responsible for informatics support for Clinical Research and Precision Medicine. Jay serves as the Scientific Coordinator of eTRIKS (European Translational Research and Information Knowledge Management Services), a public private partnership to support translational data management for projects sponsored by the Innovative Medicines Initiative as well as other Eu-based research consortia. He also co-leads the code committee of the tranSMART Foundation, a non-profit community organizer for the Open Source "tranSMART" translational research data warehouse, a system used by eTRIKS, Pfizer and many additional academic, non-profit and commercial organizations.

P6: TranSMART & eTRIKS – Current Status

SESSION STRUCTURE



INSTRUCTIONS @12:00pm

- Session Chair (Lars Greiffenberg) will invite participants to take their seats.
- Chair will open the session by introducing the presenters, their posters, along with the session structure and flow.
- PRISME Forum staff will ring the bell on each rotation (minute 13 of each presentation and then minute 15 at which time the presentation must end).
- Delegates are invited to identify the color of their badge and match to rotation called out loud by staff: "Rotation 1, 2, …n".
- Rotations will involve shift of participants' groups to the next poster on their right.

SESSION III A

ROTATION 1 - 12:05	ROTATION 2 - 12:20	ROTATION 3 - 12:35
P1 – Orange	P1 – Yellow	P1 – Red
P2 – White	P2 – Orange	P2 – Yellow
P3 – Turquoise	P3 – White	P3 – Orange
P4 – Purple	P4 – Turquoise	P4 – White
P5 – Red	P5 – Purple	P5 – Turquoise
P6 – Yellow	P6 – Red	P6 – Purple

BREAK FOR LUNCH 12:50

SESSION III B

ROTATION 4 - 13:30	ROTATION 5 - 13:45	ROTATION 6 - 14:00
P1 – Purple	P1 – Turquoise	P1 - White
P2 – Red	P2 – Purple	P2 – Turquoise
P3 – Yellow	P3 – Red	P3 – Purple
P4 – Orange	P4 – Yellow	P4 – Red
P5 – White	P5 – Orange	P5 – Yellow
P6 – Turquoise	P6 – White	P6 – Orange

POSTER ROTATIONS *(badge colors)*

ORANGE	
Google Life Sciences	Vik Bajaj
Technologie Servier	Beatrice Chapuzet
Tamr	Timothy Danford
Pfizer	Hall Gregg
H. Lundbeck	Preben Klavsen
Takeda	Tomo Matsunaga
Daiichi-Sankyo	Jim McGurk
Vencore	Oodaye Shukla
WHITE	
Bristol-Myers Squibb	Alastair Binnie
Daiichi-Sankyo	Clive Bowman
ontoforce	Hans Constandt
Takeda	Dave Ficenec
Sanofi	Olivier Gien
Vencore	Tara Grabowski
Boehringer-Ingelheim	Jasmine Saric
Ipsen	Jianying Shi
TURQUOISE	
BioReference Laboratories	Joe Donahue
F. Hoffmann-La Roche	Martin Erkens
Otsuka	George Goldsmith
AbbVie	Lars Greiffenberg
Tamr	Michael Hutnyan
H. Lundbeck	Alex Schuleit
Janssen R&D	Sandor Szalma
GNS Healthcare	Stacey Wasserman
PURPLE	
Bristol-Myers Squibb	Anastasia Christianson
University of Bristol	George Davey-Smith
Biogen	Matteo di Tommaso
Moderna	Steven Frederick
Bayer Pharma	Andreas Friese
Amgen	Arun Nayar
Praxis	Dominique Neaud
Otsuka	Ashok Upadhyay
RED	
Celgene	John Apathy
Bristol-Myers Squibb	Mike Burgess
Purdue Pharma	James Cronin
Alexion	Sebastien Lefebvre
Novo-Nordisk	Thomas Løngborn-Jensen
Allergan	Nancy Maher
Boehringer-Ingelheim	Scott Oloff
AbbVie	Alfred Stefan
The Hyve	Kees van Bochove
YELLOW	
UCB	Dan Chapman
Celegene	Klaus Hofenbietzer
Shire	Fabien Jolly
Alexion	Martin Leach
Foundation Medicine	Eric Neumann
Takeda	David Sedlock
Novo-Nordisk	Jens Noack Skærbæck
Pfizer	Susie Stephens

SESSION IV
PLENARY PRESENTATIONS

Session Chair: Anastasia Christianson

Head of Translational R&D IT, *Bristol-Myers Squibb*

Hans Constandt

CEO and Co-Founder, *Ontoforce*

Hans Constandt, CEO of ONTOFORCE, has a background in medicine, biotechnology, business modeling, bioinformatics, portfolio management, knowledge management, global teams and business liaison in the pharmaceutical industry. His expertise in corporate, academic and startups in the domain of drug discovery, translational science, ICT, data science, semantic web, sustainable business models and fundraising serves to bring innovative solutions to reality in healthcare and life sciences. Hans is a strong believer that the smart web technologies can help patients live longer and better lives.

Linked Data at Work in Translational Science

A lot of companies are facing the challenge of linking and characterizing biomedical data of human studies, biological specimens, cell lines, microarrays and other research data to enhance resource sharing which ultimately accelerates experimentation and encourages interdisciplinary collaboration.

DISQOVER uses an ontology-centric architecture to organize and visualize internal data in conjunction with third party and public data. It builds semantically linked data repositories allowing users to 'hop' from one data type to another.

The core search engine has been optimized and is now more powerful, more scalable and faster compared to traditional semantic web architectures. Users can search in different data types, follow links between data concepts while the data is being harnessed from many different heterogeneous data sources in a federated way. The intuitive visualization engine allows everybody to easily build complex semantic queries across healthcare and life sciences data from many different data sources. Results are stored in a reusable information network and can be visualized using lists, tables, tree-views, world maps, timelines, etc. depending on the data type or property.

At the end of the executed search, all search steps can be stored, exported, re-used or shared in multiple formats resulting in true collaboration across harmonized data fueling translational science. In the talk we will zoom in on the architecture and execute a demo delivering an answer to a real life scientific question.

SESSION V
KEYNOTE LECTURE

Session Chair: Martin Leach

VP, R&D IT, *Alexion*

Tara Grabowsky

Chief Medical Officer, *Vencore*

Tara Grabowsky, MD is the Chief Medical Officer for Vencore Health. Since attending Stanford Medical School and completing her residency at Harvard's Brigham and Women's Hospital, she has worked as an internist in academic, underserved, and private practice settings. Now, she is utilizing this broad base of experience to lead the team's medical efforts. She has won numerous teaching awards for her instruction of Harvard medical students, interns, and residents. She worked previously for Johnson & Johnson in its regulatory affairs department. She graduated with honors from Dartmouth College.

Oodaye Shukla

Chief Data Scientist, *Vencore*

Oodaye Shukla is the Chief Data Scientist at Vencore health care. He has worked in the Telecom, Medical, and the DoD and the IC industries for over 20 years. He started his career at the Johns Hopkins Applied Physics Lab working on optical computing, optical signal processing, and optical fiber sensors. His other accomplishments include developing neural network models to assess cataract severity and determining the correlation between Choroid uniformity and Glaucoma. Along the way he worked in the telecom industry at Lucent Technologies and at Siemens Telecom working on semiconductor lasers and receivers, 1.6 TB/s optical long-haul systems, and having responsibility for a $20 Million product portfolio. He joined Lockheed Martin in 2004 managing the development of a large scale satellite communications simulation model. During Vencore's divestiture from Lockheed Martin and has been the Advanced Communications Systems Manager and more recently the Predictive Analytics & Big Data Manager. His educational background is: BSEE from The Pennsylvania State University, MSEE (Optical Signal Processing) from The John Hopkins University, MSEE (Neural Networks) University of Maryland, and an MS in Technology Management from University of Pennsylvania/The Wharton School.

A Different Industry Perspective on Multiple Input Data Analytics

Vencore has been a key player in the intelligence community for over 40 years. We work closely with governmental organizations responsible for countering terrorism, providing law enforcement and ensuring the integrity of critical national infrastructure. Vencore is currently responsible for:

- Leading the integration of 22 agencies into a single DHS intranet
- Providing software systems to Customs and Border Patrol agents for border management
- Deploying mobile and fixed communication systems, Ensuring cyber security of DHS.Gov
- Delivering deployed asset visibility to FEMA

Vencore focuses on highly complex, interconnected programs across the areas of counterterrorism, border security, emergency response and cyber security. Vencore provides solutions by applying a deep data analytics methodology that goes beyond traditional statistics or data mining. Our approach combines advanced analytical techniques and machine learning algorithms to manage large amounts of data, uncover patterns and trends, and extract maximum value and intelligence.

In our keynote address, we will discuss how we are applying the skills developed in the defense industry to the healthcare sector. We will use a warfighter operations vignette to demonstrate how we are translating the intelligence analysis process to the biotech sector. During this discussion, we will show how we are (1) finding patients with rare disease before they are diagnosed, (2) evaluating patient journey and performing demographic analysis; and (3) redefining previously-accepted epidemiology.

SESSION VI
BRINGING IT ALL TOGETHER

Session Chair: Sandor Szalma

Head, Translational Informatics & External Innovation, *Janssen Research & Development LLC*

Objective

Create a preliminary draft of a topic proposal for review by the IMI DKM SGG or another relevant project organizing body.

Breakouts

The audience will be divided in four groups. The table captains listed above will coordinate the groups' review, consideration and drafting advice for the next steps (on the basis of the day's proceedings and the instructions displayed on the big screen).

ROUND TABLE DISCUSSION STRUCTURE

GROUP A

John Apathy	Celegene
Vik Bajaj	Google Life Sciences
Mike Burgess	Bristol-Myer Squibb
Timothy Danford	Tamr
Andreas Friese	Bayer
Tara Grabowski	Vencore
Martin Leach	Alexion
Nancy Maher	Allergan
Dominique Neaud	Praxis
Scott Oloff	Boehringer-Ingelheim
Alfred Stefan	AbbVie
Susie Stephens	Pfizer
Mark Wolff	SAS Institute

GROUP B

Alastair Binnie	Bristol-Myers Squibb
Clive Bowman	Daiichi-Sankyo
James Cronin	Purdue Pharma
Steven Frederick	Moderna
Preben Klavsen	H. Lundbeck
Sebastien Lefebvre	Alexion
Tomo Matsunaga	Takeda
Jens Noack Skærbæck	Novo-Nordisk
Jiangying Shi	Ipsen
George Davey-Smith	University of Bristol
Albert Wang	Bristol-Myers Squibb
Stacey Wasserman	GNS Healthcare
Ashok Upadhyay	Otsuka

GROUP C

Anastasia Christenson	Bristol-Myers Squibb
Matteo di Tommaso	Biogen
Beatrice Chapuzet	Servier
Joe Donahue	BioReference Labs
David Ficenec	Takeda
Olivier Gien	Sanofi
Klaus Hofenbitzer	Celgene
Iya Khalil	GNS Healthcare
Thomas Løngborn-Jensen	Novo-Nordisk
Arun Nayar	Amgen
Eric Neumann	Foundation Medicine
Alex Schuleit	H. Lundbeck
Phil Scordis	UCB

GROUP D

Dan Chapman	UCB
Hans Constandt	Ontoforce
Kaushal Desai	Bristol-Myers Squibb
Martin Erkens	F. Hoffmann-La Roche
George Goldsmith	Otsuka
Hall Gregg	Pfizer
Michael Hutnyan	Vencore
Fabien Jolly	Shire
Jim McGurk	Daiichi-Sankyo
David Sedlock	Takeda
Oodaye Shukla	Vencore
Kees van Bochove	The Hyve