Th 20th International Gene Forum

August 25—27, 2021, Tartu, Estonia

Organiser:
Institute of Genomics, University of Tartu

Acknowledgements:
Estonian Genome Foundation
Illumina

Contents

Programme

Programme

Wednesday, August 25th, 2021

14:00-14:15 OPENING OF THE CONFERENCE

Andres Metspalu,
Chairman of the Board, Estonian Genome Foundation, Estonia
"20 years of Gene Forum conference in Estonia: what has changed?"

Welcome by **Toomas Asser,** Rector, University of Tartu, Estonia

14:15-16:45 SESSION I *Genomics for healthcare*
(Moderator: Mait Metspalu)

Professor Sir Peter Donnelly, Genomics PLC, UK
„The potential for genomics to empower a new, prevention-first agenda for healthcare"

Professor Samuli Ripatti,
University of Helsinki/Institute for Molecular Medicine Finland (FIMM)
"The role of genomic information in common disease prevention"

Professor Gert Matthijs,
Center for Human Genetics, KU Leuven, Belgium
"Genomics in clinical diagnostics: guidelines and practice"

Professor Cathryn Lewis, King's College London, UK
„Polygenic risk scores for psychiatric disorders"

Thursday, August 26th, 2021

10:15-12:00 **SESSION II** COVID 19
(Moderator: Pärt Peterson)

Professor Andres Merits,
Institute of Technology, University of Tartu, Estonia
"Does complex structure of RNA genome impose restrictions for natural variation and evolution speed of SARS-CoV-2?"

Professor Adrian Hayday Francis Crick,
Institute and King's College, UK
"Learning immunology from COVID-19"

Andrea Ganna, Senior researcher at the Institute for Molecular Medicine Finland / instructor at the Harvard Medical School / instructor at the Massachusetts General Hospital, USA
"The covid-19 host genetics initiative"

12:00- 14:00 **Lunch Break**

14:00-16:45 **SESSION III** *Pharmacogenomics*
(Moderator: Lili Milani)

Dr. Charity Nofziger, PharmGenetix GmbH, Niederalm, Austria
"Solving the CYP2D6 Incidentalome"

Professor Ewan Pearson, University of Dundee, Scotland
"Genes, Drugs and Diabetes"

Dr. Jesse Swen, LUMC, Leiden, Netherland
"Challenges and opportunities for pre-emptive pharmacogenetic testing"

15:30-15:45 **Break**

15:45- 16.45 **SESSION IV** *Microbiome and health*
(Moderator: Elin Org)

Dr. Michael Zimmermann,
Structural and Computational Biology Unit, EMBL, Germany
"Identifying gut microbiome contributions to drug metabolism and toxicity"

Dr. Jingyuan Fu Professor at the University of Gröningen, Netherland
"From genome-wide association analysis to metagenome-wide association"

16:45-17:15 **Professor Arne Merilai,**
Institute of Cultural Research, University of Tartu, Estonia
"Poetics in Genes"

Friday, August 27th, 2021

10:00-13:00 **SESSION V** *Gene regulation and immunology*
(Moderators: Kaur Alasoo and Ana Rebane)

Dr. Judith Zaugg, EMBL, Germany
„Gene Regulation and Systems Epigenetics as Tools to Study Disease Mechanisms"

Urmo Võsa, Research fellow at the Institute of Genomics, University of Tartu, Estonia
"Leveraging large-scale eQTL mapping meta-analyses to prioritize trait-relevant genes in blood"

Dr. Shai Shen-Orr,
Associate Professor, Technion — Israel Institute of Technology, Israel
„Human immune monitoring coming of age"

11:30-11:45 **Break**

11:45-12.45 **Dr. Tilman Sanchez-Elsner,**
Associate Professor University of Southampton, UK
„The LPS-induction of protein expression in monocytic cells is associated to mRNA binding to ribosomes, RNA sequence features and microRNA targeting"

Dr. Richard Oram, Associate Professor and Diabetes UK Harry Keen Fellow, University of Exeter, UK
"Can we predict the future? Improved classification and prediction of type 1 diabetes and celiac disease with polygenic scores"

12.45-13.15 **SESSION VI** *Population genomics*

Dr. Matteo Fumagalli, Lecturer at the Imperial College, UK
"Deep learning in population genomics: can AI solve the unsolvable?"

13.15-13.30 **Closing remarks**

Speakers

SIR PETER DONNELY

Professor Sir Peter Donnelly is CEO of Genomics plc and Emeritus Professor of Statistical Science at the University of Oxford where he was Director of the Wellcome Centre for Human Genetics from 2009-2017.

Following a successful academic career in statistical and population genetics, Peter was centrally involved in many of the major human genetics projects over the last 20 years. He played a major role in the International HapMap Project, the successor to the Human Genome Project, and later chairing the landmark Wellcome Trust Case Control Consortium and its successor (a large international collaboration studying the genetic basis of more than 20 common human diseases and conditions in over 60,000 people). More recently he led an Oxford collaboration with Illumina to sequence the genomes of 500 individuals with a range of clinical conditions to assess the potential for whole-genome sequencing in clinical medicine - a study which led to the NHS 100,000 Genomes Project, and led the genetic work culminating in the genotyping of all 500,000 participants in UK Biobank.

In 2014, Peter and 3 colleagues, including Gil McVean the then inaugural director of Oxford's Big Data Institute, founded Genomics plc, with Peter becoming the company's CEO in 2017. Genomics now employs over 100 people, with offices in Oxford and Cambridge, UK, and Cambridge Mass. It has developed an extensive data platform linking genetic variation in humans to phenotypic outcomes, and powerful statistical and machine learning tools to jointly analyse the data. Genomics uses this in two ways, in Therapeutics to find novel drug targets, and in Precision Health to develop and deploy powerful risk prediction tools for population health and clinical decision support.

The potential for genomics to empower a new, prevention-first agenda for healthcare

For all the common chronic human diseases, and the common cancers, genetic variation is one key part of the differences between individuals in disease susceptibility. Environmental and lifestyle differences also play a central role for many diseases, while in others genetics is the major risk factor. Over the last 10-15 years there has been an explosion in our knowledge of common genetic variants which contribute to disease risk, but to date this knowledge has had limited impact on patients.

For a particular disease, a polygenic risk score (PRS) combines information from hundreds of thousands, or millions, of variants in an individual's genome. Although the variants individually have only a tiny impact on risk, their combination provides a powerful tool for risk stratification. Across a population, the individuals with polygenic risk scores in the highest few percent can be at 3-5 fold increased risk of disease. For diseases like cardiovascular disease, where clinicians already predict disease risk using standard tools, the inclusion of the genetic component of risk, via PRS, substantially improves risk prediction. For other diseases, including breast cancer, PRS is by some way the most predictive risk factor. Even for individuals who carry pathogenic mutations of large effect, their overall disease risk is modulated by their PRS for the disease in question.

Polygenic risk scores have the potential to improve patient outcomes, and the sustainability of healthcare systems, by enabling much better and earlier identification of individuals at substantially increased risk of particular diseases. They herald a new era of "Genomic Prevention" in healthcare, in which the combination of the new polygenic risk scores with existing non-genetic risk factors will power a prevention first agenda for population health — the identification of high-risk individuals who are currently invisible to the system, to allow medical interventions, lifestyle changes, and targeted screening to prevent disease, or (especially for cancers) to detect it much earlier when curative interventions are available.

This is a major focus at Genomics plc. Our extensive genotype-phenotype data platform and algorithms for deriving powerful polygenic risk scores mean that we have PRSs for over 50 diseases, which compare favourably with other published and commercial scores. The talk will describe these and their potential clinical applications, as well as the implementation projects we are undertaking with the UK National Health Service and Stanford Health Care to pilot their integration into clinical pathways.

SAMULI RIPATTI

Samuli Ripatti, PhD, is a Vice Director at the Institute for Molecular Medicine Finland (FIMM), a professor of Biometry at the Faculty of Medicine (UH) and a Scholar at the Broad Institute of MIT and Harvard in Cambridge, MA, USA. He is chairing the Academy of Finland Centre of Excellence in Complex Disease Genetics and H2020-funded Intervene Consortium. His research group studies genetic variation in the Finnish population and its effects on common complex disease risks and management. His research focuses in particular on cardiometabolic diseases and cancer as models to learn about disease mechanisms and genome-based strategies for diagnosis, prevention and stratified treatment. He has played a central role in developing and testing the use of polygenic risk in prevention of cardiometabolic diseases and common cancers. He is strongly involved in doctoral training and is currently chairing Doctoral Programmes in Biomedicine and Population Health at University of Helsinki.

The role of genomic information in common disease prevention

The genetic discoveries over the past 10-15 years have not only advanced our understanding of the underlying causes in hundreds of diseases, but also taught us about the highly polygenic nature of common complex diseases. Statistical algorithms are now able to capture efficiently these genome-wide effects on disease risks and turn them into person-specific risk estimates or polygenic risk scores.

In this talk I will review the development of polygenic risk scores, talk about the latest results in applying them to estimate risks in diseases such as coronary heart disease and breast cancer, and discuss the opportunities they provide in prevention or early detection in many diseases. I also describe the experiences of returning the risk estimates individuals and the changes in positive health behavior this has helped to trigger.

GERT MATTHIJS

Gert MATTHIJS, PhD (°1963) is a molecular geneticist. He is the head of the Laboratory for Molecular Diagnostics at the Center for Human Genetics in Leuven, and Professor at the University of Leuven, Belgium. He has been involved in the diagnostics of inherited diseases since 1994. His research interest is in Congenital Disorders of Glycosylation (CDG), a group of rare inborn errors of metabolism. His group is focusing on the systematic search for novel types of CDG. The success is partly due to the fact that the Leuven group has committed itself, since 1999, to the coordination of EUROGLYCANET, a European network on CDG.

His translational research activities deal with the development and validation of novel technologies for diagnostic use. Genomic sequencing is currently being implemented at the Center for Human Genetics, for direct use in the clinic.

He has been the coordinator of EuroGentest, a network for development, harmonization and standardization of genetic testing in Europe, originally funded by the European Commission and now integrated into the European Society for Human Genetics (ESHG), and has moderated the generation of guidelines for NGS diagnostics.

Gert Matthijs is the president of the Belgian Society for Human Genetics (BeSHG). At the international policy level, he is a member of Global Genomic Medicine Initiative (G2MC).

He has also led a workgroup on legal, ethical and societal aspects of total genome analysis (in the context of the Metaforum initiative of the University of Leuven). He co-authored a book on genetics and genomics for the public: 'The Human Recipe' (University Press Leuven, 2016). This illustrates his interest in the societal aspects of genetics.

Genomics in clinical diagnostics: guidelines and practice

Next-generation sequencing (NGS) has been introduced in clinical genetic laboratories more than a decade ago. We present, on behalf of EuroGentest and the European Society of Human Genetics, guidelines for the evaluation and validation of NGS for the diagnosis of genetic disorders.

Indeed, the diagnostic use brings specific challenges like the solid validation of methods, the clinical interpretation of results and the storage and exchange of data. In 2016, NGS guidelines have been issued for custom gene panels and whole exome sequencing (WGS). An expert group has now been working on compiling, integrating and completing guidelines for whole genome sequencing. Among other statements, the guidelines deliver major definitions like 'diagnostic utility' and 'reportable range'. We believe that defining the diagnostic utility is the laboratory's first duty when preparing to offer diagnostic NGS and WGS. Importantly, the guidelines provide a basis for the analytical and clinical validation of new NGS platforms and methods. Also, the defined parameters should allow to compare the plethora of tests offered by different laboratories, and aid technical experts in the evaluation of the quality assurance. As far as 'reportable range' is concerned, we propose the use of 3 specific percentages depending on the reference (technical target, coverage of transcript in a gene panel, coverage with reference to the genome) which will again allow to compare individual results within runs, between tests and between laboratories. For the classification of variants, we refer to recommendations published by other professional societies. The abundance of variants of uncertain significance (VUS) is, from a clinical perspective, the most confounding issue, and variant classification certainly deserves further guidance. By promoting the guidelines, we hope to contribute to the harmonization and standardization of diagnostic NGS testing, which are two of the objectives of EuroGentest.

CATHRYN LEWIS

Cathryn Lewis is Professor of Genetic Epidemiology & Statistics at King's College London, and Head of Department at the Social, Genetic and Developmental Psychiatry Centre. Her academic training is in statistics, and she has been analysing genetic studies since her PhD. She co-chairs the Major Depressive Disorder Working group of the Psychiatric Genomics Consortium, and leads the NIHR Maudsley Biomedical Research Centre Biomarkers and Genomics theme. She has published over 350 papers on statistical genetics methodology and analysis. Her research programme identifies and characterises genetic variants associated with human disease, including depression, schizophrenia, stroke and in pharmacogenetics. A major research focus is risk assessment, determining how the polygenic component of mental health disorders can be measured accurately and communicated effectively.

Polygenic risk scores for psychiatric disorders

Genetic studies have enabled researchers to identify variants associated with diseases and shown unequivocally that psychiatric disorders have a polygenic genetic architecture. Polygenic risk scores combine information across the genome to capture part of an individual's genetic susceptibility to a disorder. A polygenic score can be used to estimate an individual's lifetime genetic risk of disease, and these are widely used in research studies to characterise psychiatric disorders and risk. Polygenic scores could also be used for prognosis or to predict response to treatment. However, the discriminative ability for polygenic scores is currently low, and their clinical utility has yet to be established. I will describe the progress made in psychiatric disorders, and I will discuss the potential uses for polygenic scores and the challenges that must be overcome before they are applied in a clinical setting.

ADRIAN HAYDAY

Trained in biochemistry and with a PhD in molecular virology, Adrian Hayday took up immunology at MIT where, in 1985, he and his colleagues described the entirely unanticipated T cell receptor γ-chain genes. Since then he has shown γδ T cell biology to be overtly distinct from conventional T cell biology, and provided the first evidence that γδ T cells naturally protect against carcinogenesis. Since then he has identified molecular mechanisms by which γδ T cells discriminate tumours from normal tissue, and is committed to the cells' clinical application.

He is first or corresponding author on more than 120 out of a total of 230 publications of which over 150 describe original research. He has co-authored many patents. His awards include the William Clyde deVane Medal, Yale College's highest honour for scholarship; the UK Business of Science Award, 2017; and election as Fellow of the Royal Society, and of the Academy of Medical Sciences. He was elected to lead the British Society of Immunology (2005-09) and has co-organized major meetings including the 2014 Gordon Conference in Immunochemistry and Immunobiology, the scientific programme for the 2012 European Congress of Immunology, the 2016 international γδ T cell forum, and the first 2021 Cold Spring Harbor COVID-19 meeting.

He has chaired Wellcome Trust, CRUK, and American Cancer Society funding committees, and has served on many advisory boards, including the Pasteur Institute, Paris, the Max Planck Society, and the University of Kyoto. In the private sector he was on the SAB of MedImmune and has co-founded three biotech start-up companies.

Learning immunology from COVID-19

The immune system is implicated in evermore diverse aspects of pathophysiology, provoking increased interest in human immune-monitoring. However, we remain uncertain of what defines a healthy human immune system. Addressing this, much has been learned over the past decade from the application of several "omics" technologies following human vaccination. However, SARS-CoV-2 infection and consequent COVID-19 provided a rare opportunity to compare human immune responses in the settings of vaccination and live infection. In making those comparisons, we learned several things that are not evident from the textbooks. Those include, highly selective immunodeficiencies of cancer patients toward vaccination; and chaotic immune responses to infection displayed by highly selective components of the immune response. As well as casting new light on how the human immune system behaves in vivo, these findings can inform the application of immunotherapies and vaccines in curative settings, such as cancer and inflammatory diseases.

ANDREA GANNA

Andrea is a FIMM-EMBL-group leader at FIMM and an instructor at Harvard Medical School and Massachusetts General Hospital. His research interests lie on the intersection between epidemiology, genetics and statistics. Andrea has authored and co-authored both methodological and applied papers focused on leveraging large scale epidemiological datasets to identify novel socio-demographic, metabolic and genetic markers of common complex diseases. He has extensive expertise in statistical genetics and has been working with large-scale exome and genome sequencing data. He is co-leading two major international consortia: the COVID-19 host genetic initiative and the INTERVENE consortium. His research vision is to integrate genetic data and information from electronic health record/national health registries to enhance early detection of common diseases and public health interventions, for which he has been awarded an ERC starting grant.

The covid-19 host genetics initiative

The genetic makeup of an individual contributes to susceptibility and response to viral infection. While environmental, clinical and social factors play a role in exposure to SARS-CoV-2 and COVID-19 disease severity, host genetics may also be important. Identifying host-specific genetic factors indicate biological mechanisms of therapeutic relevance and clarify causal relationships of modifiable environmental risk factors for SARS-CoV-2 infection and outcomes. We formed a global network of researchers to investigate the role of human genetics in SARS-COV-2 infection and COVID-19 severity. We describe the results of three genome-wide association meta-analyses comprising up to 49,562 COVID-19 patients from 46 studies across 19 countries worldwide. We reported 13 genome-wide significant loci that are associated with SARS-CoV-2 infection or severe manifestations of COVID-19. Several of these loci correspond to previously documented associations to lung or autoimmune and inflammatory diseases. They also represent potentially actionable mechanisms in response to infection. We further identified smoking and body mass index as causal risk factors for severe COVID-19. The identification of novel host genetic factors associated with COVID-19, with unprecedented speed, was enabled by prioritization of shared resources and analytical frameworks. This working model of international collaboration provides a blue-print for future genetic discoveries in the event of pandemics or for any complex human disease.

EWAN PEARSON

Ewan Pearson is Professor of Diabetic Medicine at the University of Dundee, Visiting Professor at the University of Edinburgh, and Honorary Consultant in Diabetes and Endocrinology at Ninewells Hospital and Medical School in Dundee. In the School of Medicine, he is the Head of Division, Population Health & Genomics, and the Director of the Dundee Clinical Academic Track. He is an Associate Director for the Health Data Research-UK Scotland substantive site.

Professor Pearson obtained his medical degree from the University of Cambridge School of Clinical Medicine, UK. He undertook a Wellcome Trust Clinical Training fellowship with Prof Andrew Hattersley at the University of Exeter Medical School, UK and completed his PhD in the physiology and treatment of monogenic diabetes. He then moved to Dundee where he was supported by a Chief Scientist Office Clinician Scientist fellowship and, more recently, by a Wellcome Trust Investigator Award. Over the last ten years in Dundee his research interests have been in the phenotypic and genotypic determinants of drug response in diabetes, and in stratified approaches to the management of diabetes. Ewan has been awarded the Royal College of Physicians of Edinburgh Croom Lecture, an EASD Rising Star award, the Diabetes UK RD Lawrence Lecture and the EASD Minkowski Award for his work in these areas.

Genes, Drugs and Diabetes

People are all different, and this is no different when we consider people with diabetes, yet the current approaches to management of diabetes tend to treat everyone the same. The field of precision medicine aims to recognise these differences – whether at the level of their phenotype or at the molecular level. Faced with multiple, and increasing, treatment options for diabetes as well as increasing healthcare costs there is a clear need to target therapy to maximise benefit and reduce harm for every patient with diabetes.

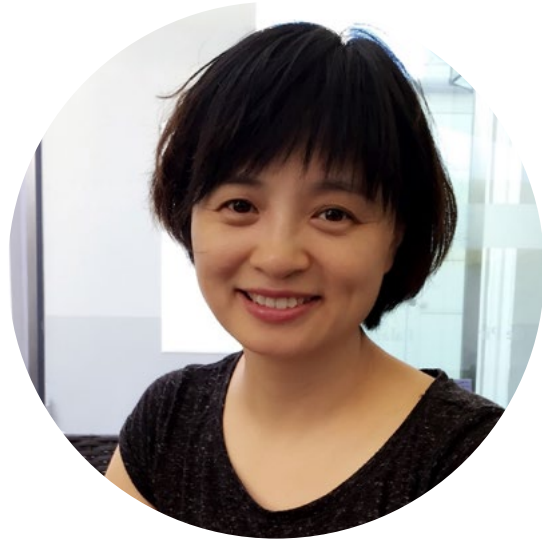
This talk will discuss advances in precision medicine and pharmacogenetics in diabetes. I will highlight recent work on how phenotypic variation matters, and how this maps to genetic variation, and will provide an overview of how genetic variants alter glycaemic response to commonly used diabetes drugs and how these inform on disease and drug mechanism. I will provide a framework for how we might implement pharmacogenomics in diabetes clinical care in the near future.

MICHAEL ZIMMERMANN

Michael Zimmermann is a group leader at EMBL in Heidelberg. Michael received a Bachelor's degree in Pharmaceutical Sciences from Basel University and a Master's degree in Biotechnology from the Ecole Supérieure de Biotechnologie in Strasbourg. He performed his Ph.D. work with Uwe Sauer at ETH Zurich, where he developed metabolomics and Systems Biology approaches to study microbial metabolism in the context of host-pathogen interactions. For his postdoctoral training, Michael joined Andrew Goodman's group at Yale University to study the molecular mechanisms of host-microbiome interactions. Michael's most recent research investigates how the human gut microbiome contributes to drug metabolism, which can result in interpersonal variation of patients' drug response and toxicity. His group employs bacterial genetics, metabolomics, gnotobiotic mouse work, and mathematical modeling to systematically map microbial drug transformations and to quantitatively separate microbial and host drug metabolism in vivo.

Identifying gut microbiome contributions to drug metabolism and toxicity

Individuals vary widely in their drug responses, which can be dangerous and expensive due to significant treatment delays and adverse effects. Growing evidence implicates the gut microbiome in this variability, however the molecular mechanisms remain mostly unknown. Using antiviral nucleoside analogues and clonazepam as examples, we recently reported experimental and computational approaches to separate host and gut microbiota contributions to drug metabolism. The resulting pharmacokinetic models identified measurable physiological, microbial and chemical parameters that dictate host and microbiome contributions to the metabolism of xenobiotics. To systematically map the drug metabolizing capacity of the gut microbiota and assess its potential contribution to drug metabolism, we further measured the ability of 76 diverse human gut bacteria to metabolize each of 271 oral drugs. We found that two thirds of these drugs are chemically modified by at least one of the tested microbes. Through combination of high-throughput bacterial genetics with mass spectrometry, we systematically identified drug-metabolizing microbial gene products. These gene products better explain the drug-metabolizing capacity of bacterial strains than their phylogenetic classification. We further demonstrate that the abundance of homologs of these gene products predict the capacity of complete human gut communities to metabolize the targeted drugs. These causal links between microbiota gene content and metabolic activities connect inter-individual microbiome variability to interpersonal differences in drug metabolism, which has translatable potential on medical therapy and drug development across multiple disease indications.



JINGYUAN FU

Dr. Jingyuan Fu is a professor of systems medicine in the University Medical Centre Groningen, the Netherlands, with a particular focus on integrative genomics and host-microbe interactions in complex diseases. She obtained a BSc in Biochemistry, a MSc in Biotechnology and Bioinformatics (with distinction), and a PhD in systems genetics *(cum laude)*. Via this route she developed her research line on systems genetics in complex traits and became an expert in integrative genomics and systems biology. Her research profile involves large-scale, population cohort-based studies on genetics, the gut microbiome, and various omics datasets. She holds ERC-CoG, NWO-VICI and several national consortia grants, such as the IN-CONTROL consortium and the Netherlands Organ-on-a-chip Initiatives.

From genome-wide association analysis to metagenome-wide association

Genome-wide association analysis has proven to be a power tool that uses the natural variation of the human genome to identify novel genes associated with individual's susceptibility of various complex diseases and traits. However, the human genome does not account for all risks. In recent years, the importance of the gut microbiome in human health has been established. While we now know a great deal about gut microbiome composition, the impact of genetic make-up of our microbiome – our "metagenome" remains unclear.

Here we presented bacterial SNPs and structural variation analysis in about 1,500 deeply phenotyped individuals from the Lifelines-DEEP cohort. We not only conducted bacterial GWAS with host's phenotypes, such as bile acids, to identify novel bacterial species and genes, but also evaluated the individual's specificity and temporal stability of bacterial genetic landscape over 4 years.

All together, our study supports the concept that humans are holobionts and that genes from the host, microbiome, virome and other members, all together contribute to the physiological function of a human body.

JUDITH ZAUGG

Judith Zaugg is a group leader at EMBL leading a group in Computational and Regulatory Genomics at EMBL Heidelberg (Germany) since 2014 (www.zaugg.embl.de). Since 2018 she has an additional faculty appointment focussed on Stem Cell-Niche Networks at the Molecular Medicine Partnership Unit (MMPU), a joint venture between EMBL and the University hospital Heidelberg (Germany). Judith studied chemical and molecular biology at ETH in Zurich (Switzerland), obtained her PhD at Cambridge University and EMBL-EBI (UK), and went to Stanford (USA) for her postdoctoral research where she was co-leading studies on epigenetic variation across individuals, and discovered that many so-called epigenetic marks do in fact have a genetic basis. Her computational group at EMBL uses systems epigenetics approaches to understand basic gene regulatory principles (e.g. Ibarra 2020), develops tools to integrate multiomics data (e.g. Berest 2019,), and applies these principles to understand disease mechanisms (e.g. Reyes-Palomares 2020). Her MMPU group focuses on understanding the interaction between the stem cell populations within the hematopoietic niche, and how aging and blood cancer affect these interactions (Garg 2019). In addition to her research, Judith Zaugg serves on the Scientific Advisory Board of the Nordic Centre of Molecular Medicine (NCMM) and on the Scientific Advisory Panel of the Hungarian Centre of Excellence in Molecular Medicine (HCEMM). She acts as academic editor at Life Science Alliance (LSA) and is the main scientific organiser of the bi-yearly EMBL|EMBO symposium on multiomics data integration (www.embo-embl-symposia.org/symposia/2021/EES21-09).

Gene Regulation and Systems Epigenetics as Tools to Study Disease Mechanisms

Phenotypic variation (including disease) across individuals has two main sources: genetic variation across individuals, and variation in environmental exposures. In the past 20 years we have made tremendous efforts in mapping common genetic variants to complex traits and diseases. Yet, the majority of these disease-associated variants lie in the non-coding part of the genome, which makes is very difficult to understand the underlying molecular mechanisms. For the environmental variation we still know very little, except that part of it will be stored in epigenetic signatures. Therefore, to gain a mechanistic understanding of phenotypic variation and potentially identify molecular intervention points of disease, it is essential to understand how genetic variants affect gene regulation, and to generate models that predict how external signals and epigenetics contribute to overall variability. Here I will present our recent work on predictive regulatory network models that aim at integrating environmental signaling with genetic and epigenetic variation in enhancers to understand complex phenotypes. I will show applications of our predictive models to pulmonary arterial hypertension, ageing of the bone marrow niche and immune-system related traits. Overall, our integrative computational tools with a focus on gene regulation provide a powerful approach to gain mechanistic insights into complex biological processes.

URMO VÕSA

Urmo Võsa is a research fellow in the Institute of Genomics, Estonian Genome Centre, University of Tartu. He received his PhD from University of Tartu in 2016 by investigating the role of microRNAs in the development of non-small cell lung cancer and their regulation by genomic variation. Urmo did his postdoc in Prof. Lude Franke's work group (University Medical Center Groningen, Groningen, Netherlands) where he investigated the regulation of gene expression in blood and focused on distal effects of trait-associated genetic variants.

His current research interests involve combining the information from large-scale cohort studies with different endophenotypes, in order to interpret the results of genome-wide association studies and to prioritize causal genes for complex traits. Together with his colleagues from University Medical Center Groningen and University of Tartu, he continues to lead the eQTLGen Consortium which leverages tens of thousands of blood samples to comprehensively characterize the genetic architecture of gene expression in blood.

Leveraging large-scale eQTL mapping meta-analyses to prioritize trait-relevant genes in blood

Despite the thousands of statistical associations discovered by genome-wide association studies (GWAS), pinpointing the causal genes for each trait is still challenging, due to the complex linkage disequilibrium patterns in the individual risk loci and polygenic nature of most traits. Local *(cis)* expression quantitative trait loci (*cis*-eQTLs) are widely used to prioritize putative causal genes in individual trait-associated loci. However, *cis*-eQTLs explain only the modest proportion of trait heritability and strong local effects are likely to be dampened by compensatory post-transcriptional buffering. In the other hand, each distal *(trans)* eQTL variant can affect many genes from all over the genome, and have a widespread impact on regulatory networks. *Trans*-eQTLs are especially interesting in the context of recently proposed 'omnigenic' model which postulates that large part of the heritability of complex traits is dominated by numerous weak *trans*-effects and that such distal effects could converge in the smaller set of 'core' genes, more directly related to the end-phenotype. However, distal eQTL effects are generally weak and therefore highly powered datasets are required to identify and interpret those in a meaningful manner.

In order to fill these gaps, we performed large-scale meta-analyses in 31,684 blood samples from 37 eQTLGen Consortium datasets, representing ~6-fold increase in the sample size compared to previous similar studies. We identified robust *cis*-eQTL effects for the large majority of tested genes and observed that these were replicable in different tissues. By testing the subset of variants previously associated with complex phenotypes, we identified numerous *trans*-eQTL effects, and performed comprehensive replication analyses in post-mortem tissues, purified cell types, cell lines and single-cell RNA-seq datasets from peripheral blood mononuclear cells. We further investigated the putative molecular mechanisms leading to distal effects and highlighted numerous examples where *trans*-eQTL effects point to interpretable biology.

Finally, we sought to identify putative 'core' genes where the effects of multiple trait-associated variants converge. To do so, we calculated the polygenic scores for >1,000 GWAS phenotypes and correlated those scores with expression of each gene. We identified potential 'core' genes correlating with the polygenic scores for several blood-related and non-blood-related traits, and outlined additional interpretable examples.

In summary, this study represents the largest eQTL mapping effort to date and provides valuable resource which can and has been used for interpretation of GWAS results. During my talk I will discuss the key insights we have learned from this project and outline how these results have been utilized in the subsequent studies.

SHAI SHEN-ORR

Shai Shen-Orr is an Associate Prof. at the Faculty of Medicine at the Technion – Israel Institute of Technology, where he heads the Systems Immunology & Precision Medicine laboratory since 2012. His research is focused on charting the immune system landscape – namely the principles by which the immune system varies over time as a function of environment and genetics. For this, his lab develops novel computational methodologies, empowering human immune monitoring for precision medicine.

Prof. Shen-Orr received a BSc from the Technion in Information Systems (1999), an M.Sc. in Bioinformatics at the Weizmann Institute of Science (2002), a Ph.D. from Harvard University in Biochemistry (2007) and performed his postdoctoral studies at Stanford University. His research has been cited numerous times and has been featured in systems biology textbooks for students.

Prof. Shen-Orr's research laid the foundation of CytoReason, a company that uses a machine learning model of the immune system; which it applies to drug development in collaboration with leading pharma companies.

Human immune monitoring coming of age

Individuals' show high inter-individual phenotypic and functional variation in immune cells, especially in older adults. This variation may describe a change in immunological state which ultimately can have clinical implications. Here I will describe two human immune monitoring studies, in which we longitudinally tracked individuals to reveal systematic patterned trajectories of immunological changes:

In the first, we analyzed the dynamics of anti-TNF response in Crohn's patients on the short time scale of weeks following drug treatment to reveal a cell-centered response network, understand how individuals vary from it and where the bottlenecks for successful response lie.

Second, we studied the longitudinal dynamics of immune system alterations in older adults over the course of nine years. Characterizing the dynamics of individual cell-types we identify an attractor point for different features on which the system converges likely due to adaptation of cellular stoichiometries to fit the life history of the individual. A high dimensional analysis of these longitudinally changing features allows to identify an immune age metric, different from chronological age, which shows clear functional immunological implications and has prognostic clinical value with respect to all-cause-mortality in healthy older adults, beyond well-established risk factors and better than other biological age metrics.

Our findings shed light on the origins and long-term dynamics of individual immune-cellular variation and provide a quantitative framework to study immune variation, allowing for clearer identification of immune based precision medicine.

TILMAN SANCHEZ-ELSNER

Tilman Sanchez-Elsner is Associate Professor in the Faculty of Medicine, University of Southampton (UK) since 2007.

Tilman studied Biology at the Universidad Complutense of Madrid, Spain, pursuing the PhD in the same University at the Department of Biochemistry and Molecular Biology (June 2002). During his PhD, Tilman focused on the transcriptional control of angiogenesis and repair in human, describing the co-operation between two transcription factors, HIF-1a (Hypoxia Inducible Factor) and Smad3 (which mediates TGF-a inducible responses). After his PhD, and being interested in a more global kind of regulation of transcription, he then moved into the field of epigenetics. He was a postdoctoral fellow (2002-2006) at the University of California at Riverside, working with Dr Frank Sauer where he described a novel non-coding RNA, crucial for the epigenetic control of the fruit fly development.

He was then interested in applying the acquired knowledge in non-coding RNAs to the human model, and more specifically, to work in Immunology and hematopoietic development. He started working in the field of microRNAs since 2006, in the lab of Dr A.L. Corbi, as a senior postdoc. In this lab, he acquired expertise in the immunological techniques, as well as in the field of MicroRNAs.

After that, he joined the University of Southampton, in 2007. Tilman has now over 14 years of independent research experience in the areas of gene expression, microRNA and macrophage biology with an interest in respiratory diseases (lung cancer, asthma, COPD) as well as inflammatory bowel diseases (IBD; Ulcerative Colitis, Crohn´s Disease). Over the years, he has studied the role of microRNAs in modulating gene expression and function in macrophages as well as T-Cells. His work discovered that dysregulation of networks in disease (asthma and IBD) modify the biology of macrophages and their ability to respond to infection and to interact with other cells (both immune cells, such as T-cells, and epithelial cells). In the past five years, he has worked closely with Prof. Christian Ottensmeier and Dr Pandurangan Vijayanand in unravelling the transcriptomic regulation of macrophages, T-Cells, B-Cells and NK cells in lung cancer.

Tilman's team has found striking evidence that macrophages (Tumour Associated, TAM) are key to recruit and maintain T-Cells and are now quite interested in establishing how this cross-talk can affect our ability to enhance the effectiveness of immunotherapy drugs targeting T-Cells. His goal is now to be able to manipulate TAMs in order to enhance the number and function of T-Cells in the tumour to boost immunotherapy in patients . We have published our microRNA, macrophage and T-Cell work in *Nature Immunology, Journal of Experimental Medicine, Journal for ImmunoTherapy of Cancer, American Journal of Respiratory and Critical Care Medicine and The Journal of Immunology* in the last two years. We are interested in further exploring the role of microRNAs and transcriptomics in inflammatory diseases (IBD, asthma) and cancer and in basic biology (role of microRNAs in regulating binding to ribosomes).

The LPS-induction of protein expression in monocytic cells is associated to mRNA binding to ribosomes, RNA sequence features and microRNA targeting

It is essential to understand cellular and molecular mechanisms regulating mRNA translation into protein in order to explain the low correlation between mRNA and protein expression levels. Although there is significant research into specific regulatory mechanisms regulating individual genes, the role of translational control at a wide transcriptome/proteome scale has not been tackled yet. In our work we have addressed several factors that we believe are involved in mRNA translation into protein, at a large scale: mRNA binding to ribosomes, RNA sequence features, and microRNA expression. We analysed lipopolysaccharide LPS-dependent changes in THP-1 cells, in mRNA/microRNA (RNA-seq and Fraq-seq binding to ribosomes) and protein expression (Isobaric proteomics) and applied machine learning/statistical inference methods of clustering and regression to elucidate what is the impact of the different regulatory elements. Our data showed that LPS not only affected mRNA transcription, but also affected the binding of mRNAs to ribosomes, resulting in changes in protein levels unexpected by only using transcriptomics data. Interestingly, mRNA binding to ribosomes were associated with distinctive cellular functions, suggesting a functional relevance to this post-transcriptional regulation. RNA sequence features (ORF length, 5' and 3'UTR length, and codon and tRNA adaptation indexes) as well as microRNAs expression also had a differential role depending on biology and binding to ribosomes. Our data indicate that mRNA features and microRNA expression are essential tools to understand translational regulation that will help interpret transcriptomics data, by improving the prediction of protein expression.

RICHARD ORAM

Richard Oram is a Diabetes UK Harry Keen Fellow specializing in Type 1 diabetes at the Institute of Biomedical and Clinical Science and NIHR Exeter Clinical Research Facility. Dr Oram's Diabetes UK funded PhD showed that most people with type 1 diabetes still make small amounts of their own insulin. He continues to study the biology of beta cell loss and the impact of persistent beta cell function on complications. His Harry Keen fellowship and the Helmsley Charitable Trust fund Dr's Oram and Tree (KCL) to study a rare subtype of extremely early onset type 1 diabetes - diagnosed under the age of 1 year. Working with Mike Weedon, he has developed a cheap, simple method to assess genetic risk in type 1 diabetes – a T1D genetic risk score (T1D GRS). He and Kash Patel showed this can differentiate type 1 from type 2 diabetes and monogenic diabetes. Collaborating with Bill Hagopian (PNRI, Seattle), Seth Sharp (DUK PhD Student) recently improved the T1D GRS to include more HLA alleles and their interactions. With Ranjan Yajnik (Pune, India), he has shown the utility of this T1D genetic risk score in Indians. With Mike Weedon, Bill Hagopian and the NIH funded SEARCH study, Richard is working of uses of the T1D GRS in an ethnically diverse US population. Funded by JDRF, he is investigating combining genetic information with longitudinal biomarkers to better predict type 1 diabetes and other autoimmune diseases from birth in the NIH TEDDY study (Ferrat Nature Medicine 2020). He is integrating genetics into better prediction of type 1 diabetes in the Trialnet Pathway to Prevention study. Dr Oram is working with Bill Hagopian to test a newborn genetic screen with autoantibody monitoring follow up to prevent childhood diabetic ketoacidosis in the CASCADE and PLEDGE studies.

Can we predict the future? Improved classification and prediction of type 1 diabetes and celiac disease with polygenic scores

My talk will discuss development of a type 1 diabetes genetic risk score. I will highlight the practical clinical utility in aiding classification of type 1 diabetes. In the second part of the talk I will discuss application of a type 1 diabetes, and if time, celiac genetic risk score, to newborn screening with a goal to predict future childhood autoimmunity in newborns.

MATTEO FUMAGALLI

Matteo Fumagalli obtained is PhD in Bioengineering from Polytechnic University of Milan, Italy, in 2011. After that, he has been an EMBO postdoctoral researcher at the University of California, Berkeley, with Rasmus Nielsen and then an HFSP research fellow at University College London with Francois Balloux. Since 2017 he is a Lecturer (Assistant Professor) in Quantitative Evolution at Imperial College London. His research focuses on the development and application of computational methods to infer demographic histories and signals of natural selection from population genomic data. He is particularly interested in the interplay between human evolution and disease susceptibility, especially for autoimmune and cardiovascular disorders. He also implements bioinformatic tools to process low-coverage high-throughput sequencing data mostly for the analysis of non-model species. More recently, his research moved to the inference of fine-scale history of admixed human population and the application of machine learning to genomic data.

Deep learning in population genomics: can AI solve the unsolvable?

The application of deep learning, a branch of machine learning based on multi-layer neural networks, to biological data appears to be a revolutionary step to address complex questions, even in population genomics and evolutionary biology. In this seminar, I will first introduce some basic concepts of deep learning and how it can be applied to population genomic data for evolutionary inferences. I will then show some recent and ongoing work to identify signatures of positive and balancing selection, adaptive archaic introgression, and assortative mating from genomic data using deep learning algorithms. I will finally outline some future research directions.

ANDRES MERITS

I obtained PhD in biology 1994 (Moscow, Russia) for work with plant viruses. During post-doctoral studies (Helsinki, Finland) I focused on the molecular biology and biotechnology of arboviruses, originally alphaviruses and subsequently also flaviviruses. These topics become central for my studies in Estonia. Most recently – and not uncommonly – our group started also to work with SARS-CoV-2. I was elected to the position of Professor of Applied Virology in Institute of Technology, University of Tartu (2007) and subsequently Academy Research Professor (2019). My current research group consists from 8 researchers, three of them with Ph.D. or equivalent degree. I have published >165 original research papers and successfully supervised 13 PhD students.

Does complex structure of RNA genome impose restrictions for natural variation and evolution speed of SARS-CoV-2?

SARS-CoV-2 is the causative agent of COVID-19, the most devastating pandemics of 21. century. Surprisingly, despite huge impact of COVID-19 on all areas of society, culture and economy the real cause of all these problem, the SARS-CoV-2 and its molecular biology, has received disproportionally little attention and, at least in Estonia, zero research funding. Neglecting the core of the problem has effectively resulted in false image of SARS-CoV-2 as completely unpredictable agent and forcing society to deal with consequences, rather than reasons, of SARS-CoV-2 related problems. Thus, emergence of new variants of concern (VOC) and other SARS-CoV-2 variants is regarded as something inevitable and unpredictable. Nevertheless, observations have revealed that "novel" properties of VOCs seem to be due to relatively small number of changes in virus genome that emerge in different combination and in response to different challenges.

Analysis of high-order structure of SARS-CoV-2 RNA demonstrates that the viral genome has extremely complex structure. Most of the residues of the SARS-CoV-2 RNAs are involved in short and long range interactions and/or in interactions with host RNAs. Most likely these interactions are functionally essential as SARS-CoV-2 has very low tolerance for synonymous mutations. Consequently, the SARS-CoV-2 genome tends, either during natural evaluation or during passaging and selection experiments in cell culture, mostly acquire non-synonymous changes. Changes resulting from selection of resistance to antiviral drug remdesivir occurred, as predicted, in RNA polymerase of the virus but were also accompanied with numerous mutations in the S-protein region. Surprisingly, the positions of these mutations are largely the same as observed in genomes of VOCs. This data indicates that SARS-CoV-2 has considerable limitations in its genetic flexibility and tends, both in real life and under laboratory conditions, mostly exploit relatively limited set of standard solutions rather than generating new and unexpected solutions.

CHARITY NOFZIGER

Dr. Charity Nofziger received her PhD in Biology in 2008 from Indiana University Purdue University of Indianapolis in Indiana, USA. From there, Dr. Nofziger accepted a post-doctoral fellowship at the Paracelsus Medical University in Salzburg Austria in the Department of Pharmacology and Toxicology where she started her work in the field of pharmacogenetics. Within the same institution, she accepted a second post-doctoral fellowship funded by Roche with a focus on the expression of drug transporters in the proximal tubule. In 2017, she switched sectors and joined PharmGenetix Gmbh, a private laboratory providing pharmacogenetic analyses and reporting services, as their Chief Science Officer.

Solving the CYP2D6 Incidentalome

Because a relatively high percentage (~25%) of clinically prescribed drugs are substrates for *CYP2D6*, accurate determination of its genotype for phenotype prediction is essential. However, this is challenging due to its inherent genetic variation, copy number variation (duplications and deletions) and hybrid formation with highly homologous pseudogenes. 137 distinct CYP2D6 alleles have been reported to date, but the functional consequence of 55% of these are either unknown, uncertain or not yet assigned. We developed an assay in a live human cell system to inform on the functional relevance of these CYP2D6 alleles.

JESSE SWEN

Jesse Swen PharmD, PhD is an associate professor of pharmacogenetics and clinical pharmacist-clinical pharmacologist at the Department of Clinical Pharmacy & Toxicology, Leiden University Medical Center. He is the chair of the laboratory of the hospital pharmacy.

The long-term central goal of my career is to improve the outcomes of drug treatment by gaining a better understanding of the genetic mechanisms that result in inter-individual variability in drug response. Our limited ability to predict drug pharmacokinetics, response, and toxicity presents a significant challenge that needs to be resolved. I aim to gain novel mechanistic insights into the origins of this limited ability by working on the following research questions:

- Dissecting the impact of rare, structural and regulatory variants on drug absorption, distribution, metabolism, and excretion (ADME)
- Elucidating the mismatch between the drug metabolizer genotype and the capacity of an individual to metabolize drugs (Phenoconversion).
- Translating pharmacogenomics to patient care.

This work is seamlessly integrated with his work as chair of the pharmacy laboratory.

I'm one of the primary investigators of the "Ubiquitous Pharmacogenomics" project (www.upgx.eu). U-PGx aims to implement pharmacogenetics across 7 European sites by genotyping 8,100 patients. In addition I am an active member of the Dutch Pharmacogenetics Working Group and the US Clinical Pharmacogenetics Implementation Consortium (CPIC).

Challenges and opportunities for pre-emptive pharmacogenetic testing

Retrospective, prospective and naturalistic studies all provide compelling evidence that genetic variation affects the way people respond to drugs. While our current understanding of how genetic variation influences drug response is already used to guide drug treatment in clinical practice, it is still limited in many ways. In this presentation I will highlight the current implementation of pharmacogenetics in the Netherlands and outline challenges and opportunities for the field with examples from our clinical practice and clinical research.

ARNE MERILAI

Arne Merilai (1961) is Professor and Chair of Estonian Literature (2011) at the University of Tartu. His primary area of research is national literature in the context of world literature informed by a comparative poetics and unified field literary theory. He has developed an innovative approach called pragmapoetics, a philosophy of poetic language usage. With more than three and a half hundred publications to his credit, Arne Merilai ranks among the most prolific scholars in Estonia.

His book Eesti ballaad 1900–1940 (The Estonian Ballad, 1900–1940) was published in 1991, and Eesti pagulaskirjandus 1944–1992: Luule (Estonian Literature in Exile 1944–1992: Poetry, with Õnne Kepp) in 1994. Eesti ballaad: Antoloogia: XVII–XX sajand (Estonian Ballad: Anthology: 17th–20th Century), Poeetika: Gümnaasiumiopik (Poetics: Textbook for High School, with Anneli Saro and Epp Annus), and Pragmapoeetika: Kahe konteksti teooria (Pragmapoetics: The Theory of Two Contexts) were released in 2003. Besides editing the joint research volumes, he has published two books of his own selected articles, Vokimeister: Kriitilisi konstruktsioone 1990–2011 (The Spinning Wheel Maker: Critical Constructs 1990–2011, 2011), and Õnne tähendus: Kriitilisi emotsioone 1990–2010 (The Meaning of Happiness: Critical Emotions 1990–2010, 2011). A member of the Estonian Writers' Union, he has penned two collections of poetry, Merlini aare (Merlin's Treasure, 1998) and Tolmutort (Dust Bunny, 2001), a play (2006), and a novel Türann Oidipus (Oedipus the Tyrant, 2009). In culmination of the 100th anniversary celebrations of the Estonian National University (2019), a large-scale outdoor video mapping was performed according to his script on Tartu's Toome Hill Park, the Spirit of Tartu (published in 2021).

Poetics Is in Genes

The article “Poetics Is in Genes” reveals the commonality between poetics and genetics for the first time. Thus far, outside of cellular biology, attempts have been made from both (text)linguistics and semiotics to describe the genome and its interactions similarly to language. However, the approach in this manifesto relies particularly on the poetic function of language and its underlying self-referentiality as its starting point. Poetic relevance reveals itself explicitly in its relation to the cutting-edge concept of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), which thematizes abundant metrical and figurative phenomena and terms on several levels: accumulation, regularity, interval, different repetitions, rhythm; equivalency, substitution, connotation; synecdoche, metonymy, metaphor, irony, implicature, paradox; palindrome, chiasmus, ellipsis, zeugma, calembour, polysyndeton; verses, stanzas, chapters, refrains, (identical) rhymes, collage, plot, poem, composition, text, hypertext, architext, orchestration; graphic imagery, symmetry – asymmetry; homonyms, synonyms, antonyms, archaisms, neologisms; words, phrases, sentences, syntax, definitions, quotes, palimpsest; cacophony, noise, harmony; self-reflexivity of the utterance and utterer. From this perspective, life stems from primordial poetics as the latter’s first level. It is a convincing enough association to apply poetic analysis to the free interpretation process of genomes by cells. A universal law of nature is that symmetry dictates design (including asymmetry): poetics is everywhere.