THE GENOTYPE TISSUE EXPRESSION (GTEX) PROJECT COMMUNITY MEETING
ENHANCING THE USAGE OF HUMAN GENOMICS FOR THE BENEFIT OF ALL

01 WELCOME

02 PROGRAM AND POSTER LIST

03 ORGANIZING COMMITTEE

04 ABSTRACTS

05 PRACTICAL INFORMATION AND OUTCOMES

06 ORGANIZERS AND COLLABORATORS
Dear Participants,

On behalf of the meeting organizers, we would like to welcome you and thank you very much for your participation in the Genotype Tissue Expression (GTEx) Project Community Meeting (Enhancing the Usage of Human Genomics for the benefit of all). We hope that the meeting will cover your expectations and we wish you a pleasant stay in Barcelona.

The Genotype-Tissue Expression project (GTEx) has created a comprehensive public atlas of gene expression and regulation across multiple human tissues. The resource is already providing valuable insights into the tissue specific mechanisms of gene regulation, enabling studies of expression quantitative trait loci (eQTLs), and alternative splicing, and enhancing the interpretation of genome-wide association studies and disease mechanisms.

The GTEx project has completed sample collection from 960 donors, and is QC’ing a final data and analysis release that will extend the RNA sequence data set to around 18-20,000 tissues. Whole Exome and Whole Genome sequence data will also be available for all donors.

The GTEx Community meeting is held annually and is open to the scientific community who are interested in learning more about the project and data, or who are already using the data in their research. Prior meetings have been held in the United States at the Broad Institute (2014), the University of Chicago (2015), and at Stanford University (2016). This year, for the first time the meeting is being held in Europe, in Barcelona, in conjunction with the completion of the main production phase of GTEx data. We are hoping to convene a community of scientists from all over the world that are using, or that plan to benefit from using, the GTEx data.

The meeting highlights will include donor and sample collection, current data sets and data types available, analysis tools being developed for these data types, and results from applying them to the data. The relationship of the project to large scale projects and data sets, such as ENCODE, GWAS, BluePrint, FANTOM and other Cancer projects will be highlighted, and we have a wide variety of talks and posters that showcase the range of research from the broad scientific community who are using the GTEx data.

The Organizing Committee
PROGRAM

Thursday, April 20th, 2017

8:30 Registration

9:00 Opening Welcome

Montserrat Capdevila, Head of Research, Area of Research and Knowledge, la Caixa Foundation
Jordi Fàbrega, Director of Business Development, Biocat
Roderic Guigó, Professor, Centre for Genomic Regulation (CRG)
Kristin Ardlie, PI, Genotype Tissue Expression (GTEx) Project LDACC, Broad Institute

9:15 SESSION 1
Chair: Roderic Guigó, Centre for Genomic Regulation (CRG), Barcelona, Spain

The GTEx resource: Project overview, data types and production updates
Kristin Ardlie, Broad Institute, Cambridge, MA, USA

Developing a Biospecimen Collection Infrastructure for GTEx and Resulting Public Resources
Helen Moore and Abhi Rao, National Institutes of Health, Bethesda, MD, USA

Process of Postmortem GTEx Tissue Collection by RPCI
Barbara Foster, Roswell Park Cancer Institute, Buffalo, NY, USA

Functional profile of the pre- to post-mortem transition in blood
Joaquin Dopazo, Centro de Investigación Príncipe Felipe, Valencia, Spain

10:30 Morning coffee break

11:00 SESSION 2
Chair: Kristin Ardlie, Broad Institute, Cambridge, MA, USA

The molecular basis of the cellular taxonomy of the human body
Alessandra Breschi, Centre for Genomic Regulation (CRG), Barcelona, Spain

Associating gene expression to GTEx histopathology image features using neural networks
William Jones, Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Intersecting pathology images and gene expression data to understand drivers of complex phenotypes
Barbara Engelhardt, Princeton University, Princeton, NJ, USA

Use of the GTEx resource in the deconvolution of the normal human skin transcriptome to identify specific cell types and their relative abundance
Barbara Shih, University of Edinburgh, Edinburgh, United Kingdom

Derivation of a microglial transcriptional signature by network deconvolution of human central nervous system expression data
Anirudh Patir, Roslin Institute, Midlothian, United Kingdom

12:40 Lunch
**14:00**  SESSION 3  
Chair: **Barbara Engelhardt**, Princeton University, Princeton, NJ, USA  
Differential proportionality: a normalization-free approach to differential expression  
**Ionas Erb**, Centre for Genomic Regulation (CRG), Barcelona, Spain  
QRank: A novel quantile regression tool for eQTL discovery  
**Iuliana Ionita-Laza**, Columbia University, New York, NY, USA  
Identifying causal variants to understand the properties of regulatory variants and their relationship with disease  
**Andrew Brown**, University of Geneva, Geneva, Switzerland  
Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis  
**Lin Chen**, University of Chicago, Chicago, IL, USA

**15:20**  Afternoon coffee break

**15:40**  SESSION 4  
Chair: **Christopher Brown**, University of Pennsylvania, Philadelphia, PA, USA  
Lessons Learned From Building Personalized Phased Diploid Genomes of the EN-TEX Samples  
**Thomas Gingeras**, Cold Spring Harbor Laboratory, New York, NY, USA  
Haplotypic epistasis shapes human genetic variation and disease risk  
**Tuuli Lappalainen**, New York Genome Center and Columbia University, New York, NY, USA  
Population-specific imputation of gene expression improves prediction of pharmacogenomic traits for African Americans  
**Assaf Gottlieb**, University of Texas Health Center in Houston, Houston, TX, USA  
Integrating tissue specific mechanisms into GWAS summary results  
**Hae Kyung Im**, University of Chicago, Chicago, IL, USA

**17:00**  Poster session and reception

**20:00**  Conference dinner*

*Conference dinner will take place at Agüelo013 Restaurant (C/ Avinyó, 37, Barcelona)
Friday, April 21st, 2017

8:45  Registration

9:00  Welcome
  Roderic Guigó, Centre for Genomic Regulation (CRG)
  Kristin Ardlie, Broad Institute

9:05  SESSION 5
  Chair: Emmanouil Dermitzakis, University of Geneva, Geneva, Switzerland
  
  Qualitative and Quantitative Profiling of Human Tissue Proteins
  Hua Tang, Stanford University, Stanford, CA, USA

  Integrated omics for mapping of the human proteome
  Cecilia Lindskog, Uppsala Universitet, Uppsala, Sweden

  Using GTEx and high throughput proteomics data to study the relationship
  between the transcriptome and proteome in human tissues
  Alvis Brazma, EMBL European Bioinformatics Institute, Cambridge, United Kingdom

  Gene Expression patterns: from enhancer regulation to proteomics
  Simon Fishilevich, Weizmann Institute of Science, Rehovot, Israel

  Blood-based gene panel for the diagnosis of coronary artery disease
  Joao Curado, Centre for Genomic Regulation (CRG), Barcelona, Spain

10:45  Morning coffee break

11:10  SESSION 6
  Chair: Tuuli Lappalainen, New York Genome Center and Columbia University, New York, NY, USA

  Contribution of non-coding DNA to complex traits and cancer
  Emmanouil Dermitzakis, University of Geneva, Geneva, Switzerland

  The genetic regulatory landscape of the human pancreatic islet transcriptome
  Ana Viñuela, University of Geneva, Geneva, Switzerland

  Statistical and experimental methods for causal inference at complex trait
  associated loci
  Christopher Brown, University of Pennsylvania, Philadelphia, PA, USA

  Integrating eQTLs across multiple human tissues with genome-wide association
  studies uncovers new associations and causal genes for complex diseases
  Ayellet Segré, Broad Institute, Cambridge, MA, USA

12:30  Lunch

13:40  SESSION 7
  Chair: Roderic Guigó, Centre for Genomic Regulation (CRG), Barcelona, Spain

  The landscape of human sex-differential transcriptome and its consequent
  selection
  Shmuel Pietrokovski, Weizmann Institute of Science, Rehovot, Israel

  Sex specific effects of long noncoding RNA across human tissues
  Anthony Payne, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom

  Functional impact of polymorphic inversions in the human genome
  Jon Lerga-Jaso, Universitat Autònoma de Barcelona, Barcelona, Spain
Metabolic Network Analysis for Understanding the Biology of Ageing
Ravi Sudharshan, Swiss Institute of Bioinformatics, Lausanne, Switzerland

Open Debate
GTEx: where might we go next?
Manolis Dermitsakis and Kristin Ardlie

15:55 Wrap up and closing
16:00 Adjourn

NOTE: Coffee breaks and lunches will take place at the Auditorium Terrace, Floor 1
1. The GTEx data and analysis portal (live demonstration).
   Jared Nedzel.

2. Prenatal and postnatal environmental tobacco smoking and molecular signatures in children.

   Victor Moreno and Ferran Moratalla.

4. The Transcriptional Landscape of Repetitive Elements in Human Tissues.
   Gireesh K. Bogu, Ferran Reverter, Marc A. Marti-Renom, and Roderic Guigó.

5. Characterization of selenoprotein gene expression across tissues and individuals.
   Aida Ripoll, Didac Santesmasses, and Roderic Guigó.


7. Correlating in vitro cell physiology to in vivo tissue homeostasis to advance regenerative medicine.
   Aysegul Dede and Jan de Boer.

8. A Human Atlas of Tissue Specific Expression based on GTEx.
   Klas Hatje, Gregor Sturm, Nikolaos Berntenis, and Laura Badi.

   Anne Ndungu, Martijn van de Bunt, The GTEx Consortium, and Mark McCarthy.

10. A Bayesian method for multivariate meta-analysis in GWAS.
    Saikat Banerjee, Lingyao Zeng, Heribert Schunkert, and Johannes Söding.


    Shmuel Pietrokovski and Moran Gershoni.

13. Reproducibility of co-expression networks across tissues
    Alejandro Caceres and Juan R Gonzalez

    Sergio Picart-Armada, Alfonso Bull, and Alexandre Perera-Lluna.

15. Effects of post-mortem interval on gene expression across several tissues.
    Pedro Ferreira, Manuel Munoz, Ferran Reverter, Kristin Ardlie, and Roderic Guigó.


19. GTEx-based gene expression features are a key component of machine learning models for the characterization of genes associated with neurological diseases. Juan Botía, Sebastian Guelfi, Karishma D’Sa, Jana Vandrovcova, John A Hardy, Michael Weale, and Mina Ryten.

20. The role of non-coding variants in prostate cancer. Katherine Hartmann.

21. Leveraging the GTEx histological data: phenotype extraction. Manuel Muñoz Aguirre and Roderic Guigó.

22. Dominant isoform switch events across tissues and in tumours. Nuno A. Fonseca, Sergio Santos, and Alvis Brazma.

23. A GTEx re-annotation to identify high-quality healthy samples, leveraged in innovative transcriptomics analyses. Frederic Bastian, Anne Niknejad, Amina Echchiki, Julien Roux, Bgee Team, and Marc Robinson-Rechavi.


27. The emerging, positive role of somatic retrotransposition in cell identity, phenotype variation and reprogramming. Valerio Orlando and Piero Carninci.


29. The Ensembl Cis-Regulatory Annotation. Daniel Zerbino.
ORGANIZING COMMITTEE

Kristin Ardlie, PI of the Genotype Tissue Expression (GTEx) Project's, Laboratory, Data Analysis, and Coordinating Center (LDACC) at the Broad Institute, Cambridge, Massachusetts, USA.

Dr. Ardlie works at the Broad Institute where she is the PI of the Genotype Tissue Expression (GTEx) Project's, Laboratory, Data Analysis, and Coordinating Center (LDACC). She oversees a team responsible for the generation, QC, and analysis of RNA sequence data from up to ~20,000 tissues, from 960 donors with whole genome sequence data, with a goal towards characterizing the relationship between genetic variation and the regulation of gene expression across multiple human tissues. Kristin received a PhD from Princeton University, completed a fellowship with the Harvard Society of Fellows, and then worked at the Whitehead Institute/MIT Center for Genome Research, and at Genomics Collaborative Inc.

Roderic Guigó is a Professor at the Universitat Pompeu Fabra (UPF) and the Chair of the Bioinformatics and Genomics Program at the Centre for Genomic Regulation (CRG), Barcelona, Spain.

Roderic Guigó is a Professor at the Universitat Pompeu Fabra (UPF) and the Chair of the Bioinformatics and Genomics Program at the Centre for Genomic Regulation (CRG). He participated in the human genome project, as well as in many other genomic and functional genomics projects, such as ENCODE, GTEx and BluePrint. He chairs with Kristin Ardlie, the Gtx Transctriptome Analysis Working group. His main research interests are in the understanding and modeling the regulated production of RNA in eukaryote cells. Roderic obtained his PhD from University of Barcelona and did post-doctoral research in Computational Genomics at Harvard and Boston Universities and Los Alamos National Laboratory in New Mexico.

Casandra Trowbridge, Senior Project Coordinator for the Genotype-Tissue Expression (GTEx) Project's Laboratory, Data Analysis, and Coordinating Center (LDACC) at the Broad Institute, Cambridge, Massachusetts, USA.

Romina Garrido, Assistant/Administrative Associate at the Centre for Genomic Regulation (CRG), Barcelona, Spain.

Marta Soler, Head of Research and Scientific Debate, Biocat, Barcelona, Spain.
We would like to acknowledge and thank the abstract review committee for their time and input:

Christopher Brown. University of Pennsylvania, Philadelphia, PA, USA.
Tuuli Lappalainen. New York Genome Center & Columbia University, New York, NY, USA.
Francois Aguet. Broad Institute, Cambridge, MA, USA.
Su Koester. National Institutes of Health, Bethesda, MD, USA.
Abhi Rao. National Institutes of Health, Bethesda, MD, USA.
Simona Volpi. National Institutes of Health, Bethesda, MD, USA.
Kristin Ardlie. Broad Institute, Cambridge, MA, USA.
Roderic Guigó. Centre for Genomic Regulation (CRG), Barcelona, Spain.
ABSTRACTS - TALKS

Thursday, April 20th, 2017

Session 1

Roderic Guigó, Group Leader at Centre for Genomic Regulation (CRG), Barcelona, Spain.

Chair of the SESSION 1

Kristin Ardlie, Dr. Ardlie leads the Laboratory, Data Analysis, and Coordinating Center (LDACC) of the NIH Common Fund’s Genotype-Tissue Expression (GTEx) Project at the Broad Institute, Cambridge, MA, USA.

The GTEx resource: Project overview, data types and production updates

Helen Moore with Abhi Rao, Branch Chief at National Institutes of Health, Bethesda, MD, USA.

Developing a Biospecimen Collection Infrastructure for GTEx and Resulting Public Resources

Genotype-Tissue Expression (GTEx), a National Institutes of Health Common Fund program, aims to determine how genetic variation influences normal gene expression in multiple human tissues. The project required collection of approximately 40 different tissue types from nearly 1000 postmortem donors. An RNA integrity number (RIN) of 6 or greater was required for each tissue for optimal downstream molecular analyses. The Biorepositories and Biospecimen Research Branch of the National Cancer Institute organized a large-scale biospecimen collection infrastructure to procure tens of thousands of quality assured, non-diseased tissues. The multidisciplinary and multi-site team for biospecimen collection included a bioethics team, tissue collection partners, a biobank and a brain bank, bioinformatics experts and software engineers, pathologists, biologists, program directors and managers, researchers, and an expert panel. The program has made publicly available several components of this model collection infrastructure: Standard Operating Procedures (SOPs) detailing collection procedures; the Comprehensive Data Resource, a custom-built distributed system to enable annotation and analysis of biospecimen data and associated clinical data; and the vocabulary system developed and utilized for the program. Additional GTEx resources available to the public include the GTEx histological image library that can be downloaded and viewed with zooming capability. Data from genomic analyses from each donor is available in the database of Genotypes and Phenotypes (dbGaP) at NIH. Additional information imported into dbGaP includes clinical, genotype, and gene expression data, pathology review, and data from expression quantitative trait loci (eQTLs). The data from GTEx are already being used in numerous projects to better understand predisposition to disease, with several high profile studies already published. The breadth of the program has afforded multiple, varied public resources from molecular analyses datasets to SOPs to banked tissues and cell lines. Researchers are encouraged to request access to utilize residual GTEx biospecimens for research purposes.

Barbara Foster, Professor of Oncology at the Roswell Park Cancer Institute, Buffalo, NY, USA.

Process of Postmortem GTEx Tissue Collection by RPCI

The procedure for donor identification, authorization for donation of tissues and collection of tissues for the GTEx project is a complex process that is driven by standard operating procedures (SOP) and Quality. Here we describe the processes implemented by the Roswell Park Cancer Institute (RPCI) Biospecimen Source Site (BSS) for the collection of ‘normal’ tissues from deceased donors for the GTEx Project. RPCI Rapid Tissue Acquisition Program (RTAP) has a partnership with the local Organ Procurement Organization, Upstate New York Transplant Services (UNYTS), to collect tissue for research purposes that cannot be obtained through standard medical procedures. For the GTEx project, RPCI utilized existing infrastructure for collection of tissues and organs for transplant to collect multiple ‘normal’ tissues from a single donor and the associated donor information.
Functional profile of the pre- to post-mortem transition in blood

Recently, efficient realistic models of signal propagation have been successfully applied to the study of signaling pathways in cancer (1). These models do not consider whole pathway scores, but rather focus on the estimation of the activity of specific canonical signaling circuits that trigger particular cell functions. Using the propagation model we have transformed primary GTEx gene expression values in blood, corresponding to pre- and post-mortem measurements, into vectors of signaling circuit activities (and cell functional activities) for each individual. Then, we applied conventional statistical tests to compare pre and post mortem activities and detected cell functions with significant changes of activity.

Firstly, several processes clearly related to "blood death" are clearly detectable among the most significant activations (e.g. Blood coagulation, Fibrinolysis, Plasminogen activation, etc.). There is a remarkable response to stress, along with the detection of DNA damage and the activation of the corresponding repair machinery. There is also a generalized deactivation of cell immunity processes. The metabolism is also affected with different functions deactivated, probably due to the post-mortem hypoxia. Probably, as a consequence of the above, there is a generalized activation of different processes leading to cell growth arrest.

As an example, three circuits from four different pathways (NF-kappa B, cAMP and HIF-1) trigger the process hierarchy (according Uniprot) of Blood coagulation and the Plasminogen activation. Collaterally, these circuits also activate two more functions: Glycolysis and Angiogenesis. Circuit activity ultimately account for the molecular mechanisms behind the functional alterations detected in the transition pre- to post-mortem blood.


Session 2

Kristin Ardlie, Dr. Ardlie leads the Laboratory, Data Analysis, and Coordinating Center (LDACC) of the NIH Common Fund’s Genotype-Tissue Expression (GTEx) Project at the Broad Institute, Cambridge, MA, USA.

Chair of the SESSION 2

Alessandra Breschi, Postdoc, Centre for Genomic Regulation (CRG), Barcelona, Spain.

The molecular basis of the cellular taxonomy of the human body

Organs and tissues are complex structures composed of millions of cells of diverse morphology and function. Tissue transcriptomes, thus, represent the average behaviour of genes across a highly heterogeneous collection of primary cells. How the transcriptional profiles of each these cells types relate to the profiles of tissues and organs is still poorly understood. Here we have monitored by RNA-Seq the transcriptome of multiple primary cells from multiple tissues. We found that their transcriptional profiles cluster into a few broad cell types: endothelial, epithelial and mesenchymal. The clustering is recapitulated using independent transcriptomic and epigenomic data, which also shows that blood cells, as well as other specialized cells, cluster separately. The influence of tissue of origin contributes very little (<4%) to the transcriptomes of primary cells. Regulation of transcription amounts plays the main role in defining these broad cell types, compared to post-transcriptional regulation (splicing), which plays a comparatively more important role in refining the characteristic transcriptomes of primary cells within a given type. We identified about 3000 genes specific to the these three cell types. These include a core set of transcription factors (TFs), showing strong co-expression patterns, and thus likely candidates to drive cell type specificity. Cell type specific genes are enriched for motifs of TFs specific to the same cell type. Cell type specific genes are mostly a vertebrate innovation, appearing early in the evolution of this lineage, with epithelial specific genes being the most evolutionary dynamic. We employed a method to estimate the proportion of cells from each type in a given tissue from gene expression values in that tissue in the GTEx collection of tissue transcriptomes. We found that, although we characterized a very limited number of primary cells from a limited number of tissues, the three basic cell types capture a large proportion (>70%) of cellular composition of many human tissues. Through the analysis of the GTEx catalogue of histological images and the associated annotations, we show that our inferred cellular composition precisely defines tissue type and captures morphological heterogeneity in the tissue samples. We identified changes in the cell type composition occurring with age and sex in a few tissues. We found that departures from the normal cellular composition of tissues correlate with histological phenotypes associated to diseases. Alterations of the cellular composition are particularly relevant in cancer, where they can be even associated to different stages of disease progression. The collection of primary cells transcriptomes produced here is a unique resource to understand tissue biology, serving as interface between tissue and single cell transcriptomics.
William Jones, PhD Student at Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

**Associating gene expression to GTEx histopathology image features using neural networks**

Genetic association studies require well-defined phenotypes. Constructing precise features is difficult in image data, so what is the best approach when analysing the GTEx high-resolution histopathology images and bulk gene expression data? We explore this question by exploiting neural networks, known for their ability to construct high-level discriminatory features between classes. Specifically, we modified InceptionNet-v3, a 220-layer Convolutional Neural Network pre-trained to distinguish everyday objects in images, to classify tissue types from image patches. To do so, we added 1024 fully-connected neurons to the final layer, and fine-tuned the network on 50 image patches from 100 images of 10 tissues. We vary the size of the sampled patches, rescaling all patches to be 299x299 pixels. We achieved 81% accuracy on held-out test data when using 128x128 patches in training and classification, and 94% when using 4096x4096 image patches, demonstrating that neural networks trained on general tasks can be repurposed for accurate biological image analysis.

Next, we asked whether the neural network extracted high-level features. We used the average 1024 final layer neuron outputs calculated from all non-overlapping 128x128 pixel patches from each lung image excluding the background as quantitative traits, and tested for their association with gene expression levels using the linear mixed model framework implemented in Limix. We report associations between individual neurons comprising this final representation and levels of individual gene transcripts. For example, transcripts of genes SLC34A2, SYT15, and ZBTB42 are statistically significantly associated to the activation of the first component p-values 0.000698, 0.0199, 0.0224 (Benjamini-Hochberg adjusted, alpha=0.05). These genes are enriched for pulmonary alveolar microlithiasis, consistent with their role in lung function, and are therefore plausible candidates. We are currently visualising the deep neurons in the trained neural network to understand which image characteristics help distinguish between tissues, and reflect gene expression state.

Barbara Engelhardt, Assistant Professor at Princeton University, Princeton, NJ, USA.

**Intersecting pathology images and gene expression data to understand drivers of complex phenotypes**

Understanding the correlations between genotype, gene expression levels, and high dimensional complex traits has been essential to studying the drivers of human complex disease and identifying effective therapeutic strategies for these traits. However, some complex traits are difficult to characterize in such a way as to make the quantification of correlations possible; one such complex trait is pathology imaging data. In this work, we use a type of deep learning, a convolutional autoencoder, to automatically extract one thousand features from each pathology image, and we use sparse canonical correlation analysis to correlate these pathology images with paired gene expression data on the same samples. Across three data sets, including two cancer tissue data sets and the GTEx data that include paired pathology imaging data and gene expression data, we find that our approach identifies the subset of genes that are differentially expressed with respect to specific image features, including cell size, extracellular matrix organization, cell wall thickness, and cell shape. We also pursue genotype association with pathology features in the GTEx data. We validate these associated genes and genotypes correlated with pathology image features using various approaches including gene ontology enrichment, tissue specific expression, and Mendelian randomization, allowing us to identify the drivers of cellular phenotypes. This work begins to explore the possibility of association mapping with phenotype data automatically derived from images.

Barbara Shih, Research Fellow of the University of Edinburgh, Edinburgh, United Kingdom.

**Use of the GTEx resource in the deconvolution of the normal human skin transcriptome to identify specific cell types and their relative abundance**

Skin has a complex anatomical architecture. At a gross level it is formed from the epidermis, dermis and subcutaneous tissue, however, there are also numerous fine structures and cell types within these layers. Genes expressed by specific cell types or pathways are frequently regulated by similar factors. Across many samples, these genes may be observed to be co-expressed, their level of expression rising and falling concurrently. We use have used a network-based coexpression approach to 'deconvolute' genes expressed by different skin cell types and processes.

Two large normal skin transcriptomic datasets were utilized, one microarray (GSE13355 and GSE30999, n=165), the other RNA-seq data (GTEx V6, n=578). Following quality control and batch correction, correlation network analysis was used to identify co-expression clusters. Annotation of the gene clusters was carried out using a combination of bioinformatics analyses, literature, expert review, as well as by similarity to previously defined co-expression modules. From a combined analysis on the two datasets, 20 gene signatures were defined. These include expression signatures for hair follicles, glands (sebaceous, sweat, apocrine), keratinocytes, melanocytes, endothelia, muscle, adipocytes, immune cells, and a number of pathway systems. We have named this resource SkinSig. SkinSig was then employed in the analysis of transcriptomic datasets derived from 18 skin conditions and aging, providing in-context interpretation of these data. For instance, where conventional analyses have shown a significant decrease in keratinization and fatty acid metabolism in aging skin, SkinSig more accurately pinpointed these changes to reduced numbers or activity in hair follicles and sebaceous glands. SkinSig also highlighted alterations in the relative cellular abundance that occurs in samples derived from disorders of the skin.

15
As well as proposing a collection of new marker genes for skin cell types, SkinSig facilitates in-context interpretation of skin transcriptomic data.

**Anirudh Patir**, PhD Student at [Roslin Institute](https://www.roslin.ed.ac.uk), Midlothian, United Kingdom.

**Derivation of a microglial transcriptional signature by network deconvolution of human central nervous system expression data**

Microglia are the macrophages of the central nervous system (CNS) with both immune and homeostatic functions, playing a central role in development and neurodegenerative diseases. However, we still know very little about what makes microglia different from other resident tissue macrophages or indeed the functional significance of their region-specific differences recently demonstrated in mice, and there is still much to understand about their role in health and disease. Here, we describe the results of our preliminary investigations into defining the transcriptomic profile of human microglia. Initially, we used the GTEx project microarray data generated from different regions of the human CNS to deconvolute the expression signature of microglia. After stringent quality control, a correlation network data was constructed using data from 143 samples. At a threshold of \( r = 0.7 \) the network contained 8,944 nodes (genes), connected by 372,198 edges. Cluster analysis identified a module of 457 genes highly enriched in known microglial genes. To verify and refine the microglial-specificity of the genes in this cluster, it was compared with microglia signatures from eleven published studies, including seven from mouse and five from human, which in themselves showed surprisingly little overlap with one another. Of the genes identified, 220 were present in one or more of the other gene lists. We are now attempting to further validate this signature by its comparison to other microglial signatures derived in a similar manner from independent datasets, including microarray data from the Allen Brain Atlas (\( n = 3,702 \)) and GTEx RNA-seq data (\( n = 1137 \)). Our aim is to create and annotate a robust microglial signature to help in defining the function of microglia across different regions of the human CNS, and in future studies to use this information to better explore their role in neurodegenerative disease.

**Session 3**

**Barbara Engelhardt**, Assistant Professor at [Princeton University](https://www.princeton.edu), Princeton, NJ, USA.

Chair of the [SESSION 3](#)

**Ionas Erb**, Postdoc at [Centre for Genomic Regulation (CRG)](https://www.crg.eu), Barcelona, Spain.

**Differential proportionality: a normalization-free approach to differential expression**

Expression data from projects like GTEx are vulnerable to artefacts coming from slight changes in data production among laboratories. Even when batch effects are negligible, sample normalization between different tissues relies on general assumptions that are rarely met entirely. Here we propose an analysis method where sample normalization is unnecessary but important insights can be obtained nevertheless. Instead of trying to recover absolute mRNA abundances, our methodology is entirely based on gene ratios, so normalization factors cancel by default. Unlike in classical compositional data analysis, we do not transform the data prior to analysis but instead consider log--ratio variance of all gene pairs separately. Small log--ratio variances (LRV) have been used to rigorously detect proportional genes in relative data before, but variance decomposition allows us to extract further information and compare proportionality between groups. We propose two definitions of differential proportionality: (1) Disjointed proportionality, where we have a difference in proportionality factors between conditions, and (2) Emergent proportionality, where genes are proportional in only one condition. In the former, the between--group variance dominates the total LRV, while in the latter, one of the group variances dominates. We apply the proposed framework to a data set from GTEx consisting of 98 samples from cerebellum and cortex, as well as to 498 samples from 16 brain tissues during human development. In doing so, we detect changes in abundance ratios of gene pairs both between tissues and between developmental time points. It strikes us how widespread and varied the phenomenon appears to be. Examples are shown where analysis based on the absolute abundance of single genes cannot detect differential expression nor correlation but analysis of their ratio shows significant differential proportionality. We believe that the approach not only circumvents technical issues but leads to new biological insights regarding the changing stoichiometry of mRNAs in development and disease.
Iuliana Ionita-Laza, Associate Professor, Columbia University, New York, NY, USA.

**QRank: A novel quantile regression tool for eQTL discovery**

Over the past decade, there has been a remarkable improvement in our understanding of the role of genetic variation in complex human diseases, especially via genome-wide association studies. However, the underlying molecular mechanisms are still poorly characterized, impeding the development of therapeutic interventions. Identifying genetic variants that influence the expression level of a gene, i.e. expression quantitative trait loci (eQTLs), can help us understand how genetic variants influence traits at the molecular level. While most eQTL studies focus on identifying mean effects on gene expression using linear regression, evidence suggests that genetic variation can impact the entire distribution of the expression level. Motivated by the potential higher order associations, several studies investigated variance eQTLs. Here we develop a Quantile Rank-score based test (QRank), which provides an easy way to identify eQTLs that are associated with the conditional quantile functions of gene expression. We have applied the proposed QRank to the Genotype-Tissue Expression (GTEx) project, and found that the proposed QRank complements the existing methods, and identifies new eQTLs with heterogeneous effects across different quantile levels. Notably, we show that the eQTLs identified by QRank but missed by linear regression are associated with greater enrichment in genome-wide significant SNPs from the GWAS catalog, and are also more likely to be tissue specific than eQTLs identified by linear regression. An R package is available on R CRAN at [https://cran.r-project.org/web/packages/QRank](https://cran.r-project.org/web/packages/QRank).

Andrew Brown, Postdoc at University of Geneva, Geneva, Switzerland.

**Identifying causal variants to understand the properties of regulatory variants and their relationship with disease**

Personalised medicine will require methods to call variants of interest from individual genomes, necessitating a deep understanding of the properties of causal variants. eQTL studies such as GTEx have identified 10,000s of regulatory variants, but understanding their properties is complicated by the fact that we do not know if we have identified the true regulatory variant or another in LD. Using RNA-seq data in 4 tissues from the Eurobats study and whole genome sequence from UK10K, we perform realistic simulations which show ~ 44% of lead eQTL variants are causal for effects on expression. Building on this, we developed the CaVEMaN method, a resampling method to estimate the probability an eQTL lead SNP is causal. We find 1,668 high confidence causal variants (probability > 0.8). Integrating open chromatin data from the Roadmap Epigenomics we observed a linear increasing relationship between the causal probability and probability of falling in open chromatin and predict 25%-76% of causal variants lie in open chromatin, depending on experiment. Integrating eQTLs with GWAS signals with RTC, we show these causal variants are highly enriched for co-segregation with GWAS signals relative to other eQTLs (16.0% sharing compared to 2.5%). We applied the method to GTEx expression data from 4 tissues to better understand tissue specific regulatory mechanisms. Integrating the GTEx v6p release with Roadmap Epigenomics, we find 24.3% of likely causal variants lie in open chromatin regions. Further work will look at identifying specific tissues where GWAS signals co-segregate with causal variants. In conclusion, we find that despite complicated outbred LD structure and high phenotype variability, it is possible to resolve causal variants in expression studies and beyond. Our results also place an upper bound on how informative ChIP-seq experiments from a single individual can be for resolving regulatory variation, which could be crucial to personalised medicine.
Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis

The impact of inherited genetic variation on gene expression in humans is well-established. The majority of known expression quantitative trait loci (eQTLs) impact expression of local genes (cis-eQTLs); more research is needed to identify effects of genetic variation on distant genes (trans-eQTLs) and understand the biological mechanisms. One common trans-eQTLs mechanism is "mediation" by a local (cis) transcript. Thus, mediation analysis can be applied to genome-wide SNP and expression data in order to identify transcripts that are "cis-mediators" of trans-eQTLs, including those "cis-hubs" involved in regulation of many trans-genes. Identifying such mediators helps us understand regulatory networks and suggests biological mechanisms underlying trans-eQTLs, both of which are relevant for understanding susceptibility to complex diseases. The multi-tissue expression data from the Genotype-Tissue Expression (GTEx) program provides a unique opportunity to study cis-mediation across human tissue types. However, the presence of complex hidden confounding effects in biological systems can make mediation analyses challenging and prone to confounding bias, particularly when conducted among diverse samples. To address this problem, we propose a new method: Genomic Mediation analysis with Adaptive Confounding adjustment (GMAC). It enables the search of a very large pool of variables, and adaptively selects potential confounding variables for each mediation test. Analyses of simulated data and GTEx data demonstrate that the adaptive selection of confounders by GMAC improves the power and precision of mediation analysis. Application of GMAC to GTEx data provides new insights into the observed patterns of cis-hubs and trans-eQTL regulation across tissue types.

Session 4

Christopher Brown, Assistant Professor at University of Pennsylvania, Philadelphia, PA, USA.

Chair of the SESSION 4

Thomas Gingeras, Professor at Cold Spring Harbor Laboratory, New York, NY, USA.

Lessons Learned From Building Personalized Phased Diploid Genomes of the EN-TEx Samples

Most current transcriptome and other functional genomics studies begin by mapping sequencing data to a standard haploid reference genome. While this approach has been very effective for measuring major trends to gene expression and regulation across individuals and tissue types, it has been largely blind to the underlying genomic differences between individuals, and especially to the phase of those variants. Consequently, the community suffers from limited power to study the role of genomic variation on such effects as allele specific binding or allele specific expression modulated by distant cis- and trans-regulatory elements. To address this critical need, as part of the ENCODE project we have sequenced the genomes of 4 human samples obtained from an ENCODE-GTEx collaboration (EN-TEx) using a large collection of genomic technologies to construct a set of personalized genomes. This includes deep coverage of Illumina short read sequencing for high quality variant identification (60X), PacBio long read sequencing for phased structural variant analysis (55x), 10X Genomics Chromium (35x) sequencing for long range variant phasing, and Hi-C short reads for chromosome-span phasing and SV analysis (100x). Using this combination of data for both alignment-based and de novo assembly techniques, we have identified millions of single nucleotide and short indel variants per genome as well as thousands of larger structural variations in each individual. We have further processed the variant and read data to establish high quality phased personalized diploid genomes, with most genes and variants fully phased at chromosome-span in each individual genome. Based on the catalog of detected variations we have further prepared individualized genome gene annotations. When combined with over 1,000 RNA-seq, ChIP-seq and other genome wide functional datasets collected from approximately 25 tissues from each donor, these new phased personal genomes provide a foundation for unprecedented exploration into the interplay between variation, tissue specific expression, and regulation with allele-specific resolution. By integrating these different functional data types in the context of an annotated personalized phased diploid genome, we identify notable effects of genetic variations on gene expression profiles that cannot be observed using the consensus human genome sequence.
Haplotype epistasis shapes human genetic variation and disease risk

Non-additive interaction between genetic variants, known as epistasis, has been hypothesized to contribute to phenotypic diversity and variant penetrance. In this work, we study haplotype epistasis, where the haplotype combination of alleles of multiple functional variants affecting the same gene modifies their functional impact. We first analyzed signs of haplotype epistasis in the general population using GTEx v6 data, complementing the eQTL data, with splicing QTLs (sQTLs) and large data set of phased rare variants from our novel method phASER. We showed by analysis of eQTL allele frequencies, allelic expression data of coding variants, and haplotypes of eQTLs and coding variants that putatively deleterious variants are depleted from the higher expressed eQTL haplotypes. Furthermore, putatively damaging variants were enriched in haplotypes that are skipped by splicing. This demonstrates that at population level purifying selection acts to reduce negative epistatic interactions. We next sought to analyze the contribution of haplotype epistasis to disease. Using the Simons Simplex Collection, we observed that in individuals with autism, rare disrupting coding variants in autism associated genes were more often found in regulatory haplotype configurations that would increase their penetrance, compared to synonymous and healthy siblings as controls. Next, we phased somatic mutations in The Cancer Genome Atlas data to their germline regulatory haplotypes and found a significant enrichment of negative haplotype epistasis in tumor suppressor genes versus control genes. These results suggest that regulatory haplotype configuration of disease-causing rare coding variants affects their penetrance via haplotype epistasis. Altogether, our results show that epistasis between variants on the same haplotype has shaped the patterns of genetic variation in humans, and that it affects penetrance of disease-causing rare variants.

Population-specific imputation of gene expression improves prediction of pharmacogenomic traits for African Americans

Genome-wide association studies (GWAS) are useful for discovering genotype-phenotype associations but are limited because they require large cohorts to identify a signal, which can be population-specific. Mapping genetic variation to genes improves power, and allows the effects of both protein coding variation as well as variation in expression to be combined into “gene level” effects. Previous work has shown that warfarin dose can be predicted using information from genetic variation that affects protein coding regions. Here, we introduce a method that improves the predicted dose by integrating tissue-specific gene expression. In particular, we leverage the Genotype-Tissue Expression (GTEx) data sets and evaluate the degree to which estimation of baseline gene expression can improve estimates of drug response. We use the gene expression imputation strategy of PrediXcan in a modified way: (1) we impute gene expression in a manner that is population-specific; and (2) we impute genes only in tissues where expression quantitative loci (eQTLs) are associated with these genes. In order to have a more interpretable model, we focus on 116 genes from the pharmacokinetic (PK) and pharmacodynamic (PD) pathways of warfarin within training and validation sets comprising both European and African-descent individuals. We build gene-tissue signatures associated with warfarin dose, and identify a signature of eleven gene-tissue pairs that significantly augment the International Warfarin Pharmacogenetics Consortium dosage-prediction algorithm in both populations. Our results demonstrate that imputed expression can improve dose prediction, in a population-specific manner.

Integrating tissue specific mechanisms into GWAS summary results

To understand the biological mechanisms underlying thousands of genetic variants robustly associated with complex traits, scalable methods that integrate GWAS and functional data generated by largescale efforts are needed. Here we propose a method termed MetaXcan that addresses this need by inferring the downstream consequences of genetically regulated components of molecular traits on complex phenotypes using summary data only. MetaXcan allows multiple causal variants and flexible multivariate models enabling the testing of a variety of complex processes under different contexts. As an example application, we trained prediction models of gene expression levels in 44 human tissues and inferred the consequences of their regulation in 40 complex phenotypes. Our examination of this broad set of human tissues revealed many novel genes and re-identified known ones with patterns of regulation in expected as well as unexpected tissues.
Friday, April 21st, 2017

Session 5

Emmanouil Dermitzakis, Professor at University of Geneva, Geneva, Switzerland.

Chair of the SESSION 2

Hua Tang, Professor, Stanford University, Stanford, CA, USA.

Qualitative and Quantitative Profiling of Human Tissue Proteins

The Enhancing GTEx (eGTEx) project aims to extend the GTEx project by generating a variety of intermediate molecular measurements on the same tissues. Using a high-through mass-spectrometry platform, we quantitatively measured protein abundance in >200 postmortem samples representing more than 20 tissues. In this talk, we will describe the pattern of proteomic variation between samples and between tissues. Integrating genomic sequencing information, preliminary analysis has identified novel variant peptides that correspond to non-synonymous variants. Along with other datasets generated by GTEx and eGTEx, this work contributes to a resource for studying how DNA variation cascade through molecular phenotypes to impact human health.

Cecilia Lindskog, Researcher, Uppsala Universitet, Uppsala, Sweden.

Integrated omics for mapping of the human proteome

Proteins are the main structural and functional constituents of every living cell, and a detailed characterization of the proteomic landscape in human tissues is crucial in order to translate the sequence information in the protein-coding genes into biological knowledge. One such initiative is the Human Protein Atlas project, mapping the human proteome using an integrated omics approach. Genome-wide mRNA expression based on RNA-seq from 37 different normal organs is complemented with RNA-data from two independent resources, including the GTEx consortium and The Functional Annotation of Mammalian Genomes 5 (FANTOM5) project. Data from the three sources is comprehensively summarized on the interactive database www.proteinatlas.org, and used for categorization of all human protein-coding genes based on expression level and tissue distribution. The independent datasets show a high degree of overlap, which supports the distribution in which almost half of the genes are expressed in all tissues and relatively few are unique for one or a few tissues. The transcriptomic analysis is combined with antibody-based protein expression profiling using standardized immunohistochemistry in 44 different normal organs and 20 types of cancer. The unique combination of transcriptomics with antibody-based proteomics allows for identification of proteins localized to certain cell types with a spatial resolution, in the context of neighboring cells. This can greatly increase our understanding of human biology, aid in elucidating the function of each protein, and provide the basis for disease-specific research and identification of potential drug target candidates. In summary, the Human Protein Atlas uses integration of large datasets from different international consortia, both on the transcriptomic and protein level, which allows researchers to explore the specific localization of all human proteins in cells, tissues and organs.

Alvis Brazma, Senior Team leader, EMBL European Bioinformatics Institute, Cambridge, United Kingdom.

Using GTEx and high throughput proteomics data to study the relationship between the transcriptome and proteome in human tissues

We studied correlation between RNA and protein abundance levels across twelve human tissues shared between three independent datasets – GTEx, Human Protein Atlas RNA dataset [Uhlen et al, 2015], and mass-spectrometry based proteomics dataset generated by the Pandey lab [Kim et al, 2014]. We show that overall the Person correlation between RNA and protein levels across the twelve common tissues for about 12,000 proteins that were able to quantify in proteome dataset is higher than been reported in most previous studies – between 0.47 (Esophagus) and 0.67 (Liver). In concordance with previous observations, we observe that the correlation between RNA and protein levels is particularly high for genes where RNA is detected above 1 FPKM. We observe that genes involved in metabolomics generally have high correlation between RNA and protein levels, while genes involved in RNA processing often have negative correlation. Some of the non-correlating genes have complex transcript structures and it is possible that their quantifications are not accurate. All three reprocessed datasets are available for download and queries from the European Bioinformatics Institute Gene Expression Atlas (www.ebi.ac.uk/gxa) [Petriszak et al, 2015].
Simon Fishilevich, PhD student, Weizmann Institute of Science, Rehovot, Israel.

Gene Expression patterns: from enhancer regulation to proteomics

The identity of different cells and tissues is governed by transcriptional regulatory programs that control the spatiotemporal patterns of gene expression. The GeneCards Suite (www.genecards.org) provides multisource information on both transcriptome and proteome expression patterns within tens of thousands of human gene web cards. A key GeneCards feature is GTEx-derived histograms of RNA expression levels in dozens of tissues. In parallel we recently introduced proteome expression data to GeneCards, derived from four mass spectrometry proteomics sources, portraying a normalized protein expression vector in 69 normal human tissues for each gene (PMID 27048349). We further calculate proteome and transcriptome differential expression, highlighting overexpressing tissues. Finally, we define GTEx-based protein-RNA expression ratios, shedding light on global aspects of expression regulation. A major challenge for understanding the regulation of gene expression is the unambiguous identification of enhancers, the distant acting regulators profusely disposed along the genome. To this end we developed GeneHancer, a novel database of human enhancers and their inferred target genes. We applied a unification redundancy-removing algorithm to 434,000 enhancers collected from four genomewide databases: the ENCODE project, the Ensembl regulatory build, FANTOM enhancer RNAs and the VISTA Enhancer Browser. We derive 285,000 unified candidate enhancers, each with a confidence score. We subsequently use five methods to infer scored enhancer target genes. Two of those methods involve GTEx data: tissue expression correlation between candidate target genes and enhancer-targeted transcription factors, and expression quantitative trait loci (eQTLs) for variants within enhancers. These results allow our VarElect tool (PMID 27357693) to interpret enhancer-contained whole genome sequencing variants in terms of disease/phenotype relations, based on high-probability enhancer target genes.

Joao Curado, Bioinformatician, Centre for Genomic Regulation (CRG), Barcelona, Spain.

Blood-based gene panel for the diagnosis of coronary artery disease

Mortality and morbidity from coronary artery disease (CAD) are a major global health burden. 4% of Catalan population suffer from CAD with 35,000 new cases diagnosed each year. These generate more than 120,000 hospitalizations and 270,000 visits to emergency departments, per year. For many patients, heart failure is the endpoint of CAD. These numbers could be improved with an earlier diagnosis: early diagnosis of CAD enables a more comprehensive management, avoidance of the transition to end-stage cardiac disease resulting in a significantly cut in the massive costs in terms of resources and lives.

In this project we leveraged our knowledge in transcriptomics and access to the GTEx data to an unmet need in clinical practice: the identification of novel biomarkers, blood-based RNA molecules, that can be used for CAD detection from a standard clinical blood sample in a non-invasive way. GTEx is the biggest public RNA-sequencing database with matched medical history to date. It includes 393 blood samples, of which 99 come from individuals with diagnosed coronary artery disease. We used this data to identify differential expressed RNA molecules (coding and non-coding; annotated or novel) that could be used as biomarkers. These potential biomarkers were categorized, characterized and ranked according to their biomarker potential and were filtered into a list of 100 potential biomarkers, used for the next steps. At the moment and in collaboration with the main Barcelona hospitals we are collecting blood samples from newly diagnosed or suspected patients where we will measure the expression of these 100 potential biomarkers by qRT-PCR. Resulting data will be used to train an algorithm that scores patients’ probability of CAD. This CAD algorithm will be validated in additional cohorts of patients evaluated for suspected CAD.

Session 6

Tuuli Lappalainen, Assistant Professor, New York Genome Center and Columbia University, New York, NY, USA

Chair of the SESSION 6

Emmanouil Dermitzakis, Professor at the University of Geneva, Geneva, Switzerland.

Contribution of non-coding DNA to complex traits and cancer

Molecular phenotypes inform us about genetic and environmental effects on cellular and tissue state. The elucidation of the genetic basis of gene expression and other cellular phenotypes is highly informative for the impact of genetic variants in the cell and the subsequent consequences in the organism. In this talk I will discuss recent advances in key areas of the analysis
of the genomics of gene expression, chromatin and cellular phenotypes in human populations and multiple tissues from the GTEx consortium and how this assists in the interpretation of regulatory networks and human disease variants. I will also discuss how these recent advances are informing us about the impact of regulatory variation in cancer.

Ana Viñuela, Postdoc at the University of Geneva, Geneva, Switzerland.

The genetic regulatory landscape of the human pancreatic islet transcriptome

Understanding the molecular properties of GWAS signals underlying common disorders requires the study of relevant cell types and tissues. However, in type 2 diabetes (T2D), obtaining enough pancreatic islets of Langerhans samples to answer these questions has been a challenge. Here we performed RNA-sequencing and genotyping in human islets from 420 cadaveric donors, and FACS sorted beta-cells in 26 of the samples to uncover genetic regulatory variants (eQTL) relevant for the tissue. We combined our samples with those provided by the GTEx consortium in 44 tissues to investigate the tissue and cell-specificity of genetic regulation and to identify more accessible tissues for the study of T2D. Through eQTL mapping we identified variants significantly affecting expression levels in cis at 6,039 genes (FDR<1%). Conditional analysis identified an additional 1,705 significant secondary eQTL signals at 1,289 genes. Each gene was assigned an islet-specificity score by comparing its expression in islets against that in 16 other tissues in GTEx. Using islet regulatory state information, we found that eQTLs for islet-specific genes were enriched in stretch enhancers compared to tissue shared eQTLs.

Integrating islet eQTL data with genetic information on 82 known T2D-associated loci identified 12 loci for T2D with at least two lines of evidence, including novel ones such as TCF7L2 and two signals at DGKB. To assess tissue-specificity of the overlap between eQTLs and T2D association signals, we compare the enrichment of T2D loci in our data with the GTEx v6p data. Across all non-islet tissues, 7 genes were linked to T2D associated loci, one of which (AP3S2) was also identified in islets. This means islet eQTLs are significantly enriched for overlap with T2D association signals (3.5-fold, p=9.2x10^{-3}) compared to all other GTEx tissues. Our results demonstrate the power of transcriptomic analysis in disease-appropriate tissues to deliver molecular insights in T2D pathophysiology.

Christopher Brown, Assistant Professor at University of Pennsylvania, Philadelphia, PA, USA

Statistical and experimental methods for causal inference at complex trait associated loci

Genome-wide association studies (GWAS) have identified thousands of loci that contribute to risk for complex diseases. The majority of the heritability of complex disease risk lies within the noncoding regions of the genome. This has led to the hypothesis that the causal variants at GWAS associated loci lead to changes in local gene expression. As a result of linkage disequilibrium and the fact that cis-regulatory elements (CREs) may target genes over large distances, it is often unclear which variant or gene affects disease risk. However, their identification will improve understanding of disease etiology and identify targets for novel therapeutic development. Recent work from efforts such as GTEx has identified genetic variation associated with gene expression variation for essentially every gene. Despite this wealth of data, the characterization of causal mechanisms at complex trait associate loci remains a significant challenge. To address this challenge, we have developed and applied high throughput computational and experimental approaches to identify candidate disease genes and the functional regulatory variants that mediate disease risk. We have focused on cardiovascular disease (CVD) and molecular trait mapping in the liver as model systems. Existing studies have focused on easily ascertained cell types, while the liver, which plays a critical role in regulating cholesterol and lipid metabolism, and where many CVD associated variants likely affect gene expression, has remained understudied. We have phenotyped liver biopsies and iPSC derived hepatocytes form more than 400 donors, collecting RNA-seq along with histone modification and transcription factor ChIP-seq data. We have used these data to identify thousands of genetic variants associated with allele-specific transcription factor binding, histone modification, gene expression, and splicing. Comparison to data from the GTEx and Roadmap Epigenomics projects demonstrate that many of these associations are specific to the liver. We demonstrate that multi-phenotype molecular trait mapping improves statistical power to detect associations and results in improved resolution at identified loci. We have integrated these data with CVD GWAS data using a novel multi-phenotype causal inference framework based on Mendelian randomization to predict the precise variants, CREs, and genes that underlie CVD risk. Using a combination of massively parallel reporter assays, genome-edited stem cells, CRISPR interference, and in vivo mouse models, we establish rs2277862-CPNE1, rs10889356-ANGPTL3, rs10889356-DOCK7, and rs10872142-FRK as causal SNP-gene sets for CVD. These results demonstrate that a molecular trait mapping framework can rapidly identify causal genes and variants contributing to complex human traits and demonstrates that, at many GWAS loci, candidate genes have been falsely implicated based on proximity to the lead SNP.

Ayellet Segrè, Computational Biologist, Broad Institute, Cambridge, MA, USA.

Integrating eQTLs across multiple human tissues with genome-wide association studies uncovers new associations and causal genes for complex diseases

Genome-wide association studies (GWAS) have lead to the discovery of thousands of common variants associated with
hundreds of complex diseases and traits. However, the causal mechanisms and genes underlying these genetic associations are largely unknown. The majority of discovered associations lie in noncoding regions, suggesting a key role for transcriptional regulation in disease etiology. Furthermore, many associations of modest effect still remain to be found. The Genotype-Tissue Expression (GTEx) whose primary goal is to identify DNA variants associated with gene expression changes (expression quantitative trait loci, eQTLs) in multiple healthy human tissues provides a unique resource to address these challenges. To this end, we developed a two-step computational method that integrates GWAS variant statistics with eQTL and pathway data to: (i) test whether eQTLs in relevant tissues significantly underlie complex disease or trait associations, (ii) detect new genetic associations of regulatory effects, and (iii) propose new causal genes and tissues of action for known and new genetic associations with complex traits. The method first tests whether eQTLs in a given tissue are enriched for modest to genome-wide significant trait associations, correcting for potential confounding effects. If enrichment is found, target genes of eQTLs with top ranked GWAS p-values are tested for enrichment in biological processes, such as signaling and metabolic pathways and mouse phenotype ontologies. We applied our method to GWAS meta-analyses of a range of complex diseases and traits (metabolic, cardiovascular, anthropometric, autoimmune and neurodegenerative), using cis-eQTLs from 44 GTEx tissues. Significant enrichment of trait associations among eQTLs was found for all traits in various tissues (p<6E-05, fold-enrichment=1.1-2.2), including expected pathogenic tissues, such as aorta artery for systolic blood pressure and hippocampus for Alzheimer’s disease. Gene set enrichment analysis of eQTL target genes in relevant tissues proposed new causal genes with regulatory effects on the tested traits, such as signaling and cardiovascular-related genes for systolic blood pressure, and cell adhesion and immune-related genes for Alzheimer’s disease. We evaluated the relative contribution of tissue-specific versus tissue-shared eQTLs to complex trait associations, and discuss the challenges of using integrative eQTL-GWAS analyses to predict pathogenic tissues given a lot of common regulation across tissues. Our work suggests that genetic regulation of gene expression significantly contributes to complex disease risk and trait variation and can be used to propose new causal genes for complex traits in relevant tissues.

**Session 7**

**Roderic Guigó**, Group Leader at Centre for Genomic Regulation (CRG), Barcelona, Spain.

Chair of the SESSION 7

**Shmuel Pietrokovski**, PI at Weizmann Institute of Science, Rehovot, Israel.

The landscape of human sex-differential transcriptome and its consequent selection

Men and women have almost identical genomes but are distinctly dimorphic, having many morphological, physiological and biochemical differences, including dissimilar disease susceptibilities. Almost all sexually dimorphic traits result from differential expression of genes present in both sexes. Such genes can thus be subject to different, and even opposing, selection constraints in the two sexes. This can impact human evolution and also lead to reduced selection on deleterious mutations that affect only one of the sexes. We present an in-depth description of human sex-differential genetic architecture and its resulting selection. Analyzing the GTEx project RNAseq data from 544 adults, we identify sex-differential expression in thousands of genes. We found both men or women biased sex differential expression in one, several, or even all 53 sampled tissues. These genes are not only expressed in the reproductive tracts, as sex-differential expression also occurs in tissues common to both sexes, including the brain, heart, and skeletal muscle. We find significant association between high sex-specificity expression and reduced selection efficiency. This enables propagation of deleterious mutations in human population that can affect the prevalence of different human diseases. In addition, the function of some of these genes is associated with several dimorphic traits, and with sex-biased diseases. Interestingly, many of the sex-specific genes that also undergo reduced selection efficiency are essential for the reproductive success of men or women. This apparent paradox might partially explain the relatively high incidence of human infertility. Our findings on the sexual transcriptome and its effects can contribute to understanding the genetic and molecular bases of human traits and disorders, specifically those with sex differential predisposition.

**Anthony Payne**, DPhil Student, Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom.

Sex specific effects of long noncoding RNA across human tissues

Long noncoding RNAs (lncRNAs) have been implicated in various sex-specific processes: the role of XIST and JPX on X chromosome inactivation is a well-known example. We assessed sex-specific differences of lncRNAs and associated processes using data from the GTEx Project.

Differential expression analysis based on sex across 42 GTEx tissues identified median 20.5 differentially expressed (DE) lncRNAs per tissue (range [4, 214]; FDR=0.01; Y chromosome excluded). The number of DE lncRNAs per tissue was correlated with sample size (r=0.738, p=2.5E-8). XIST, JPX, and RP13-36G14.4 (sense transcript within JPX intron) were DE across all tissues, while 71.6% of DE lncRNAs were DE in only one tissue.
Co-expression modules were constructed using WGCNA on all expressed genes in each tissue. Each module’s enrichment for DE IncRNAs was assessed: The Fisher combined score of differential expression p-values of IncRNAs in each module was compared to the Fisher scores of 10,000 random p-value samples. 7 modules were enriched in thyroid, 2 in amygdala, and 1 in each other tissue.

To assess potential causal influence of modules enriched for DE IncRNAs, Mendelian Randomisation (MR) was applied to module eigengenes (risk factor) and up to 118 outcomes (with ≥2 cases in a tissue). Only eigengenes without significant sex correlation were tested. One thyroid module had a significant MR effect on uremia. It contained 7 HLA genes and two DE chromosome-6 IncRNAs: HCP5 (sense overlapping MICA, which encodes a cell surface glycoprotein) and TAPSAR1 (antisense to PSMB8, a component in autoinflammation and adipocyte differentiation). The same module was enriched for 14 KEGG pathways including phagosome, cell adhesion, and various infections.

This module highlights sex DE IncRNAs in thyroid that are co-expressed with and overlap autoimmune- and cell-adhesion-associated genes. Thus, leveraging the multi-tissue nature of GTEx has allowed tissue-specific prioritisation of IncRNAs with potential sex-differential function.

**Jon Lerga-Jaso**, PhD student at **Universitat Autònoma de Barcelona**, Barcelona, Spain.

Functional impact of polymorphic inversions in the human genome

Advances in genomic techniques have generated an increasing interest in structural variants and are allowing us to characterize them at an unprecedented detail. However, there is limited information on their functional impact. This is particularly true for inversions, whose study is still challenging because of their balanced nature and the complex regions where they appear. Here we measure the effect of 45 polymorphic inversions on gene expression in lymphoblastoid cell lines from 176 individuals. Although the majority were not associated to expression differences, six inversions may influence the transcription of neighbouring genes. Moreover, we looked for potential effects in other tissues through SNPs that have been already reported as eQTLs in the GTEx project. This analysis confirmed the previous results and identified ten more candidate inversions that may affect gene expression. Interestingly, in several cases the significance of the eQTL-gene association is correlated with the linkage disequilibrium (LD) between the eQTL and the inversion, indicating that inversions could be indeed the cause of the expression changes. Finally, we compared those SNP's reported as GWAS hits to the inversions in our dataset and identified specific candidates in moderate LD ($r^2$>0.6). Few inversions also presented an enrichment of GWAS signals in their surrounding area, suggesting a potential implication in these diseases. A good example of an inversion with molecular consequences is HSINV0030, which exchanges the first exon and promoter of the chymotrypsinogen precursor genes CRB1 and CRB2. Although the inversion creates hybrid transcripts without relevant sequence modifications, we have confirmed that it affects their expression both by the mapping of GTEx-eQTLs in the inverted region and the analysis of mRNA profiles from pancreatic tissue. Thus, all together, these findings illustrate the potential functional role of inversions on the human genome and help to uncover previously missing variants related to phenotype variability.

**Ravi Sudharshan**, Doctoral Student, **Swiss Institute of Bioinformatics**, Lausanne, Switzerland.

Metabolic Network Analysis for Understanding the Biology of Ageing

Ageing is a complex process, marked by a progressive functional and physiological decline that results in increased morbidity and culminates in death. Ageing is one of the major risk factors of a plethora of human diseases such as cancer, cardiovascular and neurodegenerative diseases, thus drawing significant fraction of government spending in industrialized countries. Without a concomitant or faster increase in the length of healthy life, longer lifespan would translate to higher health care spending. In this regard, a better understanding of the biology of ageing has the potential to delay the onset of age-related diseases, extend healthy life expectancy, and thus reduce the socioeconomic burden of a greying population. The biology and mechanism of ageing is still poorly understood and the role of ageing in causing or predisposing individuals to diseases with varying pathologies remains largely unknown. The complexity of the ageing process has motivated applying a systems-oriented approach to ageing research in the last decade, involving the creation and analysis of network of biological interactions among cellular components. The close connection between ageing and metabolism is well documented, but the specific metabolic pathways that are involved in the ageing process are still hazy. In this study, we analyzed human transcriptomics data from the Genotype-Tissue Expression (GTEx) project. In particular, we employed the linear mixed effects model to determine age-related gene expression changes across different human tissues. Subsequently, we projected the age-related gene expression changes onto human genome scale metabolic model based on the constraint-based modeling of flux balance analysis. By doing so, we were able to show specific age-related alterations in metabolic pathways across different tissues. In our continuing work, we will use this knowledge to identify potential metabolic targets whose regulation may reverse the changes during ageing, and validate these targets using Caenorhabditis elegans.
1. **The GTEx data and analysis portal (live demonstration).**  
   Jared Nedzel.

2. **Prenatal and postnatal environmental tobacco smoking and molecular signatures in children.**  

Prenatal maternal smoking and early postnatal exposure to environmental tobacco smoking (ETS) are associated with a long list of adverse health outcomes in children. However, the molecular mechanisms underlying these associations are not completely understood. Current smoking in adults is associated with changes in the methylome and transcriptome of blood and epithelial lung cells, which are linked to smoking related diseases. In children, it has been shown that prenatal maternal smoking alters cord blood DNA methylation patterns, with some alterations being persistent until adolescence. The main objective of this project is to study the effect of both prenatal and early postnatal exposure to tobacco smoking on different molecular signatures in children. Two main designs will be undertaken. The first one will address tissue specificity by exploring the effect of prenatal maternal smoking on placental DNA methylation and comparing patterns with cord blood methylation. This will be done by meta-analyzing data from 1500 participants from the Pregnancy And Childhood Epigenetics (PACE) Consortium. The second study will analyze the association between both prenatal and postnatal cotinine levels (a biomarker of exposure to tobacco smoking) and the blood methylome and transcriptome (including miRNAs) of 1200 children aged 7-9y from the European HELIX project. Conditional analysis will be performed to dissect prenatal from postnatal effects, and the cross-omic signatures will be tested using multivariate analyses.

3. **Gene expression and regulation co-analysis of healthy colon tissue to explore novel mechanisms of cancer onset.**  
   Victor Moreno and Ferran Moratalla.

We aim to perform an in-depth analysis of healthy colon tissue cells with different layers of omics information to unravel novel ways of detecting proclivity of tumor development, new biomarkers of disease, or understanding the mechanisms by which they operate. For this purpose, we have used 345 GTEx RNA-seq samples from both sigmoid and transverse colon (149 and 196, respectively) to infer a gene expression network with ARACNe-AP and analyze in detail its characteristics. Moreover, GTEx SNP array data has been obtained to further study changes in gene expression network attributable to genetic variation. RNA-seq count data has been downloaded from GTEx portal (www.gtexportal.org) and after applying several filtering steps to remove lowest expressed genes in all samples, transcriptional interactions have been reconstructed between transcription factors (TF) and target genes with ARACNe-AP algorithm, applying data processing inequality method, and running 1,000 re-samplings to keep only statistically significant interactions. Once the final network will be generated, it is our goal to analyze the appearance of significant alterations in the system and identify SNPs, that could be considered responsible of these changes. Gene expression network together with SNP data analysis of GTEx samples, coupled with information on cancer mutated genes from the TCGA project could be a priceless source of information to infer drivers of tumor onset, since a large loss of transcriptional interactions have been observed in colon cancer networks, making difficult the comparison of normal and tumor networks.

4. **The Transcriptional Landscape of Repetitive Elements in Human Tissues.**  
   Gireesh K. Bogu, Ferran Reverter, Marc A. Marti-Renom, and Roderic Guigó.

More than half of the human genome contains repetitive elements and majority of their transcription is repressed. However, it is not clear how many of them are expressed and where they are expressed. To address this, as a part of GTEx project, we profiled the transcription of around 5 million repetitive elements using 8551 poly-A RNA-seq datasets from 53 tissues across 550 individuals. We report that thousands of repetitive elements originating from various repeat families are systematically transcribed across multiple human tissues in a tissue-specific manner. Especially, brain and testis show higher expression of repetitive elements compared to any other tissue. Many of the brain and testis-specific repetitive elements are from same repeat family but not subfamilies. Using a linear mixed model, on average, we found that variation in repeat expression is far greater among tissues (~57%) than among individuals (~1%). In summary, we find that repeat expression is a hallmark of tissue identity in humans. In addition, to find the association between repeat expression and the nearest gene, we applied a multivariate multilevel model that decompose variance and correlation explained by tissues and individuals.
5. **Characterization of selenoprotein gene expression across tissues and individuals.**  
*Aida Ripoll, Didac Santesmasses, and Roderic Guigó.*

Selenium (Se) is an essential trace element in humans. Its biological role is mediated by selenocysteine (Sec), the 21st proteinogenic amino acid. Sec is co-translationally incorporated into selenoproteins in response to a UGA codon, normally a stop codon, throughout a recoding mechanism that requires a dedicated machinery. The main signal for Sec insertion is a RNA secondary structure, the SECIS element, present in the 3′UTR of selenoprotein transcripts. The human genome encodes 25 selenoprotein genes, classified in 17 diverse protein families. All functionally characterized selenoproteins have oxidoreductase activity, and they serve diverse biological functions, including antioxidant defense, redox regulation, thyroid hormone metabolism, Se distribution and transport and synthesis of Sec. A large-scale characterization of selenoprotein gene expression pattern in multi-tissue transcriptome data has not previously been described. Here, we used 8555 high-quality RNA-seq samples from the Genotype-Tissue Expression (GTEx) project v6 release, covering 40 human peripheral tissues and 13 brain subregions, across individuals. Despite expression of all 25 human selenoproteins was observed in the entire set of tissues, distribution of several selenoprotein genes expression reveals a tissue-preferential expression. Thus, suggesting a regulation of the expression at tissue level. Through the analysis of the contribution of sex and age to gene expression variation, we identified two tissue-specific sex-biased genes (D12, GPx4) and a negative correlation of expression with age of certain genes (SelT, SeIT, SPS2) in most brain subregions. This study reports a characterization of human selenoprotein gene expression at an unprecedented scale which can help in understanding the biological role of selenoproteins, providing new insights into the relevance of Se in human health.

6. **Transcriptome analysis of EN-TEx data using personalized genomes.**  

The genome of every individual is different from the haploid reference commonly used for functional analysis. Results of the 1000 Genomes Projects showed that every individual carries millions of SNPs and short insertions/deletions, and thousands of structural variations, many of which affects the expression and regulation of nearby genes. At the same time our understanding of impact of this alterations is limited by using a reference genome and reference annotation without taking into account personal differences. A collaborative project between ENCODE and GTEx, the EN-Tex, aims to evaluate the impact of using personalized diploids genomes and annotations for functional analysis, in particular for analysis of gene and transcript expression across multiple conditions. In this project the genomes of 4 human samples were sequenced using combination of technologies: Illumina short reads, PacBio long reads, and 10x Genomics Chromium linked reads. For each individual, an annotated personal diploid genome was assembled using both de novo genome assembly and read mapping technologies capturing over ten megabases of variation per person from the reference. From this, we have identified genes that changed their structure and transcripts present in one allele only. Personalized files were used to perform allele specific mappings of EN-Tex long RNA-seq data from approximately 20 tissues per individual. At the same time transcriptomic data was mapped to the reference genome GRCh38 and quantified with Gencode annotation v.24. We have compared expression values obtained with these two approaches and found a high correlation between them, PCC=0.99. In average there were ~100 genes changing their expression in more than two times per sample, log2(FC)>1. In order to estimate significance of this number, we have compared gene expression changes between two randomly selected individuals from GTEx cohort, matched by age and gender.

7. **Correlating in vitro cell physiology to in vivo tissue homeostasis to advance regenerative medicine.**  
*Aysegul Dede and Jan de Boer.*

Research in our laboratory is aimed at applying basic cell biology principles in the field of regenerative medicine. We aim at translating *in vitro* cell manipulation into *in vivo* tissue regeneration. We investigate these correlations using a combination of high throughput technologies, computational modeling and experimental cell biology. For example, we previously investigated *in vitro* multipotency of human mesenchymal stem cells and found no correlation with *in vitro* ectopic bone formation. DNA microarray analysis of cells from 62 donors and comparison to their in vivo bone forming capacity showed that the CADM1 gene was highly predictive for the bone-forming capacity of hMSCs. We are currently investigating the relationship between tendon tissue physiology in vivo and tenocyte gene expression in vitro. Starting from tissue level we try to trace our steps back to understand gene regulation *in vitro* and GTEx may thus provide a different and beneficial perspective to our research.

8. **A Human Atlas of Tissue Specific Expression based on GTEx.**  
*Klas Hatje, Gregor Sturm, Nikolaos Berenten, and Laura Badi.*

We used GTEx [1] as reference atlas for gene expression in healthy human tissue to generate tissue marker gene signatures using an entropy function. These signatures were used as classifiers for several expression datasets including samples of mixed or unknown cell composition. This enabled understanding of the sources of biological variability and identification of mislabeled or contaminated tissue samples [2]. The GTEx tissue signatures were
Pancreatic islets dysfunction is central to the pathogenesis of Type 2 diabetes. Pancreatic islets play a crucial role in controlling the blood glucose levels by secreting insulin. When there is high blood glucose, pancreatic islets either undergo an adaptive response to meet the increased demands of insulin secretion or become dysfunctional. Very little is known about the molecular mechanisms underlying this adaptive response in humans. To understand the adaptive transcriptional response and underlying regulatory mechanisms occurring in a high glucose environment, we cultured a large panel of human pancreatic islets at elevated glucose concentrations and profiled their transcriptomes, open chromatin regions and several epigenetic marks. We also investigate biological processes in endocrine pancreas to be significant in the adaptive response in humans. To understand the molecular mechanisms underlying this adaptive response in humans. To understand the molecular mechanisms underlying this adaptive response in humans.

The GTEx v7 dataset greatly facilitates an in-depth analysis of transcriptomic correlates associating with type 2 diabetes (T2D) across multiple tissues, with 25% of recruited individuals having T2D and 90% of sampled tissues having more than 10 T2D cases. This is instrumental in not only identifying causal and secondary pathological processes in diabetes but can also aid in classifying relevant expression signatures serving as predictors for downstream T2D-related complications. Global patterns of gene expression were analysed using limma /voom to detect differentially expressed genes (DEGs) between T2D cases and non-diabetic controls, with PEAR factors included to account for hidden factors influencing expression. There were 255 statistically significant DEGs (FDR<0.05 per tissue) with 15 tissues having at least one DEG. Tibial artery and subcutaneous adipose had the highest number of DEGs (69 and 55 DEGs respectively) while 14 DEGs had evidence of differential expression across more than one tissue. The top four DEGs, with the largest log fold change in expression were significant only in whole pancreas (IAPP, PSPHP1, CYP3A4, ANO3). The IAPP gene, encoding an islet amyloid polypeptide, exhibited the largest log fold expression change (-1.83logFC) between cases and controls. This hormone is co-secreted with insulin from pancreatic beta cells. Next, we compared loci modulating expression (eQTLs) in three T2D-relevant tissues (whole pancreas, subcutaneous adipose and skeletal muscle). Controls were randomly down-sampled to match the number of cases in each tissue. The total number of eQTLs (FDR<0.05) detected in cases and controls was comparable in adipose and pancreas. In skeletal muscle however, controls had twice as many significant eQTLs compared to cases (1023 and 579 eQTLs respectively) with eQTLs in 399 genes shared in both and 7 genes having independent eQTLs (R2<0.2). Neither the shared nor independent eQTLs were enriched for canonical pathways or gene ontology annotations and did not overlap known GWAS loci. The 625 genes without a significant skeletal muscle eQTL in T2D cases were significantly enriched for genes in metabolic pathways (KEGG:01100, p=3.66x10^-3). These analyses provide evidence of differential gene expression and regulation in T2D. Specifically, notable differential expression in pancreatic islets, which are not separated from whole pancreas in GTEx, highlight biological processes in endocrine pancreas to be significant in the complex pathophysiology of T2D.

Meta-analyses of genome-wide association studies (GWAS) have traditionally focused on effect sizes of each genetic variant independent of all others. Joint multivariate analyses of the genetic variants, though considered more powerful, are limited by the requirement of individual level data from all cohorts. Here we describe a logistic regression approach in a Bayesian framework to create multivariate summary statistics of individual studies, subsequently combining them by meta-analysis. Our method benefits from harnessing more evidence from existing data, especially by: (1) aggregating information over many genetic variants, (2) aggregating information over many studies, and (3) distinguishing coupling from mere correlation between genetic variants and disease risk. We call our method BammGWAS (BAyesian Multivariate Meta-analysis in Genome-Wide Association Studies). BammGWAS improves precision in predicting causal loci and fine mapping causal variants in each locus.

Shmuel Pietrokovski and Moran Gershoni.

Men and women have almost identical genomes but are distinctly dimorphic, having many morphological, physiological and biochemical differences, including dissimilar disease susceptibilities. Almost all sexually dimorphic traits result from differential expression of genes present in both sexes. Such genes can thus be subject to different, and even opposing, selection constraints in the two sexes. This can impact human evolution and also lead to reduced selection on deleterious mutations that affect only one of the sexes. We present an in-depth description of human sex-differential genetic architecture and its resulting selection. Analyzing the GTEx project RNAseq data from 544 adults, we identify sex-differential expression in thousands of genes. We found both men or women biased sex differential expression in one, several, or even all 53 sampled tissues. These genes are not only expressed in the reproductive tracts, as sex-differential expression also occurs in tissues common to both sexes, including the brain, heart, and skeletal muscle. We find significant association between high sex-specificity expression and reduced selection efficiency. This enables propagation of deleterious mutations in human population that can affect the prevalence of different human diseases. In addition, the function of some of these genes is associated with several dimorphic traits, and with sex-biased diseases. Interestingly, many of the sex-specific genes that also undergo reduced selection efficiency are essential for the reproductive success of men or women. This apparent paradox might partially explain the relatively high incidence of human infertility. Our findings on the sexual transcriptome and its effects can contribute to understanding the genetic and molecular bases of human traits and disorders, specifically those with sex differential predisposition.


Alejandro Caceres and Juan R Gonzalez.

Reproducibility is a fundamental tenet of science yet perceived feeble in current biomedical research. Great effort is put into assessing the concordance between two experiments measured on the same set of individuals under controlled conditions, such as time points, exposures or tissues. However, many studies are designed to measure a population sample under a range of different conditions and, like any other study, their results are expected to be reproducible in other samples under different experimental setups. Surprisingly, there is a lack of statistical measures that assess the degree of agreement between independent studies to distinguish conditions. Take for instance the GTEx project, where gene expression data is measured in several tissues using RNA-seq, and validated with expression microarrays on the same individuals. In this study, one can infer the correlation gene network, a fundamental biological entity, for each tissue. It is therefore expected a level of reproducibility to discriminate gene networks between tissues, as derived in other studies of different individuals and experimental procedures. We propose an agreement measure on distinguishing conditions that generalizes Cohen’s kappa, in which the elements of a cross-tabulated table between conditions are pair-wise correlations between two different studies. We derive the distributional characteristics of the measure, and show how it increases monotonically with kappa while its variance allows high precision estimates of intermediate agreement. We use the measure to test the agreement to distinguish between the gene networks of four brain regions as inferred from the GTEX (RNA-seq) and BRAINEAC (microarray) projects. We find full agreement to distinguish between gene networks across tissues and fair agreement for their gene ontology enrichment status. As a conclusion, GTEx unprecedented expression data should be currently used as a benchmark to reproduce tissue specificity of gene networks obtained in independent studies.


Sergio Picart-Armanda, Alfonso Buli, and Alexandre Perera-Lluna.

The behaviour of gene expression in different tissues is a current challenge in biomedical research - relevant questions include the search for consistent, shared gene modules across tissues, the integration of networks stemming from various tissues and the estimation of gene-gene similarity matrices that take advantage of data from other tissues. To this end, the Genotype-Tissue Expression (GTEx) project is a valuable source of tissue-specific RNA-seq data. Tissue-specific networks can also be explored from a graph theory perspective. Starting from networks that have been derived using (1) the gene expression profiles in GTEx and (2) the relationship between tissues, we analyse whether proximity relationships in regulatory structure between tissues can be rediscovered through spectral signatures of their corresponding networks. We have screened 35 tissue-specific co-expression networks and described and compared their architecture and spectral properties. Additionally, we have analysed information propagation processes in these networks through diffusion kernels to gain insights of intrinsic behaviours and edge distributions in each gene network.

15. Effects of post-mortem interval on gene expression across several tissues.

Pedro Ferreira, Manuel Munoz, Ferran Reverter, Kristin Ardlie, and Roderic Guigó.

Post-mortem tissues are a valuable resource for the study of gene expression in many different diseases and disorders. Due to the unavoidable delay in the collection of tissues, post-mortem samples cannot completely escape the effects of ischemia. While DNA is known to be stable over long ischemic periods, RNA is believed to be unstable,
but this stability may vary from tissue to tissue. A systematic genome-wide tissue specific effect has not yet been fully investigated. The Genotype to Tissue Expression (GTEx) project provides a unique opportunity to study samples obtained from post-mortem donors across a wide-range of body sites. The main goal of this study is to use the collection of GTEx samples to investigate in detailed the effect of post-mortem interval (PMI) on the transcriptome state and in particular on the stability of the full range of expressed mRNA transcript levels across a large set of human tissues. We have characterized monotonic and non-monotonic changes in gene expression with relation to PMI and show that these changes are largely tissue specific. We have assessed how several quality control parameters of RNA-seq, as well as different transcript features, are affected by PMI. Beyond, gene expression we quantified PMI related changes in splicing. Finally, we propose mathematical models to take advantage of the information provided by RNA-seq to develop forensic predictors for the post-mortem interval.


About 350 million people worldwide live with a genetic disease. Many of these diseases are poorly understood, treated and handled. Gene therapy has the potential for treating genetic diseases at their roots. The concept is simple: deliver a gene and activate its expression to replace the copy that is functioning abnormally and thereby restore normal cellular activity. A drawback of many current gene therapies is that they do not precisely restrict the expression of the genes delivered to the target cells, which may result in undesired side effects. We are developing OnTarget, software for the design of short DNA sequences to specifically activate the expression of the genes delivered to the target cells. OnTarget incorporates three main capacities: 1) an underlying data repository integrating global collections of cis-regulatory annotations, both experimentally determined and computationally predicted, including promoters, enhancers, transcription factor binding sites (TFBSs), DNA accessibility, histone marks and topologically associating domains; 2) the selection of endogenous cis-regulatory regions controlling the expression of genes in specific cells or tissues; and 3) the modification of TFBSs for the modulation (enhancement or reduction) of the amount of gene expression. We anticipate that OnTarget will not only advance the analysis and understanding of genome regulation, but also accelerate the development, reduce the costs, and increase the safety and efficacy of gene therapies for many patients who are in need of treatment.

17. **Tissue discriminators in the minor spliceosome composition and in mitochondrial RNA decay with progressing ischemia.**

*Michael Sammeth, Caio Padoan, Bruna Bernardes, and Martin Oti.*

The chemistry of our body shapes our overall condition, and therefore also attributes of donors in the GTEx dataset can be mirrored by the composition of RNA molecules in each of their cells. Already during the pilot phase of the GTEx Project we reproduced the variation in mitochondria concentrations across human tissues by read counts obtained from RNA-Seq data. Employing the midpoint dataset, we extend our study of mitochondrial gene expression and correlate it particularly to the RNA decay described by the RNA integrity number (RIN) values across different post-mortem intervals (PMIs). Furthermore, we examine the connection between the medical-social metadata collected by the GTEx Project and gene expression in specific tissues of the corresponding donors. In this regard, based on the Pilot Phase dataset, we could reproduce the loss of specific TH2 cell surface markers in Whole Blood samples of donors that have been potentially infected with HIV. Overall, our investigations indicate several possibilities by which the condition of an individual can affect the molecular composition of transcriptomes in certain cell types of the body, shedding some more light on the complex convolution of factors that contribute to shape the RNA landscape of each sample.

18. **Visualising GTEx gene expression data with other baseline expression and cancer data sets through Expression Atlas.**

*Irene Papathanodorou, Nuno Fonseca, Maria Keays, Amy Tang, Elisabet Barrera Casanova, Wojciech Bazant, Anja Fullgrabe, Laura Huerta Martinez, Suhail Mohammed, Alfonso Munoz-Pomer Fuentes, Alvis Brazma, and Robert Petryszak.*

Expression Atlas1 is a database and web-service that collects, annotates, re-analyses and displays gene expression data of baseline or differential study design. We have re-analysed GTEx RNA-seq data from all 53 tissue types, mapped their annotations to the EBI’s Experimental Factor Ontology terms and clustered genes with similar expression patterns across tissues within the experiment. A user can query a gene, retrieve its expression across tissues and observe genes with similar expression patterns, if any. [http://www.ebi.ac.uk/gxa/experiments/E-MTAB-5214](http://www.ebi.ac.uk/gxa/experiments/E-MTAB-5214). We have also enabled the display of gene expression in GTEx alongside other large-scale baseline gene or protein expression studies such as ENCODE, FANTOM 5, Human Protein Atlas and Human Proteome Map. A user can query a gene or a tissue and as a result observe the expression of this gene across tissues and across different studies. This view provides an implicit comparison of the result across studies while giving clues into their
reproducibility. Finally, we have analysed GTEx using the International Cancer Genome Consortium (ICGC) analysis pipeline, in order to enable integrative views of expression data in cancer tumours, alongside the normal baseline gene expression values derived from the same tissue type in GTEx. This view can be informative, in cases where matched normal tissue from the cancer patients is absent, but also to compare gene expression from matched normal tissue in cancer patients to that from healthy individuals in the GTEx study. All data sets are available for download via text files or R objects.

19. **GTEx-based gene expression features are a key component of machine learning models for the characterization of genes associated with neurological diseases.**

Juan Botta, Sebastian Guelfi, Karishma D’Sa, Jana Vandrovcová, John A Hardy, Michael Weale, and Mina Ryten.

Over the past 5 years there has been a massive growth in genetic testing and this has had a huge and arguably disproportionate impact on our understanding of neurological disorders. In parallel there has been an equally impressive growth in the availability of high quality, large, standardized omic datasets like GTEx. In this context we ask whether we can identify the key features of a gene relevant to diseases of the central and peripheral nervous system. We apply Machine Learning (ML) on DNA and RNA features of disease genes defined by DisGeNET and expert curation to achieve this. Regarding DNA, we include ExAC pLI (probability of being intolerant to Loss of Function) as a predictor. Regarding RNA data, we include specificity of gene expression in 42 tissues as detailed in GTExV6, and gene overall and specific connectivity patterns as measured by WGCNA-based co-expression networks. We demonstrate that for many diseases gene connectivity rather than gene expression is a more useful classifier. Furthermore, this approach highlights the importance of gene expression within adipose tissue (over brain tissue) for Alzheimer’s disease (AD), in keeping with evidence linking adiposity and AD risk. A p-value of 5.79e-10 for the classifier accuracy being better than the non-information ratio suggests the utility of this approach. We conclude that, by using ML, and in particular GTEx-based gene expression features, we can efficiently generate novel insights into the location and processes driving neurogenetic diseases.

20. **The role of non-coding variants in prostate cancer.**

Katherine Hartmann.

Cancer driver mutations occur on a background of ‘normal’ variation but little is known about how common regulatory variants may interact with known driver mutations. Such regulatory variants are strongly implicated in common diseases, including cancer, by GWAS studies that find >90% of trait-associated variants in non-coding regions of the genome (Hindorff 2009). To understand the potential influence of common, non-coding variation in cancer, we begin by identifying regulatory variants for candidate genes implicated in prostate cancer – census genes established by COSMIC (Futreal 2004). Using the GTEx database we identify tissue specific regulatory variants for all 25 candidate genes. We then ask if these genes are differentially expressed in cancer versus normal tissue, which may suggest a clinical role for variation in gene expression. We also consider the haplotype structure of each locus, testing if known driver mutations occur more frequently on a given genetic background. We are particularly interested in whether regulatory variants that decrease gene expression may predispose a gene to higher rates of mutation as suggested by Lawrence et al 2013. Understanding whether non-coding variants modify driver mutations will improve our ability to account for the genetic risk of disease and treatment response.

21. **Leveraging the GTEx histological data: phenotype extraction.**

Manuel Muñoz Aguirre and Roderic Guigó.

The GTEx dataset contains histological images associated to samples from many different tissues, which have been annotated by expert pathologists with free-form text comments regarding the findings of the pathology review. These comments constitute a valuable resource of information about the phenotypes and characteristics observed in a given tissue sample, however, they are not standardized into a user-friendly format that can be employed to perform analyses related to other components of the GTEx dataset. Therefore, we developed a methodology to automatically retrieve phenotypes and terms of interest from pathology review comments and transform them into vectorized annotations.

In addition, the histological images also constitute an opportunity for phenotype extraction: we present our ongoing work on automated tissue piece segmentation from the images in order to facilitate further analyses. We also discuss the technical implications of the process of extracting relevant image features using machine learning techniques. The final aim of the latter task is to aid the pathologist’s work by providing probabilistic information about the tissue samples in the histology slides having or not a specific condition.
22. **Dominant isoform switch events across tissues and in tumours.**

*Nuno A. Fonseca, Sergio Santos, and Alvis Brazma.*

Alternative splicing is an important process in the regulation of gene expression in eukaryotes, through which a single gene can lead to different proteins. We can use RNA-seq data to identify which transcripts of each gene are being expressed in a specific condition, even though quantification of different isoforms remains a difficult problem. We have shown previously that for most genes in most tissues, one transcript is dominant, contributing to more than 80% of the human transcriptome [1]. We define a transcript of a particular gene as n-fold dominant if it is expressed at least n times more than the second most abundant transcript. We observed in GTEx that the dominant transcript of a gene tends to be the same across tissues where it is expressed, however there are a relatively small number of cases where the dominant isoform switches to another. We call these switch events. We observed 429 protein coding genes having switch events across tissues in GTEx. The results were validated in 20 (out of 32) tissues by using the Uhlén dataset [2] as control. It has been shown that the use of different isoforms has been connected to many diseases, namely to cancer [3]. We assessed the level and identified dominant 2-fold isoform switches between tumour samples (from 27 ICGC studies) and GTEx. Overall, on each sample, we observed isoform switches on 5% of the genes. Among the isoform switch events detected, we observed novel and previously reported isoform switch events (e.g., involving CTNNB1, PRICKLE1 [4], or MYH11, PPARG [5]).

23. **A GTEx re-annotation to identify high-quality healthy samples, leveraged in innovative transcriptomics analyses.**

*Frederic Bastian, Anne Niknejad, Amina Echchiki, Julien Roux, Bgee Team, and Marc Robinson-Rechavi.*

We present here the integration of GTEx data with other human and animal transcriptome data in Bgee (http://bgee.org/). We performed re-annotation to retain only healthy and non-contaminated samples, using the information available under restricted-access. We rejected all samples for 31% of subjects, deemed globally unhealthy from the pathology report (e.g., drug abuse, diabetes, BMI>35), and specific samples from another 28% of subjects with local pathologies (e.g., brain from Alzheimer patients). We also rejected samples with contamination from other tissues. In total, only 50% of samples were kept. All these samples were re-annotated to Uberon anatomy and to aging ontology terms. All corresponding RNA-Seq were analyzed consistently with other healthy RNA-seq from human and other species. We have notably developed a new method to define for each RNA-seq library a threshold for calling genes actively expressed over background transcriptional noise. This method largely outperforms methods using a naïve cutoff of, e.g., 1 RPMK or 2 TPM. This allows the integration and comparison of GTEx data with other expression data in Bgee. These processed data are being made available through the Bgee web-application, and through the R package BgeeDB, with sensitive information hidden. For each gene, Bgee provides a unique ranking of expression over anatomy. Moreover, Bgee allows to perform TopAnat analyses of enrichment of expression patterns. TopAnat is a new tool allowing to discover the organs where a set of genes is preferentially expressed. These tools now integrate and leverage the abundant expression data provided by GTEx, providing biological knowledge from such large transcriptomics datasets integrated with many smaller ones.

24. **The Role of Spliceosomal Components in Tissue Heterogeneity.**

*Caterina Coli, Panagiotis Papasaikas, The GTEx consortium, Roderic Guigò, and Juan Valcarcel.*

Generation of mature mRNAs in metazoa typically involves the removal of pre-mRNA segments referred to as introns through the process of Splicing. Definition of intronic boundaries and catalysis of their removal relies on a highly ordered, multistep cycle of recruitment, rearrangement and release of over 150 components that comprise one of the most complex molecular machineries known as the Spliceosome. Additional auxiliary factors can enhance or repress the interaction between the spliceosome and the pre-mRNA resulting in Alternative Splicing (AS), the phenomenon of shifting of intronic boundaries in different contexts. Alternative Splicing has a key role in establishing the differential expression profiles underlying cell differentiation and its misregulation is known to be involved in a wide array of human pathologies. Previous work of our lab identified a network of functional interactions among spliceosome components, auxiliary splicing factors and few chromatin modifiers based on their effects on the regulation of a small number of alternative splicing events (Papasaikas, Tejedor et al, Mol Cell 2015). In addition this work highlighted the spliceosome components as potential AS regulators. Different tissues exhibit highly distinct patterns and varying prevalence of AS. Therefore, the conservation of these functional interactions and the contribution of spliceosome components in regulating AS in particular tissues are fundamental questions for understanding alternative splicing regulation.

I am working with the large collection of data from the Genotype-Tissue Expression project (GTEx) in order to shed light on these questions. GTEx provides transcriptome quantifications for multiple samples across over 40 tissues, thus supplying a high-resolution view of AS across different cell types. The comprehensive nature of these data can also allow me to investigate functional AS interactions for different physiological states or pathological states and establish detailed connections between the presence of specific functional interaction and sequence characteristics of subsets of targets. Our work aim to shed light on the organization and versatility of the splicing regulatory circuitry across tissues by analyzing a large array of high throughput data and has the potential to greatly expand our knowledge of the mechanisms that underlie AS regulation in cell differentiation and disease.
25. Identification and analysis of splicing quantitative trait loci in GTEx.

Diego Garrido-Martín, Ferran Reverter, Miquel Calvo, and Roderic Guigó.

The identification of genetic variants associated with alternative splicing (splicing quantitative trait loci or sQTLs) in large RNA sequencing studies is still a developing field. We aimed to identify sQTLs across GTEx tissues, exploring their enrichment in biologically relevant features, sharing across tissues, effect size distribution and splicing events associated to them. Splicing should be treated as a multivariate phenotype to be recapitulated completely. Thus, for sQTL mapping we employed sQTLseekeR, a method that identifies variants associated with changes in the relative abundances of a gene’s transcript isoforms, relying on a non-parametric analogue to MANOVA. In total 15,831 protein coding genes and lincRNAs and 3,831,874 SNPs (5KB window around genes) were tested for association. On average per tissue we found 14,749 sQTLs (5% FDR) that affected 1,017 genes (sGenes). The ratio sGenes/expressed genes and the sample size per tissue showed a substantial correlation. sQTLs were significantly more exonic and fell within splice sites, RBP binding sites and GWAS hits’ vicinity more than non-sQTL SNPs. GO enrichment analysis of sGenes revealed an enrichment in immune processes and functions. Although most of the identified sQTLs had subtle effect sizes, a considerable number showed larger effects. sQTLs displayed a high degree of sharing across tissues, with tissue specificity accounting only for 3 to 27%. Most sQTLs involved complex changes in first/last exons and UTRs. In addition to these analyses, we are currently focused in the extension of sQTLseekeR's statistical framework to be able to assess significance of more than one factor through an asymptotic approximation, avoiding permutation-based approaches. We are also working in the design and implementation of a multiple testing correction schema for sQTLseekeR analogous to the proposed in FastQTL.

26. REDIportal: a comprehensive repository of A-to-I RNA editing events in GTEx samples.

Ernesto Picardi, Tiziano Flati, Silvia Gioiosa, Anna Maria D’Erchia, Claudio Lo Giudice, Tiziana Castrignanò, Giovanni Chillemi, and Graziano Pesole.

RNA editing is a relevant co-/post-transcriptional process that, in humans, involves mainly the deamination of adenosines to inosines by the family of ADAR enzymes acting on double RNA strands. A-to-I RNA editing has a plethora of biological effects depending on the RNA region involved in the modification. Changes in UTRs can lead to altered expression, whereas modifications in coding protein regions can induce amino acid replacements with more or less severe functional consequences. Interestingly, RNA editing deregulation has been linked to a variety of human disorders including neurological and neurodegenerative diseases and cancer. Indeed, A-to-I editing is required for neuronal function as many targets are key mediators of synaptic signalling. The detection of RNA editing events at genomic scale has been largely facilitated by the advent of high-throughput sequencing technologies. Using RNAseq and WGS dataset we have recently released the first RNA editing atlas in humans comprising more than 3 million of events. Although the well-known biological relevance of RNA editing in humans, many functional aspects are yet unknown. Therefore, large collections of A-to-I events are demanding to elucidate its role in physiological as well as pathological conditions. To this aim we have developed an ad hoc repository called REDIportal, that is based on a huge number of RNAseq experiments (>2500) from GTEx project. Its current release includes 4,668,508 non-redundant A-to-I events in 55 body sites of 150 GTEx individuals for which WGS data are available. By using our resource, users can interrogate genomic regions or individual genes, looking not only at the specific site but also at a plethora of additional info. REDIportal embeds JBrowse to inspect RNA editing events in their genomic context, is freely available at http://srv00.recas.bainfn.it/atlas/index.html and is supported by the Italian node of Elixir.

27. The emerging, positive role of somatic retrotransposition in cell identity, phenotype variation and reprogramming.

Valerio Orlando and Piero Carninci.

Understanding the genomic and epigenomic bases of phenotypic variation remains a central question in biology. In this context the contribution of repetitive transposable elements to human genome function and cell identity remains to be elucidated. Indeed, in contrast with general belief, recent reports indicate that mammalian somatic cells support retrotransposition, particularly in the brain. Whether this phenomenon is restricted to brain cells, TE distribution is random or follows a developmental program, its functional significance remains open questions. To this aim we investigated L1 and other TEs dynamics during human and mouse myogenesis and in cell reprogramming. We found that during myogenic program L1 retrotransposition is specifically activated under the control of both tissue specific transcription factors and epigenetic regulation involving hundreds of de novo insertions. Repeat capture and Whole Genome Sequencing analysis reveal a tissue specific integration profile with a tight correlation between de novo insertions, de novo lncRNA production and positive transcriptional regulation of recipient gene loci. Interestingly, Duchenne Muscles Dystrophy (DMD) patients show block of L1 dynamics due to HDAC deregulation. Pharmacological or gene therapy rescue of myogenic cell differentiation potential and DMD phenotype restores normal L1 expression and CNV both in mdx DMD model mice and in human DMD primary muscle cells. In another work we investigated L1 dynamics in somatic cell reprogramming model of transdifferentiation finding that L1 mobilization, neuronal gene loci colonization and lncRNA production are
required also for efficient reprogramming. Finally, we observed a progressive increasing of L1 CNV accompanying aging. Our data suggest that somatic retrotransposition is a non random, integral part of developmental programs underlying cell specialization and phenotypic variation, a so far unexplored level of complexity in human genome function. Therefore we propose to explore to saturation the tissue specific structure of the human genome, the somatic repetitive elements distribution and match it with the ncRNA output and underlying regulatory regions. The results of this work will provide an unprecedented, comprehensive picture of the contribution of the noncoding genome to phenotypic variation and tissue specialization, hopefully shedding light also on unexplored aspects of human diseases.

28. Alternative splicing produce unique transcripts in different tissues, individuals, and species. 
Reza Sodaei, Manuel Irimia, and Roderic Guigó.

Most of mammalian genes produce many different mRNA isoforms via alternative splicing (AS). However, the potential of AS to generate truly unique transcriptomes remains unknown. Here we took a highly stringent approach using GTEx data to identify AS events that produce transcripts that are unique to a specific tissue, to small groups of individuals or to the human species. First, we found that only brain, muscle/heart and testis have tissue-unique transcripts generated through AS. However, while most of the exons that are included or excluded uniquely in brain and muscle are predicted to produce alternative protein isoforms and are highly conserved, testis-unique exons usually disrupt the reading frame and are human-specific. Remarkably, the majority of brain-unique exons are microexons, with lengths ≤ 27 nucleotides. Second, we found a few dozen of exons that are uniquely skipped only in one or few individuals, and we are currently trying to identify SNVs associated to these molecular phenotypes. We did not find sex or age-related unique AS events. Finally, in line with previous studies, we find that AS can generate species-unique transcriptomic programs with wide regulatory potential. In summary, we found that AS has a limited yet significant contribution generating unique transcriptomes across different levels of biological organization (tissues, individuals and species).

29. The Ensembl Cis-Regulatory Annotation.
Daniel Zerbino.

Ensembl is one of the world’s leading sources of information on the structure and function of the genome. It already provides an up-to-date, comprehensive and consistent database that brings together genome sequences, genes, non-coding RNAs, known variants, etc. In particular, Ensembl’s Regulatory Build synthesises public epigenomic datasets produced by large-scale projects such as ENCODE, Roadmap Epigenomics or BLUEPRINT. We process them through a unified pipeline and make them available through a single interface. We also define functionally active regions across 68 human cell types (on both the GRCh37 and GRCh38 assemblies) and 6 mouse cell types, assigning them a function wherever possible. We are currently expanding our annotation to more cell types. In particular, we maintain the International Human Epigenome Consortium’s (IHEC) Epigenome Reference Registry (EpiRR) where teams from around the world record the metadata describing their epigenomic datasets before they are incorporated into IHEC’s Data Portal. We hope to expand soon to more species in collaboration with the Functional Annotation of Animal Genomes (FAANG) project.

Regulatory elements are chiefly of interest because of their action on genes; we are therefore simultaneously developing a database of cis-regulatory interactions attaching them to their target genes. Currently, two main approaches are being used to detect these interactions: genetics (e.g. eQTLs) and chromatin conformation (e.g. Hi-C). We have developed new technologies to store and display these datasets, using in particular HDF5 indexing for fast retrieval. This technology allows us to store and display all of the GTEx summary eQTL data, as opposed to only significant correlations, thus avoiding interval censoring. Over the next year, we will be integrating more eQTL and Promoter Capture Hi-C datasets, organized by tissue.

Finally, we are also developing high performance tools for high-throughput analysis. We foresee that in the future most genomic data will be confined by local legislation and it will be necessary to deploy applications across remote data centers. Using our RESTful APIs, it is already possible to retrieve our data simply and efficiently for any gene, variant or region, along with all other Ensembl annotations such as LD calculations from the 1000 Genomes dataset Phase 3, conservation scores, etc. It is thus possible to quickly develop advanced functional analysis pipelines without having to download or process massive data files, as we demonstrate with a post-GWAS analysis pipeline called POSTGAP.
PRACTICAL INFORMATION

Venue: Barcelona Biomedical Research Park - PRBB

Barcelona Biomedical Research Park - PRBB
C/ Doctor Aiguader, 88
08003 Barcelona, Spain
Auditorium

Contact persons during the event

Romina Garrido
Assistant/Administrative Associate, Centre for Genomic Regulation (CRG).
romina.garrido@crg.es | Phone: +34 629 62 55 74 (Spain)
www.crg.eu

Casandra Trowbridge
Senior Project Coordinator for the Genotype-Tissue Expression (GTEx) Project’s Laboratory, Data Analysis, and Coordinating Center (LDACC) at the Broad Institute.
ctrowbri@broadinstitute.org | Phone: 239-910-0285 (US cell number)
www.broadinstitute.org | www.gtexportal.org

Marta Soler
Head of Research and Scientific Debate, Biocat.
msolet@biocat.cat | Phone: +34 662315500 | +34 93 310 33 57
www.bdebate.org | www.biocat.cat
Dinner – Thursday, April 20th

Restaurant Agüelo013

http://aguelo013.es/
Avinyó 37,
Barcelona
Time: 8:00pm

Located in the heart of Barcelona’s Gothic Quarter. Thousands of locals keep countless memories about the days when it was one of the most typical taverns in the city. It was built in 1850 in one of the villas owned by Earl of Fonollar, Marques of Palmerola. At its beginnings in the early nineteenth century, it hosted a coal bunker, and then a wine cellar. It delivered wine in bulk which was then bottled in carafes. In 2013, the place was transformed into an avant-garde restaurant.

Directions

The restaurant is 15-20 minutes walking from the meeting venue.

You can also take the bus D20 (C/Trelawny in front of the Pullman Hotel). Drop off after 7 stops (Pg. Colom-Via Laietana stop – ID:1788). Then walk 4 minutes.
OUTCOMES

B·Debateca

On the website of B·Debate, you will find all the information related with the celebration of the meeting that includes reports, conclusions, scientific documents, interviews with the experts, speaker’s CVs, videos, images, press documentation and other related materials. We invite you to visit the section B·Debateca on www.bdebate.org

Contents of the meeting “The Genotype Tissue Expression (GTEx) Project Community Meeting”

FOLLOW US ON SOCIAL MEDIA

#Bdebate  #GTEx2017
@Bdebate
ORGANIZERS

**B·DEBATE** International Center for Scientific Debate Barcelona is a **BioCat** initiative with support from "la Caixa" Foundation. It drives first-rate international scientific debates, to foster dialogue, collaboration and open exchange of knowledge with prestigious national and international experts, to approach complex challenges of high social interest in life sciences. B·Debate sees debate as a powerful, effective way to generate knowledge and strives to help position Barcelona as a benchmark in generating knowledge and Catalonia as a country of scientific excellence.

B·Debate sees debate as a powerful, effective way to generate new knowledge. The debates are top-notch international scientific meetings featuring a selection of experts of renowned international prestige and scientists that work in Barcelona and Catalonia, moderated by scientific leaders. Since 2009 B·Debate has invited about 1200 recognized speakers and over 7,000 attendees. B·Debate seeks out answers to the challenges and needs of society in the field of life sciences, taking into account the complex, ever-changing conditions of this global world. The debates foster the integration of different disciplines of science and deal with such diverse topics as ageing, new therapeutic approaches to various diseases, innovative technology to improve knowledge of the human genome, food resources, new tools to integrate knowledge management, clinical genomics, neurosciences, climate change, and new energy sources, among others. The knowledge and results obtained through these events is spread throughout both the scientific community and general society through the various B·Debate channels and instruments.

More info: [www.bdebate.org](http://www.bdebate.org)

---

**B·Debate 2018 Call for Proposals**

Open Call for proposals for the B·Debate calendar of activities

In order to carry out its mission, B·Debate opens the call for proposals of activities to include in the 2018 calendar. The call is governed by the "2018 B·Debate Call rules and regulations".

**Deadline**

The call for proposals for activities to include in the B·Debate 2018 will remain open from March 31st until June 21st 2017 (both included).

**More Information**

For further information of the B·Debate Call for Proposals please contact:

- B·Debate
  - Baldiri Reina, 4-6, Torre 1 (PCB)
  - 08028 Barcelona
  - Phone: +34 93 310 33 30
  - bdebate@bdebate.org
  - Marta Soler – msoler@biocat.cat

---

![CRG Logo](https://example.com/crg_logo.png)

More info: [www.crg.eu](http://www.crg.eu)

The Centre for Genomic Regulation (CRG) is an international biomedical research institute of excellence, created in December 2000. It is a non-profit foundation funded by the Catalan Government through the Departments of Economy & Knowledge and Health, the Spanish Ministry of Economy, Industry and Competitiveness, the "la Caixa" Banking Foundation, and includes the participation of Pompeu Fabra University.

The mission of the CRG is to discover and advance knowledge for the benefit of society, public health and economic prosperity.

The CRG believes that the medicine of the future depends on the groundbreaking science of today. This requires an interdisciplinary scientific team focused on understanding the complexity of life from the genome to the cell to a whole organism and its interaction with the environment, offering an integrated view of genetic diseases.

The CRG is a unique centre in Spain, based in an innovative organization research model. Group leaders at the CRG are recruited internationally and receive support from the centre to set up and run their groups. An external Scientific Advisory Board, made up of 13 world leaders in the different areas, evaluates them. The result of evaluations conditions the future of the CRG scientists, no matter whether they have open-ended or time-limited contracts. This ensures the mobility and the renewal of the workforce.
CosmoCaixa offers interactive, enjoyable science and an open door for anyone who is eager to learn and understand and who never stops wondering why things are the way they are. CosmoCaixa Barcelona boasts the Geological Wall and the Amazon Flooded Forest, which features more than 100 plant and animal species that convince visitors they have been transported from the Mediterranean to the very heart of the tropical jungle. In addition to its permanent facilities and its open areas, CosmoCaixa offers a scientific and educational programme that includes exhibitions, workshops, conferences, courses and debates involving experts from all over the world.

More info: www.observaciones.lacaixa.es

**SPONSORS**

The Company of Biologists is a not-for-profit publishing organisation dedicated to supporting and inspiring the biological community. We are run by distinguished practicing scientists. We exist to profit science, not shareholders. We inspire new thinking and support the worldwide community of biologists. We do this by publishing leading peer-reviewed journals, facilitating scientific meetings and communities, providing travel grants for young researchers and by supporting societies and gatherings.

More info: www.biologists.com

More info: www.elixir-europe.org/industry

More info: www.qiagen.com

More info: www.illumina.com