

EMBO Workshop

Network inference in biology and disease

September 10-13, 2019 - TIGEM, Naples

<http://meetings.embo.org/event/19-network-inference>

Public summary

Organizer: Christian Neri

Co-organizers: Diego di Bernardo, Claudia Manzoni, Erich Wanker, X. William Yang.

Session I: New methods in network inference

Chair: Diego di Bernardo, TIGEM, Puozzoli

Design and quantitative understanding of biological systems

Luis Serrano Pubul, CRG, Barcelona, Spain

Summary of talk: Here, we determined the relative importance of different transcriptional mechanisms in the genome-reduced bacterium *Mycoplasma pneumoniae*, by employing an array of experimental techniques under multiple genetic and environmental perturbations. Of the 143 genes tested (21% of the bacterium's annotated proteins), only 55% showed an altered phenotype, highlighting the robustness of biological systems. We identified nine transcription factors and their targets, representing 43% of the genome, and 16 regulators that indirectly affect transcription. Only 20% of transcriptional regulation is mediated by canonical TFs when responding to perturbations. Using a Random Forest, we quantified the non-redundant contribution of different mechanisms such as supercoiling, metabolic control, RNA degradation, and chromosome topology to transcriptional changes. Model-predicted gene changes correlate well with experimental data in 95% of the tested perturbations, explaining up to 70% of the total variance when also considering noise. This analysis highlights the importance of considering non-TF-mediated regulation when engineering bacteria.

Continuous time Bayesian networks to infer global regulatory networks in humans

Fabio Stella, Dpt. of Informatics, Systems and Communication, University Milan-Bicocca, Milan, Italy

Summary of talk: Continuous time Bayesian networks are a special kind of probabilistic graphical models that allow to model complex stochastic processes. In particular, continuous time Bayesian networks are efficient to model highly structured stochastic processes where each variable of the considered system depends on a limited number of other variables. We introduce continuous time Bayesian networks and show experimental evidence of their effectiveness to learn from biological data. We present recent developments to learn from nonstationary systems, i.e. systems where the dependence structure can change through time. In particular, we introduce non-stationary continuous time Bayesian networks and show how they compare to alternative models when learning from data. Finally, we sketch very recent development about solving one of the main limitations of continuous time Bayesian networks, i.e. their memoryless nature. In particular, we present how continuous time Bayesian networks can be given memory, which is a fundamental feature of biological systems, by using phase type distributions.

Perspectives in the field: Learning gene regulatory networks from biological data, proved to be effective to highlight pathways of many diseases. However, it is increasingly understood that learning from data may be ineffective when asking "why type questions", i.e. when asking questions related to which is the effect of interventions to the system under study, or which are the causes of a genetic disease. The same type of question is asked in medicine when we ask whether a treatment is effective or not to make the patient recovering from a given disease. When the goal of the study is to answer why type questions than, learning from data alone, at the current state of the art, usually fails. However, recent developments in Bayesian networks, and much more in causal networks, opened a new era, an era in which artificial intelligence models will be in the conditions to answer many, not all, questions of the why type. The possibility to coherently combine domain expert knowledge, together with the available data, will reveal the most effective way to address and solve difficult decision problems in medicine, biology and in many other domains.

Deep decoding of network attacking mutations rewiring cancer

Rune Linding, Humboldt-Universität, Berlin, Germany

Summary of talk: Signaling systems in multi-cellular organisms are vital for cell-cell communication, tissue organization and disease. Cancer genomics has unraveled a surprisingly large set of novel gene lesions from tumors. Our previous studies have globally explored the rewiring of cell signaling networks underlying malignant transformation caused by kinases and other signaling proteins. By generating quantitative time-/state-series data and subsequently using these as input for deep learning based computational modeling - our lab work identify the principal changes in the genome, cell signaling and phenotypes of cells harboring genetic mutations; we validate these models by forward prediction of experimentally observed phenotypic responses to drug and genetic perturbations. We are currently deploying such forecasting models on data collected from patient derived (PDX) tumors to describe how the cell signaling networks are mechanistically, dynamically and differentially utilized in cancers. Finally, we are working to utilize deep learning to predict novel treatment and diagnostic strategies for tumors harboring different genetic lesions. In conclusion, our studies aims to unravel the fundamental rewiring of cell signaling networks in cancer and will serve as a major breakthrough in our basic understanding of their impact on the disease, paving the way for future clinical applications and tumor specific cancer therapy.

Perspectives in the field: A strategic focus of the community moving forward will be to continue to develop computational tools and to deploy these on genome-scale quantitative data obtained by, for example, mass spectrometry, genomic, and phenotypic screens and single cells to understand the principles of how spatio and temporal assembly of mammalian signaling networks transmit and process information at a systems level in order to alter cell behavior. We are now seeing the advancement of network medicine in which complex therapies targeting signaling networks associated with human disease are being both developed and deployed at an increasing pace. It is very exciting times for both researchers and clinicians.

Network-based prediction of biological interactions and functional annotations

Istvan Kovacs, Northwestern University, Boston, USA

Summary of talk: Life emerges from a complex interplay of molecular interactions. Despite exceptional experimental efforts to map out the human interactome, the continued data incompleteness limits our ability to understand the molecular roots of human disease. In the talk I will show how to fight data incompleteness and biases with novel computational methods, leading to experimentally testable, large-scale predictions. Besides molecular interactions, our approach can reliably predict a broad spectrum of functional associations, including disease associations, pathway membership and genetic interactions, as well as toxic and synergistic drug combinations.

Short talk: Integrative network-based approach identifies gene communities in COPD

Roberta Marino, University of Turin, Italy

Summary of talk: Chronic Obstructive Pulmonary Disease (COPD) is characterised by exacerbation phases alternating to stable conditions. To identify gene communities modulated in these phases, we applied a multi-network strategy, integrating a gene expression network with different layers of omics information. We implemented the pipeline for blood gene expression data from the AERIS clinical study (NCT01360398), which has observed 127 COPD patients for two years. For each condition (exacerbation and stable state) a gene co-expression network was built and integrated with a co-regulation network, a human protein-protein interaction network, a transcription factor and a microRNA co-targeting network. We applied the Infomap algorithm to detect gene communities. Because of the stochasticity of the algorithm we introduced a robustness step assessment, by repeating multiple times the community detection to compute a final consensus clustering and determine the presence of relevant gene communities. Using our pipeline, we identified co-regulated genes in blood samples taken at regular visits and at exacerbation visits. Specific metabolic functional responses are remarkably enriched in both conditions, while co-regulation of a limited set of functions and potential targets of miRNA were enriched/characteristic of exacerbation phases. These last ones are not detected by classical GSEA (Gene Set Enrichment Analysis).

Short talk: High-Content Mouse Behavioral Phenotyping identifies novel putative treatments for Neuropsychiatric Diseases

Mukesh Bansal, VP Data Science, Psychogenics, NJ, USA

Summary of talk: At Psychogenics we developed SmartCube®, a proprietary high-throughput and high-content in vivo screening platform. SmartCube consists of an automated behavioral/physiological system presenting designed to present a sequence of challenges to a mouse and automatically

quantify over 2000 features. Our unbiased approach has several applications ranging from phenotypic profiling of animal models of neuropsychiatric disorders to drug discovery by screening novel compound libraries and comparing their behavioral responses to known neuropsychiatric treatments. We developed Drug-induced Behavioral Signature Analysis (DBSA) which compares drug-induced behavioral signatures to signatures of animal models in order to predict compounds with potential therapeutic efficacy. This high-throughput *in silico* method allowed us to screen a large library of proprietary and reference compounds tested at different (relevant) doses against animal models of neuropsychiatric disorders including Huntington Disease (Q175 knock-in), as well as models of Alzheimer's Disease (transgenic tauopathy rTg4510 and a pharmacomimetic model based on scopolamine treatment). Our screens identified compounds predicted to reverse the behavioral changes of the AD model and successfully identified positive controls such as Galantamine, a known cognitive enhancer along with several other candidate cognitive enhancers. We subsequently validated the top four proprietary compounds in a Novel Object Recognition assay, demonstrating increase in cognition performance.

Session II: Network inference in diseases - Part I

Chair: [Christian Neri](#), Institut de Biologie Paris-Seine, Sorbonne Université, Inserm, Paris

Genetic approaches to dissect genetics of complex disorders

[Raffaele Ferrari](#), University College London, UK

Summary of talk: Complex disorders are the result of multiple risk factors from genetic to environmental contributing to the ultimate phenotype. It is important to consider that many (if not all) of the environmental factors are difficult to identify, measure and control. Conversely, genetics has seen the rise of incrementally improving technologies allowing the study of the genome and to gradually generate large amount of data at increased speed and decreased costs. The genome is a measurable variable and techniques to analyse genetics data have become increasingly sophisticated in order to best interpret genetics information. Genetics of complex disorders is less straight forward than monogenic diseases. It is in fact never the case that one mutation in a particular gene, but rather a combination of multiple genetic markers that contribute, with variable effect size, that lead or contribute to disease; thus it is more the matter of a so-called genetic architecture that predisposes to the development of a particular trait or phenotype.

Perspectives in the field: Genetics data generation and analytical pipelines are constantly improving to provide an initial platform to, not only identify genes on which to model disease biology and pathogenesis, but also use the genetic make-up of an individual to predict disease. It is now truly inspiring times where interdisciplinary cross-talks have started to make the most out of each other's expertise in the scientific community: *in silico* and *in vivo* models to understand disease and validate risk-pathways is the critical way to go in order, for example, to complement diagnostic criteria and improve patients selection for clinical trials. Genetics is a necessary (but not sufficient) component of the next-generation holistic approach to disease dissection and personalised medicine.

With a little help from friends: Applications of weighted network analysis to Huntington's Disease

[Peter Langfelder](#), UCLA, Los Angeles, CA

Summary of talk: After a brief review of main features of Huntington's disease (HD), I presented some of the salient points of analysis of a large RNA-seq data set from striatum and cerebellum of a series of mouse models of HD with increasing mutation load. Individual gene differential expression (DE) analysis identified thousands of DE genes in striatum (the brain region most strongly affected by HD) and fewer but still numerous genes in cerebellum which is relatively unaffected by HD. After a short introduction to Weighted Correlation Network Analysis (WGCNA) and specifically its consensus network analysis variant, I described some of the WGCNA-identified modules of co-expressed genes with strong dependence on HD-causing mutation. In addition to modules, WGCNA also identifies module hub genes which are natural targets for further follow-up. In the last part, I have briefly sketched a striatum-cerebellum consensus network analysis that identifies modules of genes that are co-expressed in both brain regions. Among other modules, the analysis found 3 modules that are downregulated with increasing mutation load in striatum but upregulated in cerebellum. Two of them are enriched in genes specific to *Drd2*-expressing medium spiny neurons, one of the predominant neuron types in striatum. Upregulation of these genes in cerebellum points to disruption of cell type maintenance expression programs in HD. I concluded the presentation by briefly listing WGCNA assumptions and limitations.

Perspectives in the field: A large variety of network approaches to analysis of large data have been shown useful in elucidating important insights in a variety of biological settings. Despite the many successes, numerous challenges remain in analysis methods, interpretation and tighter integration with biologists. Among others, network analysis methods remain somewhat blunt instruments that are often adept at reflecting underlying biology in broad strokes but with finer detail sometimes obscured or missing. By their nature, many network methods are susceptible to being led astray by misunderstood or overlooked sources of variation, suboptimal experimental designs and small sample sizes of the underlying data. These and other methodological challenges and a plethora of application will keep network analysis an active research area for the foreseeable future.

Biological networks, pathways and models: databases for Systems Biology at EMBL-EBI

Pablo Porras, EMBL-EBI

Summary of talk: The European Bioinformatics Institute (EMBL-EBI) has amongst its missions the delivery of services in the form of tools and databases for the benefit of the scientific community working on Bioinformatics. All tools and databases from EMBL-EBI are completely free and open access. In this talk I presented some of the resources that have a specific focus in areas of data and knowledge with a systems perspective.

The IntAct database (www.ebi.ac.uk/intact) focuses its efforts in the representation of physical molecular interactions extracted from experimental evidence found in the literature. It is currently the largest data provider of its kind in terms of binary interactions represented, covering over 950K interactions in 300 different organisms. IntAct uses a deep-detail curation strategy developed as partner of the IMEx Consortium (www.imexconsortium.org). This model allows capturing not only binding events, but also the full experimental setup in which these were detected, along with specific details such as binding affinities, interaction surfaces or mutations influencing interaction outcome, if available in the original publication.

Reactome (www.reactome.org) is a database for human biological pathways authored by domain experts in collaboration with expert curators. The pathway maps are integrated in a hierarchical structure and formed by individual reactions, all of them fully traceable to publication references. Reactome allows navigation through these maps and also features a number of analysis and visualization tools that allow to analyse omics data taking pathway context into account.

BioModels (www.ebi.ac.uk/biomodels) is a repository of mathematical models of biological and biomedical systems. It hosts a vast selection of existing literature-based, physiologically and pharmaceutically relevant mechanistic models in standard formats. This enables the scientific community to find, re-run and re-use existing models, extending their life and usability beyond their publication in scientific literature. The BioModels team have recently launched a new website with extended search capabilities, allowing for parameter searches and additional model formats support.

Additionally, the Complex Portal (www.ebi.ac.uk/complexportal) and OmicsDI (www.omicsdi.org) were also presented. The Complex Portal is an encyclopaedia of known multi-molecular complexes, featuring data about their internal topology, structure, function and relation to disease. Our team has recently completed the full yeast complexome, an invaluable resource for the study of complexes from a systematic perspective. OmicsDI is a repository of omics datasets where every dataset is tagged with specific metadata related to origin, data type, knowledge domain and some statistics about interest and usability.

Perspectives in the field: Data repositories and tools such as the ones hosted at EMBL-EBI remain critical for the day-to-day of systems biology research. Our teams aim to reach to the community in order to meet and even anticipate its needs in a rapidly evolving field with big data at its core. Fostering and curating data from both omics-scale and low-scale data sets remains an essential activity that critically influences our understanding of biology. Future perspectives point at a rapidly increasing importance of such databases and tools, but also to more complex challenges for data providers. We can only meet this challenge and fulfil our responsibility with the engagement of the research community.

Short talk: The genomic underpinnings of oscillatory biomarkers support successful memory encoding in humans

Stefano Berto, University of Texas Southwestern Medical Center, Dallas, TX

Summary of talk: Our understanding of human brain activity is primarily limited to non-invasive methods such as functional imaging that are then used to make correlations to other metrics such as human behavior or genetics. Such limitations are unable to directly address the molecular mechanisms of human brain activity. Our preliminary work has provided evidence to link human brain gene expression with functional imaging data in the resting state highlighting genes and molecular pathways related to neuronal activity. We take such correlations one step closer to directly test whether gene expression can be linked to human brain activity during a specific task. We

hypothesized that gene expression signatures underlie electrophysiological processes driving tasks such as episodic memory. Uncovering these signatures will offer a remarkable opportunity to study human higher cognitive abilities at the molecular level, opening up avenues to discovering therapeutic targets. The goal is to carry out a within subjects study that links gene expression data from surgical biopsies to anatomical and functional information available from the same patients through intracranial EEG recordings carried out during a working memory task and fMRI fALFF signals. Correlations between gene expression patterns and brain activity revealed information about the function of the putative connected genes, providing identification of the first known human genes implicated in memory encoding. In addition, using co-expression network analysis, we isolated memory-specific modules associated with neuropsychiatric disorders as well as ion channel activity. We further revealed that SME-specific modules are enriched for both excitatory and inhibitory neurons using single-nuclei transcriptomic data from resected tissue. Overall, this study generated the first within-subjects human dataset highlighting a list of candidate genes linked with memory encoding recorded by intracranial EEG.

Perspectives in the field: The neuroimaging genetics/genomics approaches allowed us to interrogate the relationship between human brain activity and genetic/genomic pathways. Such translational medicine approach prioritized genes which expression and genetic variation contribute to measurements of higher cognition. These initial results highlighted the genetic and genomic architecture of the human brain activity. Nevertheless, the full extent to which different cell-type contribute to brain activity remain largely unexplored. New genomic technologies, such as single cell RNA-seq, have expanded our understating of the human brain. An intriguing perspective will be the link between brain activity and different cell-types with the potential of uncovering cell-type specific genes to further test in task related experimentation. Understanding the contribution of cell-type to brain activity will provide key insights into the underlying biology of human higher cognition.

Short talk: Tumor homeostates revealed by synthetic locus control regions

Gaetano Gargiulo, Max Delbrück Center for Molecular Medicine, Berlin

Summary of talk: Our understanding of developmental and disease homeostasis relies on genetic and lineage tracing experiments to enable mapping cell types and their fate during tissue homeostasis without perturbing it. Yet, there is no established technology to perform such procedures in a multifactorial setting such as cancer. I discussed that our lab recently developed a method to generate synthetic reporters for genetic tracing of complex phenotypes. I illustrated how we can gain control on solid tumor heterogeneity and reveal drivers and modulators of cell identity and fate changes. Virtually, this method is applicable to any research question in biology involving genetic tracing.

Perspectives in the field: In my view, gaining experimental control on complexity by combining systems and synthetic biology is timely and a needful thing as the massive – yet, descriptive - identification of novel cell types in development and disease increases through single-cell omics. In cancer, I anticipate that genetic tracing of complex phenotypes will enable us to better model tumor complexity in experimental settings and discover more accurate targets and biomarkers for intervention.

Session III: Network inference in diseases - Part II

Chair: [Erich Wanker](#), Max Delbrück Center for Molecular Medicine, Berlin

Deciphering cell fate and signaling in cancer through statistical and other modeling approaches

Emmanuel Barillot, Institut Curie, Dpt. of Computational Systems Biology of Cancer, Paris, France.

Summary of talk: The advent of precision medicine has been made possible thanks to the decreasing costs of high-throughput molecular profiling technologies. Nevertheless, it relies on several pillars that still have to be strengthened, at least in oncology which was the focus of my talk. Our knowledge and understanding of molecular networks governing tumorigenesis and tumor progression has progressed significantly during the past two decades, including the interplay between tumor cells and their microenvironment. However, the currently available theoretical tumor models still need improvement. We therefore recently have developed an Atlas of Cancer Signaling Networks from the curation of literature, which covers the main cellular functions involved in cancer. These integrated networks were applied for comparison of immunoreactive and proliferative ovarian cancer subtypes using transcriptomic, copy number and mutation multi-omics data.

Perspectives in the field: The recent progress in the mathematical modeling of such networks starts bridging the gap with clinical applications. We have designed a methodology for integration of mutation, copy number and expression data in a boolean network modeling framework for predicting

essential phenotypic features of tumors upon treatment. Successful integration of metabolic and signaling networks will allow a broader application of systems biology data for interpretation and retrieval of intervention points to tackle simultaneously the key players coordinating signaling and metabolism in human cancers.

Integrative mouse genetics and systems biology to validate molecular targets in Huntington's disease

X. William Yang, University of California Los Angeles, CA.

Summary of talk: Integrative systems biology is a powerful approach to elucidate novel molecular networks in the pathogenesis of Huntington's disease (HD). Using large-scale RNA-sequencing and proteomics approaches in an allelic series of HD knock-in mice, we have previously identified Huntingtin (Htt) CAG-length dependent weighted gene co-expression network (WGCNA) modules, which implicated multiple molecular pathways in HD pathogenesis (Langfelder et al., Nature Neuroscience 19: 623-633, 2016). In my presentation, I described our implementation of an unbiased, in vivo mouse Germline Network Validation (GNV) platform to evaluate a substantial number of Htt molecular network hub genes for their putative causal roles in modifying disease transcriptomes in vivo. To this end, we established a pipeline to introduce heterozygous knockout alleles of selected GNV candidate genes into the Q140 mutant Htt knockin mice. We then performed RNA-seq analyses of all the resulting genotypes at 6-months of age to evaluate whether a genetic reduction of one allele of a GNV gene can lead to transcriptome-wide or module-specific modification of mutant Htt-induced gene expression changes. To date, we have completed RNA-seq and bioinformatic analyses of 52 GNV gene perturbations of Q140 transcriptomes. I presented examples of *Foxp1* heterozygosity that worsens mutant Htt induced downregulation of striatal medium spiny neuron genes, and a few other GNV genes that modify the oligodendrocyte gene expression in HD mice. Our ongoing study may help to elucidate critical genes that interact with mutant Htt to either rescue or exacerbate the disease-related transcriptomes in the HD mice.

Perspectives in the field: A major challenge in applying systems biology to study brain disorders is to test a large number of candidate genes for their potential roles in disease modification in relevant in vivo models. With the rapid advance in mouse genetics (i.e. CRISPR/Cas9 technology), the traditional means of phenotyping (e.g. behavior and neuropathology) is labor-intensive, poorly scalable and often has poor reproducibility. Our study demonstrates for the first time a robust, reproducible and scalable integrative mouse genetic and systems biology platform that allows the unbiased evaluation of candidate modifiers of disease-relevant molecular networks in the mammalian brain.

Integration of network layers for modeling the temporal dynamics of genetic cooperativity in Huntington's disease

Christian Neri, Institut de Biologie Paris-Seine, Brain-C Lab, Paris, France.

Summary of talk: Understanding the molecular and cellular pathogenesis of neurodegenerative diseases (ND) and prioritizing disease targets for early drug trials may greatly benefit from systems modeling studies in which the production of highly dimensional data (e.g. time series data, multi-layer data, multisource data, and any combination thereof) provide solid grounds for modeling the dynamics of NDs at high definition. However, models generated from the analysis of multidimensional data may retain a large target space or lack biological precision, or both. I discussed how our laboratory develops a series of machine learning pipelines (BioGemix framework) that support a high level of data integration while ensuring biological precision, which may enhance target prioritization. I showed how one of these pipelines, namely Weighted Edge Network Analysis (WENA: Bigan *et al*, Bioinformatics 2019), can generate a reliable multi-layer network model of the temporal dynamics of Huntington's disease (HD) in the brain of HD knock-in mice. WENA analysis notably predicts that a small network of cellular homeostasis regulators intertwined with cellular senescence inducers is specifically recruited for responding to the disease in the striatum of symptomatic HD mice, which is not seen earlier nor in the cortex. I then illustrated the value of *C. elegans* and human stem cell models of HD for biological follow-up studies in which mechanistic studies can be performed to understand how HD may alter the plasticity of stress response and cell repair and the balance between compensation and decompensation over time. In the context of the most recent findings on NDs as well as on aging, the discriminative power of this approach and tools is validated for research on NDs other than HD and has strong potential for research on aging and longevity.

Perspectives in the field: The production of highly dimensional data in ND research can greatly enhance disease modeling and target prioritization, particularly as it comes to select the mechanisms that support a proper level of cellular compensation in the early phases of ND processes and to develop early drug trials in order to delay clinical conversion from presymptomatic to manifest disease stages. Along these lines, the development of machine learning methods that ensure

predictive reliability as well as biological precision and that is coupled with model validation is foreseen to be a key component of research on brain maintenance, aging and neurodegeneration. Depending on the number of testable hypotheses retained by these methods, data production and machine learning needs to be integrated with quantitative biology and R&D, illustrating how multidisciplinary is now essential to the field of molecular neurosciences, and, more largely, may be essential to any field of biology in which in-depth molecular profiling studies are needed to understand how intra- and inter-cellular signaling processes may work in a context-dependent manner in the course of development, maintenance and aging.

Short talk: Inferring networks for bone marrow hematopoietic stem/progenitor supportive stromal cells in a developmental context

Pierre Charbord, Dpt. of Developmental Biology UMR 7622, Sorbonne Université, Paris

Summary of talk: Using different algorithms (Weighted Gene Correlation Network Analysis, multivariate information-based inductive causation) and a large panel of transcriptomes of stromal cells we have inferred the gene networks characteristic of different populations of Hematopoietic Stem Cell-supportive bone marrow stromal cells. Such results should help identify the most connected hub genes critical for the maintenance of Hematopoietic Stem Cell niches.

Perspectives in the field: Our strategy, defining first the gene organization underlying a functional phenotype then identifying the cell sets contributing to that phenotype, may be generalized to the study of any transcriptomes contrasting two opposing phenotypes.

Session IV: Network approaches in individual cells

Chair: [Raffaele Ferrari](#), UCL, London, UK

Reverse engineering single cell data: challenges and perspectives

Barbara Di Camillo, Information Engineering Department, University of Padova

Summary of talk: My research activity is centered in the development and application of advanced modeling, data mining and machine learning methods for high-throughput biological data analysis in the field of Bioinformatics and Systems Biology. In particular, I have developed and applied different advanced data mining and machine learning methods for robust biomarker discovery, predictive modeling and clustering of next generation sequencing (NGS) data. These methods include the prediction of biological annotation and its integration in the learning phase. Moreover, I have also a great expertise in the development and application of differential equation based models, Boolean and Bayesian Networks for deterministic and stochastic modeling of transcriptional networks and signaling pathways, reverse engineering of transcriptional networks and integration of genetic, phenotypic and environmental risk factors via Systems Genetics approaches. In this talk, I have reviewed the main challenges and opportunities of reverse engineering single cell data and have shown some example of how to apply it to decipher gene regulation and cell-cell interaction. Tissues and organs are complex systems made of multiple cells exhibiting different system states, which may vary from cell to cell because of the presence of heterogeneous subpopulations, cells at different stages of the cell cycle/development, and/or cells responding to different stimuli within their microenvironment. Moreover, cells are spatially and temporally organized and able to communicate and interact with each other to orchestrate self-assembly and response to stimuli as a whole.

Perspectives in the field: A crucial issue in genomic studies is indeed the elucidation of how genes change expression and interact at single cell level and how cells interact to each other. In general, given a system whose elements regulate each other, inference of the regulatory network from the observed dynamics of the system is denoted as reverse engineering. In this context, single cell RNA sequencing (scRNA-seq) has emerged as a powerful tool to study individual cell transcriptomes on a large scale. The high resolution of this technology offers, beyond the advantages of monitoring, at least in principle, the entire genome, the unique opportunity to dissect the transcriptional landscape of single cells. However, the bioinformatics analysis of such new kind of data is challenging. Indeed, scRNA-seq count data shows many differences compared to bulk RNA-seq count data, making the application of existing reverse engineering analysis methods not straightforward or even inappropriate.

Network inference from single-cell data - single cell genomics

Jean-Philippe Vert, Institut Curie, PSL Research University, ENS and MINES ParisTech, Paris, France

Summary of talk: I'm expert in the machine learning approach to computational and systems biology, and presented in my talk some recent work on gene regulatory network reconstruction from single-cell genomics data. I had four important messages: 1) single-cell transcriptomics data can bring important information regarding gene regulation that may be hidden in standard, bulk transcriptomics studies; 2)

the possibility to collect data for many individual cells, even of limited quality, is a blessing for machine learning-based inference techniques; 3) algorithms for network inference from single-cell data can borrow ideas from techniques used in standard, bulk transcriptomics studies, but must be adapted to the specificities of the data, such as the need to infer the "velocity" of each cell in the space of gene expression profiles; 4) integrating other omics or image data at the single-cell level is a promising direction to infer a more complete modeling of cellular behaviors.

Perspectives in the field: Overall, large-scale productions of data at the single-cell level in projects such as the human cell atlas of the Lifespan initiative is likely to drastically improve the quality of network inference tools.

Single-cell co-expression networks in cancer

Diego di Bernardo, TIGEM, Pozzuoli (Naples), Italy

Summary of talk: Our research aims at developing and applying experimental protocols and computational algorithms to elucidate mechanisms of genetic diseases and develop therapeutic treatments via small molecules. State-of-the-art single-cell RNA-sequencing approaches allow automatic collection of RNA content from single-cells but require a starting sample of at least 10³-10⁴ cells. Single-cell transcriptional profiling enables innovative diagnostic and therapeutic strategies for cancer patients that consider tumour intrinsic heterogeneity. Reverse-engineering approaches can exploit this heterogeneity to identify novel druggable targets and biomarkers of drug sensitivity. Here we measured gene expression profiles (GEPs) at the single cell level (scRNA-seq) for more than 40,000 cells from 33 distinct breast cancer cell lines (CCLs) including all the known subtypes, and for primary cancer cells from 7 patients. We detected pathways specifically active in a cell line from scRNA-seq by a novel measure of gene co-expression based on Rényi Multi-Information and identified genomic mutations predictive of pathway activity. Finally, by coupling existing data on drug sensitivity in CCLs with scRNA-seq, we successfully predicted drug sensitivity to 485 anticancer agents at the single cell level using exclusively gene expression data.

Perspectives in the field: One of the main challenges in the field of post-genomic research is developing methods to extract information from the vast amounts of data generated by High Throughput technologies. The aim is to elucidate biological mechanisms at the 'network' level, i.e. how genes, proteins non-coding RNAs and metabolites interact with each other to perform a specific function. Indeed, Systems Biology approaches can be used to develop novel pharmacological treatments for rare genetic diseases by identifying the most suitable compounds with the highest therapeutic efficacy.

Single cell genomics: using finer lenses to unravel features of human immunity

Alexandra-Chloé Villani, Massachusetts General Hospital, Boston, USA and Broad Institute of MIT and Harvard, Cambridge, USA

Summary of talk: Achieving a detailed understanding of the composition and function of the human immune system at the scale of the fundamental unit of life — the cell — is essential to determining the prerequisites of health and disease. Historically, leukocyte populations have been defined by a combination of morphology, localization, functions, developmental origins, and the expression of a restricted set of markers. However, these strategies are inherently biased and are often inadequate. Single-cell genomics analyses now provide an unbiased, data-driven way of systematically detecting cellular states and subtypes, and can reveal diverse facets of cellular identity, from discrete cell types to continuous dynamic transitions, which may be challenging to define by a handful of pre-defined markers since key markers may not be included or even known. We combined single cell multi-'omics' strategies together with in-depth follow-up profiling, phenotypic characterization, and functional studies of prospectively isolated human immune cell subsets, as defined by single-cell RNA sequencing data, to overcome such limitations. Our analyses revealed the detailed biological landscape of some human immune cell populations, enabling, for example, the discovery, re-classification, and characterization of several novel subsets of the mononuclear phagocyte system in health and disease.

Perspectives in the field: This strategy reopens the definition of cell type, allowing for a more sophisticated and complete view of a cell, as well as for unraveling the cellular context of genetic susceptibility variants. Such an approach, which forms the basis for constructing a comprehensive human immune cell atlas, was discussed at the meeting, together with progress achieved towards completing a draft of the human blood atlas and examples of applying this strategy to the study of immune disorders. Collectively, such revised cell taxonomy will enable more accurate functional and developmental analyses as well as immune monitoring in health and disease.

Short talk: miRAMINT, modeling RNA-seq time-series data through variable selection and surface matching narrows the implication of miRNAs in the brain of Huntington's disease knock-in mice

Lucile Mégret, Institut de Biologie Paris-Seine, Brain-C Lab, Paris, France

Summary of talk: MicroRNA (miRNA) regulation is believed to play a significant role in modulating Huntington's disease (HD) pathogenesis. This possibility is supported by studies of individual miRNAs and by system level studies of RNA-seq time series data in the allelic series of HD knock-in mice. I revisited this question by using miRAMINT, a pipeline that combines network, tree-based and plane-matching analyzes for modelling the CAG-repeat-length- and age-dependent dynamics of miRNA regulation at high precision. I showed that MiRAMINT retained a small number of CAG-length- and time-dependent miRNA-mRNA pairs in striatum, including miRNAs and mRNAs previously associated with neuronal development and maintenance and with HD pathogenesis. Data prioritization emphasizes striatal miRNAs of particular interest, e.g. those paired with high-amplitude change of target mRNA levels. The miRNA-mRNA pairs highlighted by miRAMINT differ from those reported in previous bioinformatics analyzes of miRNA regulation in the mouse or human HD brain. These findings strongly suggest that miRNA regulation may have a limited global role in regulating the dynamics of gene expression in the brain of HD mice.

Perspectives in the field: These findings provide a precisely-built resource, that is shape-matched explanatory miRNA-mRNA relationships implicated in neuronal integrity, neurotransmission and cell survival, for future studies in which to investigate how the mouse striatum may dynamically use miRNAs to compute responses to HD. MiRAMINT is part of BioGemix, a framework that we develop in our lab for precision machine-learning studies in aging and neurodegenerative disease.

Short talk: Controlling directed protein Interaction networks in cancer

Krishna Kanhaiya, Abo Academy University, Finland

Summary of talk: My area of interest is computational systems biology and bioinformatics which involves, network pharmacology, cancer systems biology, computational modeling of complex biological systems, analysis of biological network, construction of bioinformatics data analysis pipeline, database search, biomedical data analysis, machine learning in biology, statistical methods for biomedical research, and data visualization. The main message of my talk/research is that target controlling of a set of cancer-specific essential proteins is possible if it has known direct interactions to a drug target proteins and it can be driven from initial(drug targets) state to any desired final (essential proteins) state in finite time. Therefore, instead of trying to achieve full control of entire cancer's network, which in itself is highly complex, our approach aims for a targeted or partial control approach, particularly for controlling those cancer essential proteins. Our analysis showed that to control all of the essential proteins in these cancer networks, we require direct intervention over only 6.6% - 13% of the entire networks' nodes. In turn, to achieve full control of these networks, it required around 70% of the networks' nodes to be directly controlled by outside intervention. Thus, our method generates up to a 10-fold decrease in the control effort, while maintaining a high likelihood of an overall similar effect.

Perspectives in the field: My work is based on structural target controllability of the cancer networks. Here, I develop a novel and efficient approaches for the target controllability of cancer networks and demonstrate it for the analysis of various types of cancer, and identified the respective sets of driver (drug-target) proteins for controlling the networks. So, it's a novel example of how the network science and modeling approach can be successfully implemented in biological systems.

Session V: Network approaches for complex biological systems

Chair: [Claudia Manzoni](#), University of Reading, Reading, UK

Protein-protein interaction networks to interpret functional genomics

Claudia Manzoni, University of Reading, UK

Summary of talk: The field of genetics has seen an incredible growth during the past 15 years and we have now reached a point where whole-genome genetics (both genotyping arrays and sequencing technology) is relatively fast and reasonably cheap to be performed. However, genetics is not sufficient by itself. In order to provide meaningful insights and advance the field of biomedicine, genetics has to be translated into mechanistic knowledge, verified and applied to the functional setting.

This is not always possible, thus creating a bottleneck that prevents rapid intervention for disease diagnosis and development of therapeutic measures. There are different reasons for this difficult translation of genetics into functional knowledge. Among those, and of particular relevance in the case

of complex neurodegenerations, is the fact that familial forms of disease are often associated with a plethora of different genes that, when mutated, are able to trigger the complex neurodegeneration. It is difficult to model multiple, different genes in association with the same disorder through classical approaches. This is why we have implemented a more holistic disease modelling approach, using protein interaction networks, to isolate the similarities (in terms of pathways and functions) across the genes involved in the same complex neurodegeneration. We build multiple layers interactomes around the genes of interest (mutated in disease) and we merge those interactomes in a unique network, isolating the nodes that are responsible for keeping the cohesion of the graph. We have demonstrated that these nodes are the key element to identify shared pathways and functions across the mutated genes to identify communal mechanisms sustaining the disease phenotype.

Perspectives in the field: Genetics has to be translated into testable hypothesis for disease modelling, this would be possible when more holistic and network-based approach will be applied to large genomics datasets to collapse multiple genetic component of risk and evaluate their cumulative effect. We are in the need of shifting the perspective from the “one-gene : one-disease” to the “one-pathway : one-disease” approach. Similarly, as researchers, we will be more and more exposed to the need of communicating across different fields of expertise. We therefore need to disclose the assumptions that are hidden behind the generation and interpretation of our data, particularly to facilitate across field discussion and collaborations.

ncRNA networks in cancer

Pavel Sumazin, Baylor College of Medicine, Houston, USA

Summary of talk: Long-noncoding RNA (lncRNA) are differentially expressed across tissues (Guo et al., 2009) and commonly dysregulated in tumors (Yan et al., 2015). Evidence from thousands of GWAS and eQTL studies suggests that genetic variants at the loci of hundreds of lncRNAs are associated with human diseases (Hon et al., 2017). However, while improving profiling technologies have helped to considerably improve our understanding of lncRNA biology, the catalog of human lncRNAs is still incomplete, and the consequences of lncRNA dysregulation in human diseases are known for only a few lncRNA species (Huarte, 2015). Similarly, the role of lncRNAs in each tissue and the pathophysiological consequences of their differential expression and dysregulation in cancer are known for only a handful of lncRNAs (Yan et al., 2015). Their ranks may include key genes for tissue maintenance and differentiation, as well as diagnostic biomarkers and therapeutic targets for a variety of diseases and should be systematically explored. We have developed models for predicting lncRNA targets that can be populated using RNA-Seq profiles of diseased and normal tissues. Focusing on cancer, we predicted lncRNA targets based on these models and inferred function for lncRNAs that target key cancer genes and pathways. To date, we have characterized and assigned cancer-specific roles to half a dozen lncRNAs, identifying lncRNAs that act as tumor suppressors and oncogenes, and lncRNAs whose targeting alters cancer cell and tumor response to cancer therapy.

Perspectives in the field: We argue that the lncRNA space includes multiple predictive biomarkers and therapeutic targets that could be used to identify and target cancer cells with high selectivity and low toxicity. We envision that studies into regulatory networks that focus on lncRNAs and their target cancer genes and pathways will have numerous translational applications in cancer biology.

References:

- Guo, J., Miao, Y., Xiao, B., Huan, R., Jiang, Z., Meng, D., & Wang, Y. (2009). Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. *Journal of gastroenterology and hepatology*, 24(4), 652-657.
- Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J., Gough, J., Severin, J. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644), 199-204.
- Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nature medicine*, 21(11), 1253-1261.
- Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S. D., He, Q. (2015). Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell*, 28(4), 529-540.

Modelling and control of pluripotency networks

Lucia Marucci, University of Bristol, Bristol, UK

Summary of talk: Cellular functions, including cell-decision making, are orchestrated by the dynamic interplay among gene network regulations, signalling pathways activity and the extracellular environment, giving rise to highly complex gene expression patterns. In this talk, I discussed Systems and Synthetic biology approaches to understand and control gene networks regulating pluripotency and differentiation of mouse Embryonic Stem Cell (mESCs). I showed results from Stochastic Differential Equation models formalising interactions among master pluripotency regulators, which can explain the emergency of pluripotency regulators' heterogeneity in mESC isogenic populations as a result of bistability and stochasticity. I also discussed limitations of such models: uncertainties in model

structure/parameters, or in gene network interactions, can lead to incorrect predictions. I then introduced alternative Synthetic Biology approaches, which combine microfluidics/microscopy platforms with closed loop strategies to automatically regulate and control gene expression in mammalian cells. I showed the feasibility of controlling, via fluorescence reporters, inducible genes at both transcriptional and post-translational levels to reach and maintain a target expression (e.g. 50% of maximal expression) in 30 hours time-lapse control experiments. The platform enables control of both exogenous genes and endogenous signalling pathways (e.g. canonical Wnt pathway) in various mammalian cell lines.

Perspectives in the field: I believe that automatic feedback control and microfluidics/microscopy platforms, allowing precise regulation of cellular dynamics, fast modulation of cell media and environment as well as collection of single-cell imaging data, can in the future be exploited to: i) infer gene network regulations by tagging relevant genes in pathways of interest; ii) understand bifurcations giving rise to non-linear gene expression dynamics even in absence of a mathematical model of the system, and iii) facilitate the design of combination therapy strategies in cancer research.

Short talk: Yeast Complexome - The Complex Portal rising to the challenge

Birgit Meldal¹, Livia Perfetto¹, Edith Wong², Sandra Orchard¹, Henning Hermjakob¹, Pablo Porras¹
¹European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

²Department of Genetics, Stanford University, 3165 Porter Drive, Palo Alto, CA 94304, USA.

Summary of talk: Protein complexes carry out almost the entire signaling and functional processes in the cell. Conventional approaches to identify complexes include experimental methods based on quantitative proteomics studies and in-silico methods that consider protein complexes as densely connected regions within interaction graphs (mainly protein-protein interaction networks). However, these methods have inherent limitations and require a gold standard.

The Complex Portal (www.ebi.ac.uk/complexportal) is an encyclopaedia of macromolecular complexes from 20 model organisms that is manually curated based on literature evidence. For more than 2700 complexes (October 2019) it provides details about protein composition, structure, function (as Gene Ontology annotations) and roles in pathways and diseases. In 2018 we completed the first draft of all *Saccharomyces cerevisiae* complexes, also known as the 'yeast complexome' covering approx. 30% of the yeast proteome. The list of yeast known and predicted complexes was compiled in close collaboration with Saccharomyces Genome Database (SGD) and UniProt curators. The 'yeast complexome' contains nearly 600 well defined complexes including super-structures like polymerases and ribosomal processomes where complexes act as participants of larger complexes. This dataset represents a new gold standard to validate complexes predicted by *in silico* methods. We compared the 'yeast complexome' from the Complex Portal to the previous gold standard, the CYC2008 dataset (Pu *et al.* 2009) containing 408 manually curated complexes. We have compared the overlap of participants between complexes in both datasets using three Jaccard Index (JI) thresholds: 0.5, 0.75 and 1.0. The Complex Portal shows a great overlap with and offers more coverage than the CYC2008 dataset: the two datasets share approx. 500, 350 and 280 complexes at Jaccard Indices 0.5, 0.75 and 1.0 respectively. Secondly, we tested both manually curated datasets to validate complexes predicted by topological analyses of IntAct network and complexes in the YHTP2008 dataset, predicted from large-scale experiments (Pu *et al.* 2009). The overlap between manually curated and predicted complexes is very low/small (<50 complexes) at any JI threshold.

Perspectives in the field: Complex Portal as a gold standard opens opportunities for the development of more sensitive and sophisticated prediction tools for identification of complex-complex interactions, regulatory interactions and predictions of molecular functions at system level. Additional complexomes are in preparation, including human, mouse and *C.elegans*, with that of *Escherichia coli* scheduled to be completed in 2019.

Short talk: Effect of mechanical compression on the gene expression profile of human bronchial epithelial cells

Margherita De Marzio, Harvard, Boston, USA

Summary of talk: Asthma is the most common chronic airway diseases and affects more than 300 million children and adults worldwide. Despite the advancements in the identification of the potential causes inducing this complex disorder, the molecular mechanism triggering its onset and driving its development is still unknown. Recent studies have highlighted the role of compressive forces on the airway epithelium as a potential co-factor in the activation and chronicity of this disease. Due to bronchoconstriction, asthmatic bronchi are narrower than healthy people and the epithelial tissue is exposed to excessive mechanical compression resulting from the airflow obstruction. Data on adults with atopic asthma and additional in vivo and in vitro experiments have revealed that increased epithelial compression can induce airway wall remodeling, a structural alteration of the airway

epithelial tissue typical of chronic asthmatic patients, in the absence of any inflammatory process and that pressure can increase the production of transforming growth factor β (TGF β) and endothelin, two regulatory mediators whose concentration is enhanced in asthmatic airways.

These results strongly suggest that imbalances in airway pressure may have a key role in the initiation of the molecular processes leading to the asthmatic phenotype.

In this study, we went one step further towards understanding the molecular changes induced by mechanical stimuli. In collaboration with Prof. Jeffrey Fredberg and Prof. Jin-Ah Park at the Harvard T. H. Chan School of Public Health, we exposed healthy human bronchial epithelial cells (HBECs) to compression with 30 cm H₂O pressure for 3 hours and we determined their post-compression expression profile using RNA-Sequencing.

By combining differential expression analysis, pathway overrepresentation test and network approach, we investigated the transcriptional alterations induced at two time points, immediately after releasing pressure and 21 hours later, and we used a betweenness centrality measure called “flow centrality” to identify the potential genes driving the cascade processes that connects these time points. Our results suggested that compression activates a two-step cellular response: an immediate but transient response involving membrane signaling proteins, and a long-time, steady-increase, effect on genes associated to cell shape, extracellular matrix reorganization and tissue remodeling. More strikingly, we compared the transcriptome of compressed HBECs with RNA-Seq expression profile of asthmatic HBECs and we found that the transcriptional response at 21 hours post-pressure resembles the genetic signature of asthmatic airway remodeling.

Finally, by using a protein-protein interaction (PPI) network, flow centrality was able to determine the genetic bottlenecks of communications between the molecular perturbations occurring at the two time points and it detected genes involved in cytokine production, wound healing and collagen production as the key mediators of compression-induced transcriptional changes.

Perspectives in the field: Growing experimental evidences support the idea that asthmatic bronchoconstriction is not only triggered by inflammatory response but it itself acts as a trigger for asthmatic inflammation and tissue remodeling. Confirmation of a key role of mechanotransduction in the pathogenesis of asthma would imply that the traditional picture of asthma as a “cause and effect” disease starting from the allergic response and ending with chronic inflammation and respiratory difficulties would give way to a feedback loop process where inflammation, remodeling and bronchoconstriction are all equal contributors to the activation and perpetuation of the disease.

While a tremendous amount of experimental and computational work has been already done, two main future strategies could improve significantly our understanding of the biological mechanisms involved in the asthmatic syndrome: a comprehensive in-vivo study of human bronchial epithelial tissue in compressed healthy and asthmatic airways, that would rule out potential artifacts coming from in-vitro cultured studies, and RNA-sequencing at the single-cell level of the airways epithelium, a technology that would allow to discern eventual differences in the transcriptional changes of the multiple cell types composing the tissue, such as ciliated, goblet and basal cells. Furthermore, in terms of clinical applications, these studies could provide new contributions for the development of targeted drugs inhibiting the specific molecular components driving asthma development and exacerbation. Indeed, a huge amount of clinical studies have already shown that the synthetization of drugs.

Session VI: Network approaches in brain system biology

Chair: [X. William Yang](#), UCLA, CA

Assembling global neural networks of the mouse brain

[Hong-Wei Dong](#), [Mark&Mary Stevens](#) *Neuroimaging & Informatics Institute, Keck School of Medicine, USC, USA*

Summary of talk: The major missions of the BRAIN Initiative is to map the interconnections among all regions of the C57Bl/6 mouse brain, to generate a corresponding comprehensive connectome map that represents the interconnections in a common neuroanatomic frame, and to understand how the different brain regions assemble into functional networks based on these connections. The biological significance of assembling a brain-wide wiring diagram is tantamount to that of the Human Genome Project. Our Mouse Connectome Project (MCP, www.MouseConnectome.org) is on the frontier to carry on this mission. In my talk, I highlighted the general methodological approach of our project and reported the tremendous progress that we have made in the last 10 years. In short, we have systematically constructed neural networks of the entire neocortex, basal ganglia, hippocampus, amygdalar, and other forebrain areas.

Perspectives in the field: We anticipate to completing our construction of the whole brain wiring diagram in the next 5 years. Meanwhile, as part of the BRAIN Initiative Cell Census Network (BICCN, biccn.org), we have begun to tackle another major challenge to catalog neuronal cell types in the

mouse brain based on their anatomical locations, connectivity, molecular identities and physiological properties. The ensuing challenge is to analyze the vast connectomics and cell type connectivity information in a way that is most conducive to generating novel behavioral hypotheses for experimental testing.

Multilayer network integration and spatial reconstruction of brain circuits/networks *Manlio De Domenico, CoMuNe Lab, Fondazione Bruno Kessler, Trento, Italy.*

Summary of talk: Understanding how the human brain is structured, and how its architecture is related to function, is of paramount importance for a variety of applications, including but not limited to new ways to prevent, deal with, and cure brain diseases, such as Alzheimer's or Parkinson's, and psychiatric disorders, such as schizophrenia. The recent advances in structural and functional neuroimaging, together with the increasing attitude toward interdisciplinary approaches involving computer science, mathematics, and physics, are fostering interesting results from computational neuroscience that are quite often based on the analysis of complex network representation of the human brain. In recent years, this representation experienced a theoretical and computational revolution that is breaching neuroscience, allowing us to cope with the increasing complexity of the human brain across multiple scales and in multiple dimensions and to model structural and functional connectivity from new perspectives, often combined with each other. In this talk, I review the main achievements obtained from interdisciplinary research based on magnetic resonance imaging and establish de facto, the birth of multilayer network analysis and modeling of the human brain.

Perspectives in the field: Despite the deluge of available data sets, from healthy subjects to patients affected by neuro-degenerative diseases and psychiatric disorders, and the many results obtained from existing analytical approaches, still much work is needed. In fact, methods for the (functional) reconstruction of brain circuits suffers from several statistical flaws which requires novel ideas and groundbreaking methodologies to be solved. More reliable reconstructions – based, for instance, on generative models including spatial information and geometry – and integration of multivariate information will deepen our understanding of the human brain in life and disease.

Short talk: PINOT: A transparent data mining pipeline for mapping confidence-weighted protein-protein interaction networks

James E. Tomkins, UCL, UK

Summary of talk: Ongoing efforts to detect protein-protein interactions (PPI) has resulted in an abundance of PPI data which are actively curated into numerous repositories, such as IntAct and BioGRID. Considering this wealth of data collectively to gather insight into protein interaction profiles in the wider cellular context is challenging, largely due to dataset inconsistencies.

We have recently developed a PPI query resource, named Protein Interaction Network Online Tool (PINOT), which collates data from several repositories at the time of query and processes these data through a systematic pipeline of quality control steps and confidence scoring procedures. Each entry is assigned a confidence score based on the number of distinct methods used for PPI detection and the number of publications that report the interaction. In addition, this information is fully traceable back to the source publication.

This straightforward-to-use tool adds value to the existing toolbox of PPI query resources and facilitates the construction of PPI networks. It provides a route for rapidly surveying literature-derived PPI data for proteins of interest which can form the foundation for further analyses. PINOT has been applied to numerous different research questions, including interactomic studies of structurally related and disease related proteins, in order to decipher commonality and distinction in interaction and functional profiles.

Perspectives in the field: Decades of evidence based PPI data are readily available within the public domain. These data are likely to hold promise for building a holistic view of the subcellular environment and understanding this dynamic network of molecules in the context of biological functions. In addition, these data are useful for the interpretation of novel datasets and for prioritising PPIs for follow-up studies, enabling a streamlined transition of *in silico* to 'wet lab' analysis. This approach enhances time and cost efficiency in the relation to validating PPI hits from a high-throughput screen, for example. To prevent neglect of these literature-derived data and to optimise their utility, PPI query resources, such as PINOT, are of notable value to the research community.

Transcriptional networks specific to human cognition

Genevieve Konopka, UT Southwestern Medical Center, Dallas, Texas, USA

Summary of talk: I discussed the use of comparative genomics across primate brain tissue to identify signatures of genes expression specific to human brains. By applying weighted gene co-expression network analyses (WGCNA) to these datasets, hub genes within human-specific modules can be

prioritized for functional follow-up studies. I presented one such example where the transcription factor CLOCK was a top hub gene in a human frontal pole specific module. I then went on to show that manipulating CLOCK expression in human neurons led to changes in gene expression consistent with an extra-circadian role for CLOCK as well as the regulation of additional genes with human-specific expression in the brain. I then discussed the importance of a cell type-specific approach for gene expression profiling across primate brains. I demonstrated how bulk tissue approaches have been underpowered to detect certain cell types. New methods that use single-cell expression datasets and deconvolution can be applied to previous bulk datasets to illustrate this point. I also showed how the integration of human disease gene variants and expression datasets with the comparative primate datasets can uncover how certain cell types that have evolved in the human brain are at risk in neuropsychiatric disorders.

Perspectives in the field: Future directions for these approaches include using single-cell profiling in comparative primate studies as well as in tissue from patients with various brain disorders. One of the challenges for the field is to connect various levels of gene regulation: epigenetics, alternative splicing, post-translational modifications etc. at a cell type level in the developing and diseased brain. Another challenge is to determine how to prioritize and functionally follow-up on human variant or expression data in model systems.

Translational Profiling of Neuronal Cell-type-Specific Transcriptomics During Aging and in Age-dependent Diseases

Myriam Heiman, MIT, Picower and Broad Institutes, Cambridge, USA

Summary of talk: In my presentation, I described my group's recent efforts to perform cell type-specific mRNA transcriptional and translational profiling in affected brain regions in Huntington's disease and mouse models of Huntington's disease. For transcriptional profiling we have applied new single nuclear RNA-Seq technologies, and for translational profiling we have used a cell type-specific tagged ribosome approach (TRAP). These studies have revealed both conserved and divergent responses of different cell types to mutations in Huntingtin, the Huntington's disease gene. Among the most vulnerable cell types in Huntington's disease, striatal medium spiny neurons, we implicate new pathways that point to novel therapeutic strategies for Huntington's disease and demonstrate a large degree of conservation between human and mouse striatal neuron cell type-specific responses to mutant Huntingtin.

Perspectives in the field: Future directions for us, and the field in general, will be to use both new systems-level and large scale genetic screening and perturbation approaches to analyze these large sets of correlative data generated by technologies such as single cell and single nuclear RNA-Seq in order to predict which molecular changes are most important to the mechanisms driving disease biology.