# 4th Disease Maps Community Meeting

Sevilla, 2-4 October 2019

INB · elixir SPAIN · FEDER Fondo Europeo de Desarrollo Regional · UNIÓN EUROPEA "Una manera de hacer Europa" · GOBIERNO DE ESPAÑA · MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES · Instituto de Salud Carlos III · JUNTA DE ANDALUCIA · Fundación Progreso y Salud · CONSEJERÍA DE SALUD Y FAMILIAS · Hospital Universitario Virgen del Rocío · systems medicine disease maps

# Wednesday 2nd October

### 09:00 Disease Maps Members/Coordinators Meeting

### 10:30 Coffee Break

---

## Session 1: Using existing data to create disease maps

### 11:00 Welcome

*Joaquin Dopazo, Clinical Bioinformatics Area, FPS, Sevilla, Spain*

### 11:10 Opening Key Lecture 1

Analysis of multi-omics data to improve the understanding of Parkinson's disease
*Inge Jonassen, University of Bergen, Norway*

### 12:10 Short talks

Towards automatic assembly and management of evolving disease models modularisation, merging, layout, comparison and versioning
*Alexander Mazein, University of Luxembourg, Luxembourg*

Common disease mechanisms at play? A modelling challenge
*Rudi Balling, University of Luxembourg, LCSB, Luxembourg*

Disease map for aHUS – molecular mechanisms and hypotheses
*Tatiana Serebriyskaya, RacursBioMed Ltd., Russia*

A disease map and a database for cystic fibrosis
*Catarina Pereira, University of Lisboa Faculty of Sciences, BioISI – Biosystems and Integrative Science Institute, Lisboa, Portugal*

### 14:00 Lunch and Poster viewing

---

## Session 2: Computational models

### 15:30 Key Lecture 2

Toward whole-cell computational models for precision medicine
*Jonathan Karr, Icahn School of Medicine at Mount Sinai, New York, USA*

### 16:30 Short talks

A multiscale modelling platform to simulate drug synergies in different cell population architectures.
*Miguel Ponce de Leon, Barcelona Supercomputing Center, Barcelona, Spain*

The metabolite abundances in coherence with the activities of signalling pathway circuits.
*Cankut Cubuk, Clinical Bioinformatics Area, FPS, Sevilla, Spain*

Executable Disease Maps – Addressing the challenges of large scale dynamical modelling
*Anna Niarakis, GenHotel, University of Evry, Paris-Saclay, France*

## *Thursday 3rd October*

### Session 3: Multi-view data integration

**9:00 Key Lecture:**

Logic modelling to integrate disease maps and various omics data
*Julio Saez-Rodriguez, RWTH-Aachen University Hospital, Aachen, Germany*

**10:00 Highlighted talk**

Elixir Sponsor talk

**10:30 Coffee break and poster viewing**

---

**11:00 Selected short talks**

Atlas of Cancer Signaling Network: a resource of multi-scale biological maps to study disease mechanisms
*Luis Cristobal Monraz Gomez, Institut Curie, France*

Interactive visualization of phenotypic and genotypic information in disease maps for identification of mechanistic biomarkers
*Sascha Herzinger, University of Luxembourg, Luxembourg Centre for Systems Biomedicine, Luxembourg*

Reactome Pathway Analysis and Visualization
*Henning Hermjakob, EMBL-EBI, United Kingdom*

**12:30 Poster session**

Flash presentations

**13:30 Lunch and poster viewing**

---

### Session 4: Discussions

**15:00 Introduction**

**15:30 Discussion topics**

Four discussion sessions (Topics to be determined)

**17:30 Presentation preparation and final discussion**

**18:30 Results presentation**

**20:30 Gala Dinner**

## *Friday 4th October*

**Session 5: Clinical application of disease maps**

**09:00 Key Lecture 4**

Computational approaches to tackle chemoresistance in high-grade serous ovarian cancer

*Sampsa Hautaniemi, Faculty of Medicine, University of Helsinki, Finland*

**10:00 The DisGeNET platform's 10 anniversary**

Developing a knowledge base for disease genomics

*Janet Piñero, GRIB (IMIM-UPF), Spain*

**10:30 Coffee break**

---

**11:00 Short talks**

Machine learning and mechanistic models for drug repositioning in rare diseases

*Marina Esteban-Medina, Clinical Bioinformatics Area, FPS, Sevilla, Spain*

Connecting metabolic biomarkers with biological pathways and clinical data to enable omics data interpretation

*Denise N. Slenter, Maastricht University, Netherlands*

Generation of process diagrams for computational modelling: lessons from the classroom

*Tom C. Freeman, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, UK*

**13:00 Poster viewing**

**13:00 Lunch**

---

Hands-on session

*Tutorials of the following tools and resources*

**Tutorials of the following tools and resources**

**14:30 Newt**

**14:30 BiNoM, ACSN**

**15:30 MINERVA**

**15:30 Hipathia**

(Depending on computers and rooms available, will be distributed in one or two parallel hands-on sessions)

**17:00 Meeting closure**

# Abstracts

## Towards automatic assembly and management of evolving disease models: modularisation, merging, layout, comparison and versioning

*Alexander Mazein[1], Anna Niarakis[2], Marek Ostaszewski[1], Inna Kuperstein[3], Jan Hasenauer[4] and Andrei Zinovyev[3]*

*University of Luxembourg[1], Univ Evry, Université Paris-Saclay[2], Institut Curie, Paris, France[3], University of Bonn, Germany[4]*

**Author keywords:** Disease mechanisms, systems biomedicine, knowledge representation

Using our collective experience in the development of resources for asthma, rheumatoid arthritis, Parkinson's disease, cancer and other diseases, we would like to discuss challenges related to the development of high-quality representations of disease mechanisms in such a way so they are reproducible, executable as computational models and modularised to reusable and easily-updatable blocks. We outline approaches and technologies for enabling the conversion between standard formats, automatic compatibility check, merging, comparison of similar subpathways and versioning in case more than one version of a cell-specific adaptation is needed. With the importance of readability and accessibility by humans, automatic layout algorithms are identified as one of the main obstacles for future progress in the field.

## Common disease mechanisms at play? A modeling challenge

*Rudi Balling and Feng He*

*LCSB - Luxembourg Centre for Systems Biomedicine*

**Author keywords:** energy metabolism, Chronic diseases, Common mechanisms

Life expectancy has increased during the last century, and so has the disease spectrum. Whereas 100 years ago, acute and chronic infections were the main cause of morbidity and mortality, we are now witnessing a dramatic increase in chronic, age related diseases. In addition to diabetes, allergies, cardiovascular disease and cancer, we also see an increase in the prevalence of neurodegenerative and neuropsychiatric diseases. Whereas this increase is thought to be a result of our sedentary life style and a hypercaloric nutrition, the detailed molecular mechanisms are not well understood. We have identified the function of a familial Parkinson disease gene as an important player in controlling central energy metabolism. Based on this insight a hypothesis is presented that links the energetic and biosynthetic constraints and contributions of glycolysis and oxidative phosphorylation to the pathogenesis of a wide range of different chronic diseases and their comorbidities. A proposal is made for a community driven project to carry out a detailed cross-disease comparison of the regulation and function of the central metabolism. The disease map community could play an important role to ensure the rapid integration of the obtained results into the various disease maps and as a forum for further hypothesis generation and validation.

## Disease map for aHUS – molecular mechanisms and hypotheses

*Tatiana Serebriyskaya and Olga Migulina*
*RacursBioMed Ltd, Russia*

**Author keywords:** Rare diseases, Atypical hemolytic uremic syndrome (aHUS), Pathways perturbations

Insight of molecular mechanisms underlying rare diseases provides knowledge about the normal function of the disease-associated genes, increases number and accuracy of genetic tests, improvements in diagnosis and development of new gene-targeted treatments. Disease-associated signaling pathway maps allow to represent detailed molecular mechanisms of disease supporting investigation of pathogenesis and suggesting possible ways of therapeutic intervention.

Atypical hemolytic uremic syndrome (aHUS) is rare hereditary disease which characterized by the triad of microangiopathic hemolytic anemia, thrombocytopenia, and acute kidney injury. aHUS is heterogenous disease that is caused by dysregulation of complement pathway. Pathogenic mutations identified in complement genes (C3, CD46/MCP, CFB, CFH, CFHR1, CFHR3, CFHR5, CFI, DGKE, and THBD) can cause the disease in over 50% of aHUS patients [PMID: 24161037; PMID: 25608561; PMID: 20556434]. But involvement of other genes and pathways which regulate complement system (such as coagulation pathway) was also discussed in disease pathogenesis [PMID: 24029428; PMID: 30377230].

Previously we've reported about new manually curated comprehensive database for aHUS-associated gene variants [DOI: 10.29245/2572-9411/2018/1.1168].

**Objectives:** aim of this work is to develop aHUS-specific disease pathway map which represents aHUS-associated perturbations in molecular signaling pathways. Additionally, we incorporate into map information about key pathogenic genetic variants, disease-associated changes in protein activities and levels and other etiologic triggers.

**Methods:** disease-associated pathway map is created manually based on analysis of scientific publications about signaling pathways affected in aHUS. The map was created in SBGN format that allows to reconstruct dysregulated pathways in detail, and in Cytoscape format (binary interactions) which made it available for the network analysis methods.

**Results:** in this work we've developed a novel approach to represent possible mechanisms of aHUS pathogenesis and different types of aHUS-associated data in large comprehensive disease pathway map. Different map formats make it available as for analysis and interpretation of existing experimental and clinical data, and for proposing further hypotheses of disease development and possible treatment options.

## A disease map and a database for cystic fibrosis

*Catarina Pereira[1], Alexander Mazein[2], Margarida Amaral[1] and Andre Falcao[3]*

*BioISI – Biosystems and Integrative Science Institute, Lisboa, Portugal1, University of Luxembourg, Luxembourg Centre for Systems Biomedicine2, LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal[3]*

**Author keywords:** disease map, cystic fibrosis, database, SBGN, MySQL, CFTR, F508del

**Background:** Cystic fibrosis (CF) is a life-threatening autosomal recessive disease caused by more than 2,000 mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene, generating CF disease phenotypes. It is still unclear how mutations in a single gene generate variability in symptoms and disease severity in individuals with CF. The application of bioinformatics and systems biology tools in the integration and representation of CF heterogenous data help interpret its biological significance and the finding of new therapeutic targets.

**Aim:** this work aims at presenting the CF-MAP, a comprehensive disease map summarizing the biological pathways in which CFTR protein is a key player, and the CF database (CFdb), which provides a way to access additional information for each interaction present on the map.

**Methods:** The CF-MAP was constructed with the yEd software and is represented in an activity flow language accordingly to Systems Biology Graphical Notation. The knowledge represented involves extensive manual curation of literature and participation of CF experts. The CFdb was modelled as a relational database and implemented using MySQL and the website was implemented using PHP.

**Results:** this map integrates major wt- and F508del-CFTR traffic pathways including endoplasmic reticulum (ER) quality control (ERQC), ER-associated protein degradation (ERAD) machinery, stabilization and activation of CFTR at the membrane and its endocytosis to degradation or recycling. The CFdb is based on experimental and 'omics' data from the literature.

**Conclusions: the database (CFdb) and the disease map (CF-MAP) will turn CF knowledge accessible and exchangeable to researchers, allowing analysis at the molecular, functional and physiological levels of CF and possible novel treatment strategies. These resources will be openly available on the web, aiming at being a reference CF data repository with comprehensive user input capabilities for continuous updating.**

## A multiscale modelling platform to simulate drug synergies in different cell population architectures

*Miguel Ponce de Leon[1], Arnau Montagud[1], Miguel Vazquez[1], Åsmund Flobak[2] and Alfonso Valencia[1]*
*Barcelona Supercomputing Center, Spain[1], Norwegian University of Science and Technology[2]*

**Author keywords:** multiscale modelling, agent-based modelling, Boolean modelling, drug synergy, cáncer

Understanding the mechanisms of cancer resistance is key to improve clinical treatments and ultimately extend patients life expectancy [1]. The emergence of resistant cells is a complex phenomenon in the intersection of different scales; molecular, cellular and inter-cellular interactions, in an environmental architecture that directly effects response to drugs [2–4].

Multiscale simulations can help study this phenomenon by integrating models of processes that take place at these different scales. We are developing a multiscale modelling framework combining agent-based and Boolean models, to cover from genes' activity and cells' phenotypes to physical interactions among cells and with their environment. The multiscale models provide a genotype-to-phenotype mapping framework, which allow the exploration of genetic variations, environmental conditions (e.g. presence of oxygen or signalling molecules or structure of the extracellular matrix), and variations of architecture.

We will present the current status of our multiscale model, with simulations of wild type cell growth and consequences of treatment with different drugs. Our modelling framework is based on PhysiBoSS [5], consisting in a stochastic Boolean model of cell fates [6] embedded in a flexible agent-based model of population dynamics [7]. Our results using an AGS-carcinoma-cell-specific Boolean model [8] allow us to identify combinations of drugs that avoid the resistances created by known cancer mutations in the model. These simulations are in good agreement with a set of experimental results, not used as priors.

Furthermore, we explored the effect of different cell population architectures in the response to combination of drugs that target signalling pathways by simulating cells in 2D plate-like monolayers and 3D organoid-like spheroids.

In summary, our initial results in this multiscale bottom-up simulation framework running in MareNostrum IV using HPC capacities, indicate that it is possible to explore the space of combinations of mutations, drugs and architecture in a given model of tumour to identify candidate mechanisms potentially responsible for the drug synergies leading to a reduced number of testable hypotheses.

References:

[1]. Kessler DA, et al. Cancer Res. 2014; 74:4663–4670
[2]. Kim E, et al. PLoS Biol. 2018; 16:e2002930
[3]. Horvath P, et al. Nat. Rev. Drug Discov. 2016; 15:751–769
[4]. Jabs J, et al. Mol. Syst. Biol. 2017; 13:955
[5]. Letort G, et al. Bioinformatics 2018; bty766
[6]. Stoll G, et al. Bioinformatics 2017; 33:2226–2228
[7]. Ghaffarizadeh A, et al. PLOS Comput. Biol. 2018; 14:e1005991
[8]. Flobak Å, et al. PLOS Comput. Biol. 2015; 11:e1004426

## The metabolite abundances in coherence with the activities of signaling pathway circuits

*Cankut Cubuk[1], Carlos Loucera[1], Marta R. Hidalgo[2] and Joaquin Dopazo[1]*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[1], Príncipe Felipe Research Center, Valencia, Spain[2]*

**Author keywords:** mechanistic models, networks, pathways, machine learning, functional annotations

The deviant amount of metabolites are the main reflectors for the metabolic irregularities that are behind the initiation and progression of cancer. However, because of the lack of knowledge on the functional roles of metabolites, it is still hard to understand their mechanistic interactions with cellular functions and to elucidate their oncogenic activities [1].

Most of the cell fate is governed by functional events triggered by the signaling pathways. It has been shown that mechanistic models of signaling pathways can infer the activity of the different elementary functionalities in the cell from gene expression measurements [2]. In order to understand how the different metabolites are related with the cell decisions we have used a method, Metabolica, which uses transcriptomic and/or genomic data to predict the activity of metabolic reactions and then propagates the metabolic flux over the metabolic hypergraph that accounts for the production of a metabolite [3]. The Metabolica is an extended version of Metabolizer [4], which has been successfully used to predict essential genes in cancer cell lines [5], The Metabolica calculates the relative metabolite abundances per sample, which can be used in a wide range of downstream analyses, and, specifically, for the annotation of metabolites with their potential functional roles. Using a large dataset of gene expression data (TCGA), we inferred an approximation to the causal relationship between the estimated metabolite abundances with Metabolica and the molecular mechanisms that are mediated by the activities of signal transduction circuits [2] applying multi-task learning in the context of Gaussian Processes (GP) [6]. Both the signaling and metabolic pathway methods used in this study input the same type of data, gene expression, but they apply slightly different propagation algorithms and use totally different pathway topologies and gene contents. Finally, we annotate the metabolites with the functionality triggered by the circuits whose activity they predict.

Here, we propose an approach that combines the results of mechanistic models of signaling and metabolic pathway. In this study, a total of 494 metabolites; 317 and 393 metabolites in kidney and breast carcinoma, respectively, were annotated [Figure 1]. The cellular functions were also grouped under ten cancer hallmarks [7] to uncover small-molecule dependencies of cancer cells.

References

[1]. Nowicki, S. and Gottlieb, E. (2015). Oncometabolites: tailoring our genes. FEBS Journal, 282(15), pp.2796-2805.

[2]. Hidalgo, M., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J. and Dopazo, J. (2016). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget, 8(3).

[3]. Cubuk C., Loucera C., Hidalgo M., et al. (2019) METABOLICA: A systems biology approach to prioritize METABOLItes in CAncer (manuscript in preparation).

[4]. Cubuk, C., Hidalgo, M.R., Amadoz, A., Rian, K., Salavert, F., and et al. (2018). Differential metabolic activity and discovery of therapeutic targets using summarized metabolic pathway models. npj Systems Biology and Applications, 5, 7.

[5]. Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., and et al. (2018). Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. Cancer Research, 78(21):6059–6072.

[6]. Rasmussen C.E. (2004) Gaussian Processes in Machine Learning. In: Bousquet O., von Luxburg U., Rätsch G. (eds) Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science, vol 3176. Springer, Berlin, Heidelberg

[7]. Hanahan, D. and Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. Cell, 144(5), pp.646-674.

## Executable Disease Maps – Addressing the challenges of large scale dynamical modelling

*Anna Niarakis*

*GenHotel, University of Evry, Paris-Saclay, France*

**Author keywords:** executable disease maps, large scale boolean models, automatic inference, rheumatoid arthritis

Building large scale dynamic models can be a tedious and time-consuming work that requires not only the construction of the regulatory graph but also the writing of the logical formulae.

In this work I will present CaSQ, a tool that will ease the construction of large scale Boolean models, taking advantage of the similarities that are shared between molecular interaction maps and dynamic models. First of all, the molecular maps are process description representations that are well referenced and annotated providing a critical source of knowledge. The maps also contain information about the interactions, catalyses, activations and inhibitions of the network, information that is essential for the building of a computational model. For the conversion of a molecular map to a Boolean model, we decided to utilise systems biology standards for both network representation (Systems Biology Graphical Notation: SBGN) and model construction (Systems Biology Marked Up Language-qualitative: SBML-Qual) so that CaSQ could be interoperable with other tools and modelling software. For the inference of the logical formulae, we based our assumptions on topology and semantics of the molecular maps. With the use of CaSQ, we can now obtain large scale Boolean models that can be executed using popular modelling tools such as GINsim and Cell Collective.

However, the problem of analyzing large scale Boolean models is real and constitutes a challenge in the field. For coping with size and complexity one can adopt a modularized approach or perform reductions and create different versions of the original model. I will use a state of the art molecular map for Rheumatoid Arthritis developed in our lab for showcasing the modularized approach.

We work closely with the team of Tomas Helikar for importing large scale SBML-qual models in Cell Collective that would retain layout and annotations and also with the team of Denis Thieffry for analysing stable states using GINsim. The goal is to propose a seamless pipeline for producing executable Boolean models starting from molecular interaction maps, which can be subsequently analysed in depth using various tools for computational modelling. CaSQ, the tool that we develop can play the role of a bridge, bringing together two distinct communities, curators and modellers, in an effort to produce models of better quality, accuracy and reusability.

References:

V. Singh et al. “CaSQ, a tool for automated inference of Boolean models from molecular interaction maps”. Under preparation, 2019a

V. Singh et al. “RA signalling map: A state of the art interactive knowledge base for Rheumatoid Arthritis”. Under preparation, 2019b

V. Singh et al. “Computational Systems Biology Approach for the Study of Rheumatoid Arthritis: From a Molecular Map to a Dynamical Model”. In: Genomics and Computational Biology (2017). doi: 10.18547/gcb.2018.vol4.iss1.e100050.

A. Naldi et al. “Logical Modeling and Analysis of Cellular Regulatory Networks With GINsim 3.0” . In: Frontiers in Physiology (2018). doi.org/10.3389/fphys.2018.00646

T. Helikar et al. “The Cell Collective: Toward an open and collaborative approach to systems biology”. In: BMC Systems Biology (2012). doi: 10.1186/1752-0509-6-96.

## Atlas of Cancer Signaling Network: a resource of multi-scale biological maps to study disease mechanisms

*Luis Cristóbal Monraz Gómez, Maria Kondratova, Andrei Zinovyev, Emmanuel Barillot, Inna Kuperstein*

*Institut Curie, Paris, France*

**Author keywords:** Biocuration, Cancer, Data visualization, Enrichment analysis, Knowledge formalization, Modelling, Molecular network, Network navigation, Signaling network map, Signatures, Structural analysis, Systems biology, Systems medicine

We present here the second edition of Atlas of Cancer Signaling Network (ACSN2.0, https://acsn.curie.fr). ACSN is a web-based resource of multi-scale biological maps depicting molecular processes in cancer cell and tumor microenvironment. The core of the Atlas is a set of interconnected cancer-related signaling and metabolic network maps. Molecular mechanisms are depicted on the maps at the level of biochemical interactions, forming a large seamless network of above 8000 reactions covering close to 3000 proteins and 800 genes and based on more than 4500 scientific publications. Constructing and updating ACSN involves careful manual curation of molecular biology literature and the participation of experts in the corresponding fields.

The maps of ACSN2.0 are interconnected, the regulatory loops within cancer cell and between cancer cell and tumor microenvironment are systematically depicted. The cross-talk between signaling mechanisms and metabolic processes in the cancer cells is explicitly depicted thanks to new feature of the Atlas: ACSN2.0 is now connected to RECON metabolic network, the largest graphical representation of human metabolism.

The Atlas is a "geographic-like" interactive "world map" of molecular interactions leading the hallmarks of cancer as described by Hanahan and Weinberg. The Atlas is created with the use of systems biology standards and amenable for computational analysis. As of today, ACSN2.0 is composed of 13 comprehensive maps of molecular interactions. There are six maps covering signaling processes involved in cancer cell and four maps describing tumor microenvironment. In addition, there are 3 cell type-specific maps describing signaling within different cells types frequently surrounding and interacting with cancer cells. This feature of ACSN2.0 reflects the complexity of tumor microenvironment. The resource includes tools for map navigation, visualization and analysis of molecular data in the context of signaling network maps.

## Collaborative construction and visual analysis of biological pathways with Newt 2.0

*Ugur Dogrusoz*

*Bilkent University, Turkey*

**Author keywords:** Systems biology, Biological pathways, Disease maps, Pathway editor, Graph visualization, Software tool, Web based software

Information about pathways is becoming increasingly available in detailed, computable standard formats such as SBGN and BioPAX. Effective visualization of this kind of information is a key recurring necessity for biological data analysis, especially for -omic data. Data analysis tools are rapidly migrating to web based platforms; hence, there is a substantial need for sophisticated web based pathway viewers that support these platforms and other use cases.

Towards this goal, we have been developing a web based editing tool named Newt for biological patways in SBGN as well as a simplified notation. Newt features nesting pathways to arbitrary depth to represent molecular complexes and cellular locations with ability to collapse, automatic pathway layout, editing and highlighting facilities to enable focus on part of a larger map, and the ability to inspect pathway members for detailed information. It can be used within a web browser without any installation and can be readily embedded into web pages.

Newt fills an important gap by making the large and fast-growing corpus of rich pathway data accessible to web based platforms. It can be used in a variety of contexts and in multiple scenarios ranging from visualization of the results of a single study in a web page to building data analysis platforms. Using a local graph database and validation mechanisms, Newt was recently turned into a collaborative construction environment, where remote users can collaborate on building and analyzing pathways.

## Reactome Pathway Analysis and Visualization

*Henning Hermjakob*

*EMBL-EBI, United Kingdom*

**Author keywords:** Pathways, Visualisation, Data analysis

Reactome (https://reactome.org) is a free, open-source, open-data, curated and peer-reviewed knowledge base of biomolecular pathways, currently covering 10,833 protein coding genes supported by 30,027 literature references. Pathways are arranged in a hierarchical structure, allowing the user to navigate from high level concepts like immune system to detailed pathway diagrams showing biomolecular events like membrane transport or phosphorylation. For the higher levels of the hierarchy, Reactome now provides scalable, interactive textbook-style diagrams in SVG format, which are also freely downloadable and editable. Repeated diagram elements like 'mitochondrion' or 'receptor' are freely available as a library of graphic elements at https://reactome.org/icon-lib. Detailed lower-level diagrams are downloadable in editable PPTX format as sets of interconnected objects, as well as in standard png format. Pathway analysis capabilities have been extended to include quantitative GSEA analysis, an R interface, and a new, visually attractive genome-wide results overview based on Voronoi maps.

## The DisGeNET platform's 10 anniversary: developing a knowledge base for disease genomics

*Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch, Francesco Ronzano, Emilio Centeno5, Ferran Sanz and Laura I. Furlong*

*GRIB (IMIM-UPF), Spain*

**Author keywords:** disease genomics, human disease, linked open data, genetic variation, disease-gene associations

The DisGeNET platform contains information about the genetic determinants of human diseases and traits. Initially developed as a Cytoscape plugin, it has evolved into different formats and applications, and it now undergoes its 6th release. The current version of DisGeNET contains more than 600,000 gene-disease associations, between 17,000 genes and 24,000 diseases and phenotypes. The platform also includes 210,000 variant-disease associations, between 117,000 variants and 10,000 diseases and traits. The data has been obtained by integrating information from a dozen repositories with associations extracted from Medline abstracts using state-of-the-art text mining technologies, with a tool developed for this purpose. DisGeNET offers a suite of bioinformatics tools to facilitate the exploration and analysis of data: a web interface, a Cytoscape App, a new API, an R package and Linked Data resources. The DisGeNET-RDF SPARQL endpoint is one of the Elixir Recommended Interoperability Resources. DisGeNET is an established resource, with approximately 20,000 users per year, used to address a variety of research questions in disease genomics, disease comorbidity, drug development and toxicity studies, and thus, it constitutes a powerful tool to facilitate translational research, and boost precision medicine.

## The druggable landscape of Familial melanoma: Using machine learning and mechanistic models to screen for putative functional drug-targets

*Marina Esteban-Medina, María Peña-Chilet, Carlos Loucera and Joaquin Dopazo*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain*

**Author keywords:** Genomics, Machine learning, Mechanistic models, Familial Melanoma, Signaling pathways, Big data, Drug-targets

**Introduction:** despite the rapid advancements in the understanding of the disease, Familial melanoma (FM) is still a major clinical challenge in the Western world. There is an urgent need for identification of effective combinations of drugs which induce more efficient responses and prevent disease progression. Merging the knowledge about mechanisms of disease and drug action, together with the arising machine learning methodologies, here we present a mechanistic model approach to identify and define the disease influence map, gaining insight into the landscape of functionally relevant genes of the disease, including those currently outside of the disease map of action.

**Material and Methods:** we selected ORPHANET/OMIM (ORPHANET:618) well-known genes responsible for FM, then using KEGG DB, we constructed the disease influence map by extracting all the circuits (receptor-effector) containing those genes. Since we were interested in finding drug combinations with potential for clinical translation, we selected well-characterized drugs approved by the Food and Drug Administration (FDA) or in late clinical trials. We implemented Multi-Output Random Forest regression, from GTEx expression data to circuits activity, to infer the effect of the selected drug-targets over the constructed disease influence map. The inference is backed by a repeated cross-validation strategy which constructs the relevance distribution for each target, while the MORF hyperparameters are selected by means of Tree-structured Parzen Estimator (TPE).

**Results:** based on the 12 genes implicated in FM, the analysis of pathways involved in Familial melanoma allowed the identification of 48 circuits in 7 pathways that triggered the disease pathogenesis and constitute the disease influence map. Moreover, the application of multi-output regression methodologies revealed relevant potential therapeutic targets which could further be studied.

**Conclusions:** we have proposed a machine learning methodology for the prediction of potentially causal relationships between drug targets and cell activities related to disease phenotypes, using the constructed influence map of the disease as functional frame. These findings would lead to a more efficient drug selection for patients treatment.

## Connecting metabolic biomarkers with biological pathways and clinical data to enable omics data interpretation

*Denise N. Slenter, Irene Hemel, Josien Landman, Chris T. Evelo and Egon L. Willighagen*

*Maastricht University, Netherlands*

**Author keywords:** Metabolic Biomarkers, Biological Pathways, WikiPathways, Cytoscape, Pathway and Network Analysis, Genetically Inheritable Metabolic Disorders, (Metabol)Omics, Data Integration

Metabolomics has rebooted our scientific attention towards small molecular compounds in living organisms [1], measuring endogenously formed metabolites and exogenous parent compounds with one integrated approach. Metabolite levels often correlate with disease phenotypes [2] and are thus very useful as biomarkers. Changes in metabolites and their concentrations are relevant for not just metabolic diseases, but also various forms of cancer, and the symbiotic interaction of the gut microbiome with the (human) body. Linking the rich metabolomics data to the context of the involved biological processes, allows mechanistic explanation with pathway and network analysis. However, this is complicated by the chemical diversity of mammalian metabolism, with over 10 thousand metabolites currently known. First, it is estimated that only 30% of the measured and characterised metabolites that can be identified. Second, some metabolomics platforms measure compound classes (e.g. a group of lipids or carbohydrates) instead. Moreover, all have structure and charge state-specific database identifiers.

With this study we aim to improve the link between disease and biological pathway knowledge. We address several of the above raised issues with a two-step approach. First, we connect metabolic biomarker information from the IEM-base [3] to the pathways in WikiPathways [4] using the RDF format [5] and SPARQL-queries. The second step extends the first with metabolic reference values and patient data, which is visualised and analysed with Cytoscape [6] and automated with the statistical software R. We have tested our approach on several genetically inheritable metabolic disorders related to the urea cycle and pyrimidine metabolism, where several overlapping biomarkers are present. With our approach we can visualize the alteration of a metabolic pathway in a patient compared to clinically relevant parameters and the downstream effects of a genetic inborn error of metabolism. Because of the diverse content of our pathways, the presented approach opens up the possibility of integration with other omics data, enabling a straightforward interpretation of disease causes and potentially finding targets relevant for treatment.

References

[1]. Kell D. et al. Metabolomics, 2016, 12:148.

[2]. Orešič M. et al. Expert Rev Mol Diagn, 2006: 575-585.

[3]. Lee J. et al. Genetics in Medicine, 2018, 151.

[4]. Slenter D. et al. Nucl. Acids Res. 2018, D661–D667.

[5]. Waagmeester A. et al. PLoS comp. bio. 2016, e1004989.

[6] Shannon P. et al. Gen. Res. 2003, 2498-504.

## Interactive visualization of phenotypic and genotypic information in disease maps for identification of mechanistic biomarkers

*Sascha Herzinger, Piotr Gawron, David Hoksza, Wei Gu, Venkata Satagopam, Reinhard Schneider and Marek Ostaszewski*

*University of Luxembourg, Luxembourg Centre for Systems Biomedicine, Luxembourg*

**Author keywords:** disease maps, visualization of molecular mechanisms, visual analytics, interoperability, tranSMART

Research projects on human disease introduce increasingly complex profiling of patients, both on the phenotypic and genotypic level. Phenotypic data encompass medical and family history, lifestyle questionnaires or standardized clinical tests. Genotypic data are collected using sequencing results of growing precision, uncovering disease-associated mutations. Both these data sources give an invaluable insight into the nature of a given pathology, but are challenging to integrate, and made understandable and accessible to clinical researchers.

We will present VarSmart, a translational medicine workflow based on interactive interfaces of tranSMART [PMID:24303286] and MINERVA [PMID:28725475], supported by a variant store using ElasticSearch technology [www.elasticsearch.org]. Using SmartR [PMID:28334291] as an interfacing technology, VarSmart retrieves the individual identifiers from cohorts selected in i2b2 tranSMART. Then it queries ElasticSearch API for variants corresponding to these individuals and visualizes them on a disease map in an associated MINERVA Platform. The entire setup allows for GUI-based data selection and exploration, without direct access to the underlying sources, masking the data complexity from clinical researchers.

**The pipeline will be demonstrated using two publicly available resources. The clinical results shown are in whole based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga, in particular the BRCA dataset, curated as a part of the eTRiKS project. The other resource is the Atlas of Cancer Signaling Networks, the largest diagram of cancer-related mechanisms. We will discuss the benefits and challenges of such an interactive pipeline from the point of view of streamlining analysis of complex clinical and genomic data, but also concerning the calculation of statistical power of the generated hypotheses.**

## Generation of process diagrams for computational modelling: lessons from the classroom

*Tom C. Freeman*

*The Roslin Institute and Royal (Dick) School of Veterinary Studies, Easter Bush Campus, The University of Edinburgh, Edinburgh, UK.*

Pathway modelling is challenging on many fronts. Firstly, biological systems are intrinsically complex, comprising of a myriad of molecular species (proteins, genes, biochemicals) forming an intricate network of interactions. Whilst there are many publications that describe elements of a given system, language is a limiting medium with which to describe such complexity, even when the components and the interactions between them are described concisely, which is rarely the case. Furthermore, all graphical languages have their limitations in displaying the nature of biomolecular species and the interactions between them, a curator must learn to deal with these. Finally, when reading the literature or examining existing models, a curator must know what information is relevant and learn how to deal with missing, ambiguous or sometimes contradictory information, displaying their accumulated knowledge as an annotated, 'readable' diagram. How then do find people to take on the task of creating models and train them to be good pathway curators?

For the last 10 years in Edinburgh we have been running an elective course for final year biomedical BSc students called 'Making sense of disease pathways'. Near to the start of the course each student is given an area of biology to model or are asked to select an area of their own choice, and during the 10 weeks of the course's duration they must produce a high-quality process diagram of their system of interest. I will describe my experience of running this course and of training students who have no previous of pathway modelling to become expert pathway curators. I will discuss some the issues that have been encountered and show some the diversity of the 50 or so models produced to date. I will also highlight the positive learning experiences for the student and describe how the course is assessed. Under-graduate and post-graduate students are a largely untapped resource for model curation and the take home message of this talk is to highlight their potential use in creating the models we need.

# Poster communications

## 1. The Adult Neurogenesis Map

*Rupert Overall*

*German Center for Neurodegenerative Diseases (DZNE), Germany*

**Author keywords:** Adult neurogenesis, Brain, Ageing, Phenotypes

The hippocampus is a key brain structure for learning and memory. It not only processes input from the environment, but also fundamentally influences behavior. This means that the neural network in the hippocampus is a core part of an information loop connecting environmental stimulus and response. It is particularly intriguing that this special brain region is also home to a population of neural stem cells which allow the environmentally-regulated creation of new neurons, throughout the life of the organism, that add an extra level of flexibility to hippocampal performance. We have previously shown that the regulation of the stem cell pool and the generation of new neurons are under complex genetic control. We also maintain a structured database of all genes reported to affect adult hippocampal neurogenesis in some way. We are now extending this effort to build a map encompassing behavioral phenotypes and environmental stimuli. The map, as a web application, enables interactive browsing and complex searching of the knowledgebase, and also provides a platform for Boolean modelling. A working draft of the Adult Neurogenesis Map is presented and challenges for future development of this project are discussed.

## 2. Maintenance and Enrichment of Disease Maps in Biological Expression Language

*Charles Hoyt, Daniel Domingo-Fernández, Sandra Spalek, Esther Wollert, Rana Aldisi, Lingling Xu, Yojana Gadiya, Trusha Adeshara, Stephan Gebel and Martin Hofmann-Apitius*

*Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany*

**Author keywords:** Knowledge Graphs, Curation, Enrichment, Quality Assurance

While numerous projects have dedicated extensive efforts into generating disease maps, most struggle with maintenance and updating. Moreover, projects' focuses and members' prior knowledge can lead into biases or gaps in coverage of relevant knowledge.

We propose an enrichment workflow [1] designed to facilitate the maintenance and updating of disease maps while simultaneously reducing bias by filling these gaps (https://github.com/bel-enrichment). While we present the application of this workflow to several disease maps in the Biological Expression Language (BEL), which encodes qualitative causal, correlative, associative, and ontological relationships between biological entities across multiple modes and scales that includes provenance, experimental context, and biological context, it is general to any systems biology modeling language.

We previously generated NeuroMMSig [2], an inventory of 124 candidate pathophysiological mechanisms, represented as graphs, for Alzheimer's disease, 65 in Parkinson's disease, and 31 in epilepsy. After a four week campaign with four curators, we were able to increase the number of nodes by 5x and edges by 7x while still maintaining focus and context specificity. As we have made several publications [3, 4] based on NeuroMMSig before its update, we are currently working to reproduce previous results and make comparisons.

We modified the enrichment workflow to take topics as input (e.g., genes, chemicals, biological processes) and prioritize literature from a database of automatically extracted relations. We applied both these workflows in two projects. In the first project, we modeled the role of the modification and aggregation of the Tau protein neurodegenerative diseases and related tauopathies. In the second, we modeled the role of proteostasis in healthy aging and neurodegenerative disease.

Finally, we not only provide the resources curated during the course of these two projects at https://github.com/pharmacome/knowledge, but we also present several software packages we have developed to facilitate curators in organizing and ensuring the quality of their work. In order to automate quality control, we generated a curation workflow using GitHub and Travis-CI (https://github.com/pybel/pybel-git). In order to facilitate the organization of BEL documents within git repositories and direct access through Python packages, we developed the bel-repository Python package (https://github.com/pybel/bel-repository).

References:

[1]. Hoyt, C. T., et al. (2019). Re-curation and rational enrichment of knowledge graphs in Biological Expression Language. Database, 2019, 536409.

https://doi.org/10.1093/database/baz068

[2]. Domingo-Fernández, D., et al. (2017). Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment. Bioinformatics, 33(22), 3679–3681. https://doi.org/10.1093/bioinformatics/btx399

[3]. Khanna, S., et al (2018). Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms. Scientific Reports, 8(1), 11173. https://doi.org/10.1038/s41598-018-29433-3

[4]. Hoyt, C. T., et al. (2018). A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities. Database, 2018. https://doi.org/10.1093/database/bay050

## 3. A mechanistic approach for lncRNA functional annotation

*Inmaculada Álamo-Álvarez, María Peña-Chilet, Carlos Loucera and Joaquin Dopazo*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain*

**Author keywords:** lncRNA, mechanistic model, functional annotation

The non-coding genome has traditionally been considered junk DNA and thought to be transcriptionally silent or functionally irrelevant; however, recent research has detected plenty of functionally significant elements in the non-coding genome, including long non-coding RNAs (lncRNAs). The Encyclopedia of DNA Elements (ENCODE) study reported the existence of over 9640 lncRNA loci in the human genome. Some of these transcripts are linked with human diseases of complex etiology, including cancer or diabetes. Although we have insight into small RNA mechanistic pathways, lncRNA are still poorly characterized due to their heterogeneity. Mechanistic models that link gene expression to cell functional activity are gaining popularity and have shown to be useful in modeling complex disease processes. Here we propose an approach based on mechanistic signaling pathways that allow the assignation of functional roles to these transcripts.

We used the lncRNA expression data normalized by FPKM (fragments per kilobase of exon model per million reads mapped) that can be downloaded from mitranscriptome.org database. We selected 5487 samples with expression data for 12382 transcripts labeled as lncRNA, which had been processed from TCGA data. We also used FPKM-normalized RNA-Seq gene expression data for the same samples, directly downloaded from GDC. Normalized gene expression values were used to compute signaling circuit activation values in the mechanistic model implemented in the HiPathia method. Pearson's R correlations between lncRNA transcript expression levels and signaling circuit activation values were computed across samples, and corrected by multiple comparisons using FDR. Signaling circuits trigger different cell functional activities which were annotated with GO terms, Uniprot keywords and, using text mining tool CHAT, with Hallmarks of Cancer.

We obtained about 7.5 million correlations between lncRNA and signaling circuits activation values with an adjusted p-value under 0.05, being the strongest positive correlation 0.98 and the strongest negative one -0.44. We also observed that negative correlations were neither as strong nor as frequent as the positive ones. In order to focus on a reduced but more reliable dataset, we restricted our initial analysis to those correlations with $|R| > 0.8$, which resulted into a more manageable dataset of 166 pairs lncRNA-signaling circuit relationships. Then, we assigned to these lncRNA a putative functional role taken from the correlated signaling circuits. This work suggests that mechanistic models could provide a useful tool to understand the functional implications of lncRNA activity and annotate them.

## 4. Drosophila melanogaster as a signaling pathway model for Duchenne Muscular Dystrophy

*Sergio Romera-Giner[1], Joaquin Dopazo[2] and Marta R. Hidalgo[1]*

*Príncipe Felipe Research Center, Valencia, Spain[1], Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[2]*

**Author keywords:** Drosophila melanogaster, signaling pathways, Hipathia, Duchenne muscular dystrophy

Drosophila melanogaster is one of the main model organisms used for the study of certain human diseases [1], mainly due to the genetic similarities between the two species and its easy maintenance and breeding under lab conditions. Nevertheless, despite the vast amount of collected information about the fruit fly, there are still many key aspects mostly unknown about its cell's inner workings that could be crucial for a better understanding and treatment of several diseases. One of them is the role of cell signaling pathways in a disease context, one of the most promising lines of work in the field of precision medicine.

Hipathia [2] is a methodology that estimates the activation of cell signaling pathways from gene expression data. Recently, Drosophila melanogaster has been added as an available organism to be analyzed with Hipathia, allowing for a new insight into fruit fly cell processes.

To test the potential of Drosophila signaling pathways as models of disease, we have applied Hipathia to study the Duchenne muscular dystrophy (DMD), a complex disease believed to be related to specific signaling pathways [3]. For this task, we have analyzed gene expression data from a Drosophila melanogaster DMD model and from human DMD patients, comparing diseased to healthy subjects of both species. The results for both species have been later compared to find common behaviours in the differential activation of the disease-related signaling pathways, stating Drosophila melanogaster as a suitable model for future DMD studies.

Bibliography:

[1] UB Pandey et al., 2011, DOI: 10.1124/pr.110.003293

[2] MR Hidalgo et al., 2017, DOI: 10.18632/oncotarget.14107

[3] S Bhatnagar, 2010, DOI: 10.1007/s00109-009-0550-4

## 5. Towards representation of genetic alterations in Glioblastoma Multiform

*Romain Tching Chi Yen*

*ITTM Solutions SA, Luxembourg*

**Author keywords:** Glioblastoma, GBM, Cancer, Genetic alterations, Mutations

Glioblastoma Multiform (GBM) is the most common and deadliest type of brain cancer, with a median life expectancy of 14 months after diagnosis despite treatment. There are around 3 new cases per 100,000 people and per year. Because of its high heterogeneity and resilience, the mechanisms through which the tumour develops and becomes resistant to treatment are still poorly understood. As part of the efforts of the Marie-Curie H2020 GLIOTRAIN project to better understand the molecular mechanisms of the disease, we are currently building a disease map for GBM. A challenging aspect of such a cancer map is the modelling of genetic mutations and alterations underlying and driving the disease. Indeed, genetic mutations and alterations driving tumour development are stochastic events selected for the survival advantages they provide, yet they are also very heterogeneous across patients in GBM (and in cancer in general). Be it SNPs, insertions/deletions, copy number alterations, promoter hypermethylation, etc. all the way to gain or loss of whole chromosomal arms, modelling them within the framework of Disease Maps is not trivial and need some consideration. In addition, it has been frequently observed that some of these mutations are mutually exclusive, where two (or more) genes have a high mutation frequency in the population yet are almost never mutated simultaneously in cells. This adds to the complexity of representing genetic alterations in a Disease Map. However, since they are the driving factor behind GBM and most cancers, including them in cancer-focusing Disease Maps is an important matter that needs to be addressed. Here we present our work on the GBM Disease Map to model genetic alterations, showing and discussing our approach to a problem relevant to all cancers and somatic mutations-driven Disease Maps.

## 6. Using multi-omics to depict the response of human fibroblasts to X-ray radiation

*Martín Garrido-Rodríguez[1], Rosario Morrugares[1], María Peña-Chilet[2], Eduardo Muñoz[1,] Marco A Calzado[1] and Joaquin Dopazo[2]*

*Maimónides Biomedical Research Institute of Córdoba (IMIBIC), Córdoba, Spain[1], Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[2]*

**Author keywords:** Multi-omics, Phosphoproteomics, DNA damage, Signaling

Radiodermatitis is a secondary effect of the skin exposure to ionizing radiation and a limiting factor in the treatment of some types of cancer by radiotherapy. Since fibroblasts play a main role in the development of the disease, we decided to investigate the response of this cell type to different doses of X-ray radiation at transcriptomic, proteomic and phosphoproteomic level. Then, this information was used to evaluate the observed changes by using mechanistic models of signaling pathways.

For this, primary human fibroblasts were cultured and exposed to two different doses of X-ray radiation. Poly-A RNA and SILAC-labeled protein samples were obtained and processed through RNA-Seq and LC-MS/MS, respectively. RNA-Seq reads were aligned to the human genome and the counts per gene were analyzed using the edgeR R package. The proteomics and phosphoproteomics changes in response to both doses of radiation were statistically evaluated through a One Sample T-Test applied to the resultant SILAC ratios. The information on gene functional interactions as described in KEGG and PhosphoSitePlus was used to construct an extended map of the pathways involved in the fibroblast response to X-rays. The signaling circuit activities were calculated using data from the different -omic levels by applying the hipathia mechanistic modelling.

As expected, the analysis of the activity of signaling circuits revealed a group of significant alterations within the pathway, which are involved in biological processes such as DNA damage, inflammation and extracellular matrix organization. Moreover, the additional signaling events added to the map allowed us to identify novel interactions not included in the canonical KEGG pathways. In this subset of novel features, we identified candidate targets that could be used to prevent the damage produced by radiation.

In conclusion, the combination of multi-omics data with an extended pathway of the response to radiation allowed us to identify novel altered signaling circuits. Those circuits may include putative druggable targets to prevent the reaction of the fibroblasts in the development of radiodermatitis.

## 7. Unraveling the heterogeneous signaling and functional landscape of single-cell RNA-seq

*Matias M. Falco[1], Carlos Loucera[1], Marta R. Hidalgo[2] and Joaquin Dopazo[1]*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[1], Príncipe Felipe Research Center, Valencia, Spain[2]*

**Author keywords:** single-cell, signaling pathways, heterogeneity

The recent development of single-cell transcriptomics provides an exceptional approach to dissect the complexity of bulk samples and shed light on the functional nature of its heterogeneity. A growing number of analyses have revealed a large degree of heterogeneity in gene expression levels across the different cells of the same samples. However, little is known on the functional consequences of this heterogeneity. The use of mechanistic models of signaling pathways has proven to be a useful tool for understanding relevant aspects of cell functionality. Here we propose to use mechanistic models of signaling pathways to deconvolute the functional landscape of a tissue using single-cell RNA-seq experiments.

Here we carried out a comprehensive study of single cell circuit activity on a glioblastoma experiment with 3589 single-cell gene expression measurements available [1].

Currently, drop-out events are inevitable in any single-cell technique, which makes necessary the initial use of imputation methods. Following, we transformed the gene expression values into signaling circuit activities using a mechanistic models of signal transduction as implemented in the Hipathia tool [2] (http://hipathia.babelomics.org/).

This analysis of signaling circuit activities revealed that the degree of heterogeneity at circuit level is higher across cells than the observed across samples from different individuals, as happened at the gene level. Also, different neoplastic clusters have been characterized according to their differences in signaling circuit activity profiles. We specifically focused on the signaling circuits that trigger cell functionalities that can be easily assimilated to cancer hallmarks.

In conclusion, we present a methodology that allows estimating the activation state of signaling circuits, defined within pathways, at the level of individual cells. With this methodology we are able to outline the profile of signalling circuit activities and, consequently, the functional landscape on any single-cell experiment.

References:

[1]. Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., & Li, G. (2017). Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. Cell reports, 21(5), 1399-1410.

[2]. Hidalgo, M. R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., & Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget, 8(3), 5160–5178. doi: 10.18632/oncotarget.14107.

## 8. CyPathia: A new Cytoscape app for interpretation of the consequences of the combined changes of gene expression levels and/ or genomic mutations in the context of signalling pathways

*Kinza Rian[1], Marta R. Hidalgo[2] and Joaquín Dopazo[1]*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[1], Príncipe Felipe Research Center, Valencia, Spain[2]*

**Author keywords:** Hipathia, pathways, Cytoscape, precision medicine

Understanding the dynamics of the cell functionality and its contribution to disease or drug action mechanisms is a main challenge for precision medicine. Mechanistic models are gaining importance for genomic data interpretation because they provide a natural link between genotype-level measurements and phenotype-scale functional behavior of the cell.

In this paper we proposed a novel Cytoscape plugin (called CyPathia) that allows the Cytoscape's community to interpret disease and drug action mechanisms, as well as to predict the potential consequences of therapeutic interventions. This plugin implements the hipathia method [1] that models cell signaling activity from gene expression and/or mutation data using biological knowledge on signal transduction as represented in pathway repositories. Currently KEGG pathways are implemented by default but user-defined pathways can also be used. We are working to include also Disease Maps pathways.

The CyPathia plugin is based on Hipathia bioconductor package [2]. CyPathia provides the Cytoscape community for the first time with the possibility of using mechanistic models. It also improves workflow reproducibility and researcher productivity by enabling popular languages (e.g JavaScript and R) to directly perform genome-scale pathway interpretation and analysis.

References:

Hidalgo, M. R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., & Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget, 8(3), 5160–5178. doi: 10.18632/oncotarget.14107

https://www.bioconductor.org/packages/release/bioc/html/hipathia.html

## 9. A parser of signaling pathways for KEGG, Wikipathway and manual-curated AF pathways to Simple interaction file format

*Kinza Rian[1], Fatima Zahraa Annassiri[2], Amal Maurady[2], María Peña-Chilet[1] and Joaquin Dopazo[1]*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[1], Laboratory of Innovation Technology LTI, ENSAT, Tangier, Morocco[2]*

**Author keywords:** signaling pathways, network parser, genes interaction

Currently, using the knowledge available in the different pathway repositories (KEGG, Reactome, Wikipathways, Disease Maps, etc) for studying the aspects of the cell functionality remains a challenge due to the different formats in which those repositories allow downloading the pathway descriptions. Moreover, this problem of format incompatibility extends to pathway edition tools. On the other nans, many network/pathway analysis software require pairwise interaction networks as input SIF (Simple Interaction Format file).

In order to face the challenge of the lack of a common standard, we present here a parser from KGML (KEGG signaling pathways)[1] , XML ( from wikipathway database) [2] and SBGN ( manual curated pathways by Newt)[3] formats to SIF files. Our parser contains a set of rules to reduce complex interactions to pairwise (or binary) relationships.

Whereas the SIF does not include the style information such as layout, node shape and colors, etc., which are mandatory for some of network/pathway analysis tools, such as Hipathia [4], our parser save this information in a simple tab file, which is named attribute file (.att format).

The parser presented here can be used as a preprocessing tool for signaling pathways downloaded from different sources to make them usable within applications that require pairwise interaction input format. Actually, SIF format can be easily imported into popular network analysis tools, such as Cytoscape, Hipathia, etc.

References:

[1]. Kanehisa M. The KEGG database. Novartis Foundation Symposia. 2002;247:91–103. doi: 10.1002/0470857897.ch8.

[2]. Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., ... Willighagen, E. L. (2018). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic acids research, 46(D1), D661–D667. doi:10.1093/nar/gkx1064

[3]. http://newteditor.org/software.html

[4]. https://www.bioconductor.org/packages/release/bioc/html/hipathia.html

## 10. Variant prioritization using a mechanistic approach in Mental and behavioural disorders

*Ana M Pérez-Gutiérrez, Rosario Carmona, Virginia Aquino-Quintans, Javier Pérez-Florido, María Peña-Chilet and Joaquín Dopazo*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain*

**Author keywords:** mechanistic model, functional annotation, variant prioritization, mental and behavioural disorders

One of the main problems with Rare Diseases (RD) is the still large number of undiagnosed cases, especially in mental and behavioral disorders. Although approximately 80% of RDs have a genetic cause, the wide range of phenotypic variability that these pathologies present, besides the lack of knowledge of the responsible mechanisms of diseases, makes it difficult to find new disease genes and to determine new molecular therapeutic targets. Indeed, prioritization strategies used so far, targeting one gene at a time, failed to detect the causative genes in many cases. Conversely, a more systems biology oriented approach could render better results, especially in pathologies with putatively complex disease mechanism. Therefore, the detection of systematically affected pathways using a mechanistic mathematical model could provide a new approach for candidate gene prioritization that allow, not only the identification of new disease genes, but also the biological processes involved in the disease shedding light on the disease mechanism.

We analysed genomic data from a cohort of 130 individuals that present some type of mental disorder (ICD-10:F01-F99). We prioritized variants according to several criteria and assigned these variants a level of evidence. In order to evaluate the effect of those variants in the human brain tissue we performed a Mechanistic Pathway Activity (MPA) analysis using Hipathia algorithm. We evaluated the potential effect that mutations harbored by each individual would have over the signaling circuits in brain healthy tissue using GTEx transcriptome data as reference. For this purpose, we performed in silico knockout and calculated the circuit activities obtaining those pathways deregulated in the simulated affected tissue. We carried out a stratified study by phenotype, attending to their HPO term, in order to detect the specific pathways altered in each phenotype subgroup.

In this work we presented a new methodology to prioritize variants that uses a mechanistic mathematical model of signaling pathway activity, as a proxy for cell functional activity, in contrast to other methods that only consider genes individually. Using this approach, we obtained the pathways systematically deregulated in mental and behavioral disorders as well as the specific signaling circuits associated to the corresponding HPO terms. Moreover, this study revealed new candidate genes that can be further validated, helping us gain insights in the complex disease mechanisms involved in mental disorders.

## 11. A variational autoencoder approach to learn the signaling pathways manifold

*Angel Serrano[1], Isabel A. Nepomuceno-Chamorro[1], Carlos Loucera[2] and Joaquín Dopazo[2]*

*University of Seville[1], Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[2]*

**Author keywords:** Genomics, Deep Learning, Mechanistic models, Signaling pathways, Manifold Learning

**Introduction:** It is known that gene expression measurements are able to obtain relevant information about the state of each tumor. Deep neural networks have shown promising results in many fields where pattern recognition plays and important role, such as image classification, language translation, or speech recognition. Furthermore, deep learning has been used in many areas related to biomedicine, from the analysis of genomic data to splicing prediction. In fact, some models are capable of learning a meaningful latent space from gene expression. We present a Variational Autoencoder (VAE) which was fed with data generated by signaling circuits activity, in order to study whether this model can capture biologically-relevant features and preserve the structure of different tissues and tumor types/subtypes.

**Material and Methods:** We implemented a VAE in order to learn a compressed yet biologically relevant representation of pathway circuits activity. The signaling activity was computed on TCGA expression data by making use of the Hipathia R package which implements a mechanistic model built on top of the KEGG pathway database. In order to gain new insights into the circuit space latent structure, we used t-distributed stochastic neighbour embedding (t-SNE) as a means to visualize the encoding layer which had 100 nodes. Then, we analyzed the reconstruction error distribution along several types and subtypes of tumors.

**Results:** Based on the tumors represented in TCGA and the analysis of the VAE encoder layer, we can assert that our autoencoder is able to preserve tissue specific patterns. Furthermore, as the error analysis showcases, the model presents good generalization capabilities.

**Conclusions:** As our model is able to learn a signaling manifold representation we are in a good position to combine our VAE with gene and metabolism-specific autoencoders. This could lead to a multi-omic approach for biomarker and drug discovery findings by embedding/encoding complete cell-lines (CL) and searching for tumor subtype-CL matches with a fully automatic methodology.

## 12. Determination of lymph eigenfrequency through simulation as a possible lymphedema non-invasive treatment

*Sebastian Diaz, Miguel Esteban Mora Gonzalez and Ricardo Jaramillo Diaz*

*Universidad Ecci, Colombia*

**Author keywords:** Lymphedema, Treatment, eigenfrequency, simulation

The lymphatic system disease that present swelling by accumulation lymph, generates health problems and psychosocial complications in affected people. The causes can be; complications in cancer treatment, mainly breast cancer surgery and radiotherapy, among other factors; filariasis in equatorial zone; and hereditary conditions. It is estimated that around 140 to 250 million of people in the world suffer this pathology. Lymphedema can be primary or secondary. Primary lymphedema affects 1/100000 persons, mainly children and adolescents. Secondary, affects 1/1000 people, adults mostly.

Current treatments generate complications which can be surgical techniques and non-surgical. Among the non-surgical techniques are complete decongestive therapy, this can lead injuries blood vessels or muscle mass and drugs as diuretics or antimicrobials that depend the patient's response. Surgical techniques include lymphatic lymphatic anastomosis (LLA), lymphatic venous lymphatic anastomosis (LVLA), vascularized lymph node transfer (VLNT). These surgeries present permeability problems. This paper proposes a non-invasive palliative treatment model based on resonance frequency of lymph and presents a simulation of radiation exposition of a computational model lymphedema.

**Methodology:** Literature was sought about the lymphatic fluid, etiology, physiology, chemical and physical characteristics, diagnosis and treatments. As of different articles about model or simulations of lymphedema and lymphatic system was performed a simulation with the shape of an ellipsoid in the software COMSOL MULTIPHYSICS that investigated lymph eigenfrequency. Was used the Acoustic-Solid Interaction, Frequency Domain multiphysics interface that combines the Pressure Acoustics, Frequency Domain and Solid Mechanics interfaces to connect the acoustic pressure variations in the fluid domain to the structural deformation in the solid domain.

**Preliminary results:** The literature sought about the lymphatic fluid; exhibit want of studies based on different radiation types non-invasive on lymphedema. Were found values necessary for physical simulation as; density, Young's modulus and Poisson ratio. The results of simulation shows that lymph eigenfrequency based on acoustic interaction be among 496,31-503.5 Hz. This generated conditions in the medium such as: sound pressure level, acoustic pressure also it evince a deformation of de structure simulated as lymphedema.

**Preliminary conclusions:** The results suggest damage or unfolding in the edema, the irradiation in the lymph eigenfrequency can be considered as alternative treatment non-invasive for lymphedema which decrease complications presents in the current treatments. In this study the form of lymphedema was assumed as ellipsoid, in studies futures will be take the form in the specific limb with the other tissues present, the response to different types of radiation such as micro-waves and ultrasound will also be evaluated.

## 13. Enabling disease maps analysis with Minerva API and plugins

*David Hoksza, Piotr Gawron, Marek Ostaszewski, Ewa Smula and Reinhard Schneider*

*University of Luxembourg, Luxembourg Centre for Systems Biomedicine, Luxembourg*

**Author keywords:** systems biology, disease maps, visualization, API

Disease maps offer a rich source of contextualized knowledge about disease molecular mechanisms which can be harvested to gain new insights. To do so one needs tools to visually explore the molecular data, to contrast it with the knowledge available in numerous bioinformatics databases and with data generated from high-throughput experiments. Such functionality is offered by MINERVA (http://r3lab.uni.lu/web/minerva-website/), a platform for web-based visual exploration of molecular networks.

However, the growth of the Disease Maps Community creates the demand for the integration of new data sources accompanied by novel analytical and visualization functionalities. To address these needs, the MINERVA platform was extended by two ways: (i) an Application Programming Interface (API) for programmatic access and manipulation of hosted data, and (ii) a plugin architecture allowing to flexibly extend the existing functionality of the web interface.

The REST API of MINERVA automates functionalities such as i) obtaining elements and reactions, including all available annotations, of hosted maps, ii) listing drugs, chemicals and miRNA targeting map elements and iii) uploading overlays such as gene expression or variation information, to a given map. With the API one can turn the originally purely web-based functionality into high-throughput computational pipelines.

The plugin architecture is supported by an extension of the API which enables custom JavaScript to interact with the respective MINERVA instance to retrieve its data, listen to its events and modify its visual state. This allows for the construction of custom plugins for advanced visualization and data integration, independent of the core functionality of MINERVA. Additionally, the plugins can use the REST API to access all functionality which is available also outside of the JavaScript API. The advantage of the plugin architecture is twofold. First, whenever a custom visual interpretation is required, users can benefit from the existing web interface and easily extend its capabilities via the plugins architecture. Second, plugins do not need to be distributed with MINERVA core application and thus provide a quick way to address user requirements without the need to respect the full development cycle.

## 14. Different layout aware formats in systems biology used by Minerva platform

*Piotr Gawron, David Hoksza and Marek Ostaszewski*

*University of Luxembourg, Luxembourg Centre for Systems Biomedicine, Luxembourg*

**Author keywords:** layout, molecular network, systems biology

The understanding of complex biological networks often relies on both a dedicated layout and a topology. Currently, there are three major competing layout-aware systems biology formats, but there are no software tools or software libraries supporting all of them. This complicates management of molecular network layouts and hinders their reuse and extension. We present a high-level overview of the layout formats in systems biology, focusing on their commonalities and differences, and introduce a new conversion module within the MINERVA platform. The module is available via a REST API and offers, besides the ability to convert between the layout-aware systems biology formats, the possibility to export the layouts into several graphical formats.

## 15. Conciliation of medicine systems disease maps and other molecular interaction networks using logical properties of ontology

*Vincent Henry[1], Giulia Bassignana[2], Violetta Zujovic[2], Fabrizio De Vico Fallani[3], Olivier Dameron[4], Ivan Moszer[5] and Olivier Colliot[3]*

*OMICtools project[1], Institut du Cerveau et de la Moelle Epinière[2], Inria – ICM[3], University of Rennes[4], IHU-A-ICM[5], France*

**Author keywords:** Systems biology, Ontology, Logical reasoning, Network

Medicine systems disease maps (DM) summarize knowledge about molecular pathophysiology into process description networks (PDN). PDN provide a network representation wherein molecules states are described and linked to processes they are involved in, and processes are linked together by the molecules they share. Direct interaction networks (DIN) are networks based on direct interactions between unprocessed molecules. The choice of the network type is a key element in systems biology analysis. PDN may be too complex for some analysis, but also the only knowledge resource available. DM provide a fine representation of molecular reactions, including information for metabolic reactions, transduction signaling and gene expression regulation. Thus, we hypothesize that DIN could be deduced from information provided by DM. More specifically, ontologies, which are able to manage knowledge and relationships using logical rules, are a suitable framework to deal with this large-scale inference. We present an ontology-driven methodology that results in the deduction and the addition of DIN relationships to PDN ones.

**Methods:** We designed the Molecular Network Ontology (MNO), which contains 42 classes that represent a) molecular reactions (i.e. binding, conversion or transcription) and b) molecular participants (e.g. gene, native gene product or converted gene product). Process classes were formally defined according to participant classes using 4 PDN relationships: “has input”, “has output”, “positively mediated by” or “negatively mediated by”. Then, based on these relationships and process classes, logical rules were designed to infer DIN relationships (e.g. “symmetrically interact with” or “positively directed to”).

**Use Case:** The macrophage signal transduction map (MSTM) is a curated DM that contains 724 molecular reactions involving 1,353 participants.

MSTM network was integrated into MNO: reactions and participants described in MSTN became individual instances of MNO classes. Edges from MSTN became PDN relationships between instances. Meta-information (labels, identifiers, crosslinks...) were kept as individual annotations. Following automatic reasoning, the patterns of MNO-inferred DIN relationships and of those provided by STRING queries were compared to validate the consistency of deductions.

**Results:** MNO could fully integrate MSTM information as individual instances of its classes. Then logical rules allowed us to enrich the initial PDN relationships with 36,442 consistent DIN relationships.

***Conclusion:*** Ontologies are an adequate framework to manipulate networks. Subsequent logical reasoning benefits from the genericity of a systems approach, leading to the inference of consistent DIN new relationships from PDN. Finally, such a methodology opens perspective to expand the choice of appropriate networks for systems biology analysis.

## 16. A tutorial of HiPathia, a mechanistic model of signaling pathways

*Kinza Rian[1], Marta R. Hidalgo[2] and Joaquin Dopazo[1]*

*Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[1]. Príncipe Felipe Research Center, Valencia, Spain[2]*

**Author keywords:** mechanistic model, signaling pathways, HiPathia

Conventional gene-based approaches ignore the modular nature of most human traits, while pathway enrichment methods produce only illustrative results of limited practical utility. Recently, new methods have emerged that change the focus from the whole pathways to the definition of elementary sub-pathways or circuits within them that have any mechanistic significance. In some cases, this involves the recodification of Process Description pathways into Influence Maps that describe how proteins interact among them to trigger or carry out cell functionalities. The activity of such circuits defined within Influence Map pathways is expected to be a better descriptor of cell functional activity than measurements for whole pathway or single gene activities.

Here we present a tutorial on the HiPathia MPA method (1), implemented in a R/Bioconductor package (http://bioconductor.org/packages/devel/bioc/html/hipathia.html) as well as in an interactive web application (http://hipathia.babelomics.org/). This tutorial demonstrates how to transform decontextualized gene expression measurements into highly-informative cell activity quantitative values and how to relate them to phenotypes. Different analyses can be carried out using circuit activities that include differential activity analysis, when two conditions are compared, or relation of circuit activities to a continuous variable. Since circuits modeled have an associated functional meaning (any of them trigger one or more cell functions, defined by Gene Ontology terms), the results provide direct clues to understand disease mechanisms or drug modes of action. It is also possible to build predictors directly based on circuit activities, which adds an interesting mechanistic dimension to the prediction process.

In addition to be used to uncover the molecular basis of phenotypes, mechanistic models can also be used to predict what would be the potential effect of one or several interventions (KOs, inhibitions, over-expressions, drugs, etc.) over the system studied. Thus, the PathAct (2) wen application (http://pathact.babelomics.org/) allows predicting from a holistic perspective what would be the effects of interventions over a specific system. The HiPathia suite provides a friendly environment to use Influence Map pathways as templates of cell functionality to provide a mechanistic interpretation of transcriptomics data.

References:

[1]. M. R. Hidalgo et al., High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget 8,5160 (Dec 22, 2017).

[2]. F. Salavert et al., Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. Nucleic Acids Res 44, W212 (Jul 08, 2016)

## 17. Causal structure learning inferring signal transduction pathways

*Helena Molina Abril[1], Isabel Nepomuceno[1], Cankut Cubuk[2], Carlos Loucera[2], Marta Hidalgo[3], Kinza Rian[2], Matías M. Falco[2], María Peña-Chilet[2] and Joaquin Dopazo[2]*

*University of Seville, Seville, Spain[1], Clinical Bioinformatics Area, Fundación Progreso y Salud, Seville, Spain[2], Príncipe Felipe Research Center, Valencia, Spain[3]*

**Author keywords:** Causal structure learning, Signal transduction Pathways, Graphs

The problem of inferring signal transduction pathways is addressed here using causal structure learning. In cases where experimental data are limited, these methods try to learn information about a causal structure from observational data. We compare here the application of several recently proposed structure learning algorithms, assuming that the initial data were generated from an unknown causal structure which can be represented by a directed acyclic graph. Constraint-based, Score-based and Hybrid methods have been tested, and compared to neural networks approaches. Results (reported first inferring only the graph structure, and secondly adding edge's directionality) show reasonable results for certain circuits but poor performance for others, illustrating the limitations of such methods and the need to further study causal assumptions violations

## 18. The Atlas of Inflammation Resolution (AIR)

*Matti Hoch and Olaf Wolkenhauer*

*Department of Systems Biology and Bioinformatics, University of Rostock*

**Author keywords:** Acute Inflammation, Inflammation Resolution, Multilevel Organisation

Inflammation is the first action in a complex healing process after pathogenic infections or injuries, initiated by our immune system. Recent advancements in omics yield a remarkable amount of data on inflammation and inflammation resolution, however, a detailed and quantitative understanding of the involved regulatory processes is still missing. Toward the development of more sufficient investigation methods, we constructed the Atlas of Inflammation Resolution (AIR), a research resource combining a molecular interaction map (MIM) of the latest information on interacting mediators in inflammatory pathways with molecular regulatory processes of clinical phenotypes.
The MIM contains clinical indication-specific molecular interactions from publicly available literature and databases. The network was generated using state-of-art network construction methods and the CytoScape platform. Furthermore, we expanded the MIM by two layers with transcriptional and miRNA regulators from established databases. To bring the AIR into clinical context, we further manually created and curated individual small MIMs (submaps) containing detailed molecule processes on selected clinical phenotypes for inflammation and inflammation resolution.
The AIR is available on the MINERVA (Molecular Interaction Network Visualization) platform providing an interface for users such as clinicians or researchers to extract information or to visualize their own data. This provides opportunities for the development and implementation of functionalities to predict key regulatory elements and changes in clinical disease phenotypes or drug interactions through analysis of context-specific -omics data. Using the AIR to elucidate mechanisms and signals of drugs or diseases that regulate the complex response of inflammation/inflammation resolution will lead to the development of a more advanced and directed therapy design.

# 4th Disease Maps Community Meeting

Sevilla, 2-4 October 2019