Highlights from the eleventh ISCB Student Council Symposium 2015

Wilkins, Katie; Hassan, Mehedi; Francescatto, Margherita; Jespersen, Jakob Berg; Gonzalo Parra, R.; Cuypers, Bart; DeBlasio, Dan; Junge, Alexander; Jigisha, Anupama; Rahman, Farzana

Published in:
B M C Bioinformatics

Link to article, DOI:
10.1186/s12859-016-0901-4

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Wilkins, K., Hassan, M., Francescatto, M., Jespersen, J. B., Gonzalo Parra, R., Cuypers, B., ... Rahman, F. (2016). Highlights from the eleventh ISCB Student Council Symposium 2015. B M C Bioinformatics, 17(Suppl. 3), 203-205. DOI: 10.1186/s12859-016-0901-4

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Highlights from the 11th ISCB Student Council Symposium 2015

Dublin, Ireland. 10 July 2015

Published: 25 February 2016

A1

Highlights from the eleventh ISCB Student Council Symposium 2015

Katie Wilkins1,2, Mehedi Hassan3, Margherita Francescatto4, Jakob Jespersen5,6, R. Gonzalo Parra7, Bart Cuypers8,9, Dan DeBlasio10, Alexander Junge11, Anupama Jigisha12, Farzana Rahman1

1Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University; Ittaca, NY, USA; 2Graduate Field of Computational Biology, Cornell University, Ittaca, NY, USA; 3School of Computing & Mathematics, University of South Wales, Cardiff, UK; 4Department of Genome Biology for Neurodegenerative Diseases, German Center for Neurodegenerative Diseases, Berlin, Germany; 5Department of Surgery, Massachusetts General Hospital, The Broad Institute of MIT and Harvard, Boston, USA; 6Department of Systems Biology, Technical University of Denmark, Kemitorvet, Denmark; 7Protein Physiology Lab, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina; 8Molecular Parasitology Unit (MPU), Institute of Tropical Medicine, Naionalestraat 153, 2000 Antwerp, Belgium; 9Advanced Database Research and Modeling (ADReM) research group, University of Antwerp, Middelheimlaan 1, 2000 Antwerp, Belgium; 10University of Arizona, Department of Computer Science, Tucson, AZ 85721, USA; 11Center for non-coding RNA in Technology and Health, Department of Veterinary Clinical and Animal Sciences, University of Copenhagen, Copenhagen, Denmark; 12University College Dublin, Dublin, Ireland

Correspondence: Farzana Rahman (farahman@iscb.org) – School of Computing & Mathematics, University of South Wales, Cardiff, UK

© 2016 The Author(s). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
described a method for dealing with heterogeneity by analyzing multiple models of heterogeneity and calculating model selection statistics to identify the most likely of these models, accounting for noisy data. This method was tested on several real data sets [11]. Although we often think of DNA as a two-dimensional sequence, the development of the high-throughput ChiA-PET method for identifying all physical contacts between distant loci is enabling modeling of DNA as a three-dimensional structure. Przemysław Szalaj presented 3D-NOME, a new computational method for modeling the three-dimensional structure of the genome [12]. Based on both ChiA-PET data and known interactions of CTCF and RNAPII, 3D-NOME builds an initial model of the nucleosome in a bottom-up fashion and then uses Monte Carlo simulations to reconstruct each level of structure in a top-down fashion.

Epigenome-wide association studies allow the high-throughput identification of epigenetic markers that contribute to human disease. Charles Edmund Breeze presented eFORGE, a tool for identifying the cell type specificity of differentially methylated positions significantly associated with disease. eFORGE identifies cell types of interest based on enrichment of overlap between significant differentially methylated positions and DNase I hypersensitive sites in each cell type, compared to matched, randomly generated background sets. Jonas Ibn-Salem used Hi-C chromatin conformation to show that paralogous genes share more enhancer elements with one another and is more likely to occur in the same topological association domain than matched, randomly selected gene pairs. This suggests that paralogs share common regulation because and cluster within the three-dimensional chromatin architecture.

As the number of genome sequences available increases, it becomes more challenging for biologists to characterize the data and use it to classify sequenced organisms. Giulia Fiscon presented a new feature extraction method for solving this problem, which she integrated into a classifier used to assign virus specimens to viral species [13]. RIG-I-like receptors (RLRs) are an important component of the immune response to viruses, detecting viral RNA and triggering the production of type I interferons. Robin van der Lee identified 10 features common to RLR pathway components and used these features to predict a role in the RLR pathway for 187 novel genes. RNAi knock-down screens showed that almost half of these genes had an impact on production of type I interferon, IFNβ [14].

Only four drugs for treating the lethal disease visceral leishmaniasis are available and the resistance is emerging rapidly in the causal parasite, Leishmania donovani. Bart Cuyvers shared a new database, integrating the genomic, epigenomic, transcriptomic, proteomic, and metabolomics, and phenotype data available for L. donovani. This database is coupled with many novel data-analysis tools, such as one used to find patterns in different omics data across different conditions [15].

In the final oral presentation of the day, Westa Domanova showed that temporal data could be used to distinguish the activity of different kinases, even those sharing the same consensus motif [16].

Award winners

Thanks to the generous contribution of the Swiss Institute of Bioinformatics, the Knowledge Integration for Development workshop was awarded to Alexander Monzon and Marek Cmero. Thanks to the generous contributions of F1000, the workshop focused on how to give an effective elevator pitch and gave attendees practical advice on how to give an effective elevator pitch and gave attendees practical advice on how to give an effective elevator pitch and gave attendees practical advice on how to give an effective elevator pitch.

Career Central workshop

The goal of Student Council Career Central is to expose students to the experiences and success stories of senior researchers. To this end, the Student Council collaborated with the Junior PI group and COBE COSI to organize an Applied Knowledge Exchange Sessions workshop dedicated to career development. The Student Council invited the first speaker of the session, Dr. Mick Watson, Director of ARK-Genomics, The Roslin Institute, University of Edinburgh, UK. He focused his advice on how to give an effective elevator pitch and gave attendees practical advice on how to give an effective elevator pitch to one another.

Conclusions

This year’s number of submissions and participants increased by about a third and a half respectively over last year’s symposium. It seems likely that this is due in part to greater availability of funding and easier acquisition of visas in Europe for many participants.

Acknowledgements

Because of space constraints we are unable to mention in this publication all the volunteers whose contributions make the Student Council Symposium a reality every year. Our recognition and appreciation goes out to all of them, since without their support the organization of such an event would simply not be possible.

We would like to thank ISCB Executive Director Diane Kovats, ISCB Conferences Director Steven Leard, ISCB Conference Administrator Pat Rodenburg, ISCB programmer Jeremy Henning, and ISCB Administrative Support Suzi Smith for their logistical support and invaluable advice. Furthermore, we thank the ISCB Board of Directors for their continued support of the ISCB Student Council in general and the Student Council Symposium in particular.

We are greatly indebted to ISMB 2015 conference chairs Dr. Alex Bateman and Dr. Janet Kelso for giving us the opportunity to organize the Student Council Symposium 2015 in Dublin. The Student Council would also like to thank our keynote speakers Dr. Ruth Nussinov and Dr. Des Higgins who generously donated their valuable time by delivering keynote addresses.

The Symposium would not be possible without the financial support of our generous sponsors. We would like to thank BioMed Central, Oxford University Press, Swiss Institute of Bioinformatics, F1000, Bina, and The Genome Analysis Center for their contributions.

We are very grateful to all the volunteer reviewers for their work on ensuring the quality of the scientific program, and to the program and travel fellowship committees for coordinating the reviewing effort.

Finally, we thank all Student Council members that have spent countless hours organizing all aspects of SCS 2015 to ensure its success.

References

1. Rahmann F, Wilkins K, Jacobsen A, Junge A, Vicedo E, DeBiasio D, Jigisha A, Di Domenico T: Highlights from the tenth ISCB Student Council Symposium 2014. BMC bioinformatics 2015, 16(Suppl 2):A1.
O1
Prioritizing a drug's targets using both gene expression and structural similarity

Griet Laenen1,2, Sander Willems3, Lieven Thoresz2, Yves Moreau1,2
1KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium; 2iMinds Medical IT Dept, Leuven, Belgium; 3Scientific Institute of Public Health (ISP), Platform of Biotechnology and Molecular Biology (PPB), Brussels, Belgium; 4KU Leuven, Dept. of Development and Regeneration @ Kulak, Kortrijk, Belgium

Correspondence: Griet Laenen (griet.laenen@esat.kuleuven.be) – iMinds Medical IT Department, Leuven, Belgium

Background
The pharmaceutical industry is facing unprecedented pressure to increase its productivity. Attrition rates in the later stages of development have risen sharply, with toxicity and lack of efficacy being the main bottlenecks [1]. To address both these safety- and efficacy-related issues, a better understanding of the complex biological response to drug treatment is vital. Although many drugs exert their therapeutic activities through the modulation of multiple targets [2], these targets are often unknown and identification among the thousands of gene products remains difficult. We propose a computational method to support the identification of putative targets of a drug by means of a dual approach combining network diffusion of gene expression with chemical structure similarity.

Methods
The first component of our method prioritizes proteins as potential targets by integrating experimental gene expression data with prior knowledge on protein interactions [3, 4]. More specifically, genes are ranked based on the transcriptional response of functionally related genes by diffusing differential expression signals following treatment over a protein interaction network. In addition, drug-protein interactions can also be predicted from structural information. Building on the similar property principle, the second component of our method prioritizes proteins as drug targets based on the interaction with compounds structurally similar to the drug of interest. To this end compound-compound similarity scores are combined with compound-protein interaction scores. Both this structure-based and expression-based approach produce a genome-wide ranking of potential targets that can eventually be fused to obtain a single ranking.

Results/Conclusion
Our method has been evaluated on a test set of small molecule drugs for which the known targets were derived from ChEMBL [5]. AUC values of up to 90% were obtained. These results indicate the predictive power of combining gene expression data and structural information for a drug of interest with known protein-protein and protein-compound interaction information respectively, to identify the targets of that drug. As such this dual method can aid in gaining a better knowledge of a candidate drug’s mode of action and its off-target effects and thus be of value in the drug development process.

References
1. Aronovsmith J, Harrison R. Drug repositioning: the business case and current strategies to repurpose shelved candidates and marketed drugs. In: Baratt MJ, Frail DE, editors. Drug Repositioning: Bringing New Life to Shelved Assets and Existing Drugs. Hoboken, NJ: John Wiley & Sons; Inc, 2012. p. 9-31.
2. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. 2008; 4(11):682-690.
3. Laenen G, Thoresz L, Börnigen D, Moreau Y. Finding the targets of a drug by integration of gene expression data with a protein interaction network. Mol. BioSyst. 2013; 9:1678-1685.
4. Laenen G, Aradeshirdavani A, Moreau Y, Thoresz L, Galahad: a web server for drug effect analysis from gene expression. Nucl. Acids Res. 2015; 43:W208-212.
5. Bento AP, Gauton A, Hersey A, Bells LJ, Chambers J, Davies M, Krüger FA, Light Y, Mak L, McLachlin S, Nowotka M, Papadatos G, Santos R, Overington JP. The ChEMBL bioactivity database: an update. Nucl. Acids Res. 2014; 42:D1083-D1090.

O2
Organism specific protein-RNA recognition: A computational analysis of protein-RNA complex structures from different organisms

Nagarajan Raju1, Sonia Pankaj Chothani1,2, C. Ramakrishnan1, Masakazu Sekijima1,2, M. Michael Gromiha1

1Department of Biotechnology, Bhuban and Jyoti Metha School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, Tamilnadu, India; 2Philips Research North America, 345 Scarborough Road, Briarcliff Manor, NY10510, USA; 3Global Scientific Information and Computing Center (GSIC), Tokyo Institute of Technology 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

Correspondence: M. Michael Gromiha (gromiha@iitm.ac.in) – Department of Biotechnology, Bhuban and Jyoti Metha School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, Tamilnadu, India

BMC Bioinformatics 2016; 17(Suppl 3):O2
Motivation
Protein-RNA interactions play essential roles in many cellular processes. It is unclear whether same RNA binding proteins from different organisms show unique patterns or variations to recognize RNA. To address this issue, we have constructed 18 sets of same protein-RNA complexes belonging to different organisms and analyzed the interactions and interacting patterns using various sequence and structure based features [1].

Results
We have investigated the recognizing elements by grouping the protein chains into five organisms such as E. coli, H. sapiens, S. cerevisiae, archaea and thermophiles [2]. We observed that positively charged residues are highly preferred in E. coli whereas aromatic residues and polar residues show preference in S. cerevisiae and thermophiles, respectively. In case of RNA, adenine and uracil are highly preferred in H. sapiens and S. cerevisiae, respectively. The neighboring residues around the binding sites are unique in different organisms (Eg. Cys-His in H. sapiens, Ala-leu in E. coli, Ser-Arg in S. cerevisiae, Gly-Arg in archaea and Asn-Lys in thermophiles). Further, molecular dynamics simulations of aspartyl tRNA synthetase complexes from E. coli, T. thermophilus and S. cerevisiae revealed the similarities and differences in structurally equivalent binding site residues to understand the recognition mechanism.

Conclusion
Sequence and structural analysis along with MD simulation showed the variations in the interactions between protein and RNA to understand the recognition mechanism of protein-RNA complexes.

References
1. Raju N, Gromiha MM. Prediction of RNA binding residues: An extensive analysis based on structure and function to select the best predictor. PLoS One 2014; 9 (3): e91140.

2. Raju N, Chothani SP, Ramakrishnan C, Sekijima M, Gromiha MM. Structure based approach for understanding organism specific recognition of protein-RNA complexes. Biology Direct. 2015; 10: 8.

O3 Detection of Heterogeneity in Single Particle Tracking Trajectories
Paddy J Slator1,2, Nigel J Burroughs1
1Systems Biology Centre, University of Warwick, Senate House, Coventry, UK; 2Systems Biology Doctoral Training Centre, University of Warwick, Senate House, Coventry, UK
Correspondence: Paddy J Slator (p.j.slator@warwick.ac.uk) – Systems Biology Doctoral Training Centre, University of Warwick, Senate House, Coventry, UK
BMC Bioinformatics 2016, Volume 17 Suppl 3: O3

Background
Single particle tracking trajectories are fundamentally stochastic, which makes the extraction of robust biological conclusions difficult. This is especially the case when trying to detect heterogeneous movement of molecules in the plasma membrane. This heterogeneity could be due to a number of biophysical processes such as: receptor clustering [1], traversing lipid microdomains [2, 3] or cytoskeletal barriers [4].

Results
Working in a Bayesian framework, we developed multiple models for heterogeneity, such as confinement in a harmonic potential well, switching between diffusion coefficients and diffusion in a fenced environment (or ‘hop’ diffusion). We implement these models using a Markov chain Monte Carlo (MCMC) methodology, developing algorithms that infer model parameters and hidden states from single trajectories. We also calculate model selection statistics, to determine the most likely model given the trajectory. Our methodology also accounts for measurement noise. For LFA-1 receptors diffusing on T cells we previously showed that 12-26 % of trajectories display clear switching between diffusive states, depending on treatment (example trajectory in Fig. 1) [5].

Analysis of the motion of GM1 lipids bound to the cholera toxin B subunit in model membranes confirmed transient trapping in harmonic potential wells. We developed an algorithm which detects hopping diffusion, and validated on simulated data (Fig. 2). We have also demonstrated that allowing for measurement noise is essential, as otherwise false detection of heterogeneity may be observed.

Conclusions
We have used Bayesian methodology to analyze single particle tracking trajectories. Rather than methods which rely on generic properties of Browninan motions, our approach allows us to test which biophysical model best fits a trajectory. With the continuing improvement in spatial and temporal resolution of trajectories, these methods will be important for biological interpretation of single particle tracking experiments.

References
1. Bray, D., Levin, M.D., Morton-Firth, C.J.: Receptor clustering as a cellular mechanism to control sensitivity. Nature 393(6680), 8588(1998).

2. Sahi, S.J., Leutenegger, M., Hilbert, M., Hell, S.W., Eggeling, C.: Fast molecular tracking maps nanoscale dynamics of plasma membrane lipids. Proceedings of the National Academy of Sciences 107(15), 6829-6834 (2010).

3. Lingwood, D., Simons, K.: Lipid rafts as a membrane-organizing principle. Science 327(5961), 46-50 (2010).

4. Kusumi, A., Suzuki, K.G.N., Kasai, R.S., Ritchie, K., Fujiwara, T.K: Hierarchical mesoscale domain organization of the plasma membrane. Trends in Biochemical Sciences 36(11), 604-615 (2011).

5. Slator, P.J., Cairo, C.W., Burroughs, N.J.: Detection of Diffusion Heterogeneity in Single Particle Tracking Trajectories Using a Hidden Markov Model with Measurement Noise Propagation. PLOS ONE 10(10), 0140759 (2015).

Fig. 1 (abstract O3). MCMC fit of two-state diffusion model with measurement noise to an LFA-1 trajectory, from [5]. Colour denotes inferred diffusion state, with green slow diffusion and blue fast diffusion. Colorbar length 100 nm
O4
3D-NOME: 3D NucleOme Multiscale Engine for data-driven modeling of three-dimensional genome architecture
Przemysław Szalaj1,2, Zhonghui Tang3, Paul Michalski3, Oskar Luo3, Xingwang Li4, Yijian Ruan1,4, Dariusz Plwczyński1
1Center for Bioinformatics and Data Analysis, Medical University of Białystok, Poland; 2I-BioStat, Hasselt University, Belgium; 3The Jackson Laboratory for Genomic Medicine, Farmington, USA; 4Centre of New Technologies, Warsaw University, Poland
Correspondence: Przemysław Szalaj (przemek.szalaj@uhasselt.be) – I-BioStat, Hasselt University, Belgium
BMC Bioinformatics 2016, 17(Suppl 3):O4

Background
Human genome is folded into three-dimensional structures. The 3D organization of the genome is thought to facilitate compartmentalization, chromatin organization and spatial interaction of genes and their regulatory elements. Recently developed high-throughput ChIA-PET method allows us to capture the genome-wide map of physical contacts between distal genomic loci.

Methods
We present the 3D-NOME, a multi-scale computational engine we developed to model the 3D organization of the genome. First, we use a bottom-up approach to create a hierarchical model representing the nucleus based on the underlying data features. Then we use Monte Carlo simulations to sequentially reconstruct all the levels in a top-down manner, i.e. we reconstruct more general levels first and we use them to guide the simulation on the following levels. This approach allows us to efficiently model the chromatin folding on a level of whole chromosomes as well as single topological domains. In our modeling we consider CTCF (which is known to be responsible for chromatin weaving) and RNA Pol II (which activates genes transcription) interactions. Taken together these two protein factors provide a comprehensive map of human genome interactions. We also consider CTCF-motif orientations in order to obtain more reliable structures. The specificity of the ChIA-PET data allows us to model the shape of individual chromatin loops and their mutual interactions within topological domains boundaries.

We demonstrate the effectiveness of 3D-NOME in building 3D genome models at the kilobase resolution using ChIA-PET data of human B-lymphocytes (GM12878 cell line) [1].

Conclusion
In this work we describe both the model construction and simulation steps of our algorithm. We do also highlight main advantages of our approach compared to existing methods for genome architecture modeling. We hope that further refinement of 3D-NOME and application to additional ChIA-PET and other types of 3D genome mapping data will help to advance our understanding of genome structural organization and functioning.

References
1. Guoliang Li, Fullwood M J, Xu H, et al. (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol 11:R22.

O5
A novel feature selection method to extract multiple adjacent solutions for viral genomic sequences classification
Giulia Fisccon1,2, Emanuel Weitschek1,2, Massimo Ciccozzi3, Paola Bertolazzi2, Giovanni Felici2
1Department of Computer, Control and Management Engineering, Sapienza University, 00185-Rome, Italy; 2Institute for Systems Analysis and Computer Science, National Research Council, 00185-Rome, Italy; 3Department of Engineering, International University UNINETTUNO, 00186-Rome, Italy; 4Istituto Superiore di Sanità, 00161-Rome, Italy
Correspondence: Giulia Fisccon (fisccon@dis.uniroma1.it) – Institute for Systems Analysis and Computer Science, National Research Council, 00185-Rome, Italy
BMC Bioinformatics 2016, 17(Suppl 3):O5

Background
Leveraging improvements of next generation technologies, genome sequencing of several samples in different conditions led to an exponential growth of biological sequences. However, these collections are not easily treatable by biologists to obtain a thorough data characterization and require a high cost-time investment. Therefore, computing strategies and specifically automatic knowledge extraction methods that optimize the analysis focusing on what data are meaningful and should be sequenced are essential [1].

Methods
Here, we present a new feature-selection algorithm based on mixed integer programming methods [2] able to extract multiple and adjacent solutions for supervised learning problems applied to biological data. We focus on those problems where the relative position of a feature (i.e., nucleotide locus) is relevant. In particular, we aim to find sets of distinctive features, which are as close as possible to each other and which appear with the same required characteristics. Our algorithm adopts a fast and effective method to evaluate the quality of the extracted sets of features and it has been successfully integrated in a rule-based classification framework [3].

Results
Our algorithm has been applied to three viral datasets (i.e., Rhino-, Influenza-, Polyomaviruses [4-6]) and enables to extract all the alternative solutions of virus specimen to species assignments, by identifying portions of sequence that are discriminant, compact, and as shorter as possible. To conclude, we succeeded in extracting a wide set of equivalent classification rules, focusing on short regions of sequences with high reliability and low computational time, in order to provide the biologists with short and highly informative genome parts to be sequenced, as well as a powerful instrument both scientifically and diagnostically, e.g., for automatic virus detection.

References
1. Chen JY and S. Lonardi S, Biological Data Mining, Chapman & Hall, 2010.
2. Bertolazzi P, Felici G, Festa P, Fisccon G, and Weitschek E. Integer programming models for feature selection: new extensions and a randomized solution algorithm. European Journal of Operational Research, 1-20, 2015, (accepted).
3. Weitschek E, Velzen R, Felici G, and Bertolazzi P. Blog 2.0: a software system for character-based species classification with dna barcode sequences. what it does, how to use it. Molecular ecology resources, 13(6):1043-1046, 2013.
The protozoan parasite Leishmania donovani is the causative agent of visceral leishmaniasis in the Indian subcontinent and poses a threat to public health due to increasing drug resistance. Only little is known about differences in clinical features between influenza A H1N1, A H3N2, and B in adult patients. Respirology, 8(2):231-233, 2003.

4. Tapparel C, Junier T, Gerlach D, Cordey S, Van Belle S, Perin L, Zdobnov EM, and Kaiser L. New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features. BMC genomics, 8(1):224, 2007.

5. Weitschek W, Lo Presti A, Drovandi G, Felici G, Ciccozzi M, Ciotti M, and Bertolazzi P. Human polymaviruses identification by logic mining techniques. Virology journal, 9(1):1-6, 2012.

6. Kaji M, Watanabe A, and Aizawa H. Differences in clinical features between influenza A H1N1, A H3N2, and B in adult patients. Respirology, 8(2):231-233, 2003.

O6 A Systems Biology Compendium for Leishmania donovani
Bart Cuypers1,*, Pieter Meysman2,*, Manu Vanaerschot3, Maya Berg4, Hideo Imamura5, Jean-Claude Dujardin6,†, Kris Laukens1,†, Jean-Claude Dujardin and Kris Laukens contributed equally to this work.

Background
The protozoan parasite Leishmania donovani is the cause of visceral leishmaniasis in the Indian subcontinent and poses a threat to public health due to increasing drug resistance. Only little is known about its peculiar molecular biology and the 'omics integration efforts conducted so far are very limited. Here we present an integratory database that contains all genomics, transcriptomics, proteomics and metabolomics experiments that are currently publicly available for Leishmania donovani. In addition, the user interface contains new analysis tools that use powerful pattern mining strategies like frequent itemset mining to allow the linking of results from different 'omics layers in new datasets.

Methods
We developed a user friendly tool to crosslink all existing L. donovani – 'omics experiments. Genomics, transcriptomics, proteomics, metabolomics and phenotypic data were collected and added to a MySQL database compendium, further complemented with publicly available data. Relations between different 'omics layers were explicitly defined and provided with a level of confidence. Python scripts were developed to preprocess, analyse and import the data. To allow comparability between different experiments the principles of the COLOMBOS bacterial expression compendium were adapted. Next to this vast data resource, a set of integrative data-analysis tools was developed based on data mining strategies. For example: One tool uses frequent itemset mining algorithms to detect which proteins and metabolites frequently exhibit the same behaviour under different conditions. Another tool converts several –omics layers to a network format that can be opened in Cytoscape [2] thus be the basis for network analysis. Django and Twitter Bootstrap frameworks were used to create a web portal to make the tools accessible to any Leishmania researcher.

Results
Excellent public genome, protein and metabolite annotation databases are already available for Leishmania (e.g. TriTrypDB [3] and GeneDB [4]). However, the added value of our tool is that it links these annotation data to 'omics experiments that are either provided by the user, or publicly available. New experiments can quickly be preprocessed, analysed and integrated in the database via its Python back end. Using the compendium and its tools, we characterized the development and drug-resistance of Leishmania donovani in a system biology context. The genomes of more than 200 strains were examined for associations with phenotypical features and a subset was linked to transcriptomics, proteomics and metabolomics results. The compendium and its scripts were designed to be generic and can therefore be used for other organisms with only minor adaptations to the original setup.

O7 Unravelling signal coordination from large scale phosphorylation kinetic data
Westa Domanova1,2, James R. Krycer1,3, Rima Chaudhuri1,3, Pengyi Yang4, Fatemeh Vafaeef5,6, Daniel J. Fazakerley1,3, Sean J. Humphrey7, David E. James1,3,4, Zdenka Kuncic8,9

Background
A growing body of evidence is emerging that the temporal behavior of cellular signaling molecules controls biological responses. Phosphorylation, one of the most prevalent signaling modifications, occurs rapidly in response to environmental changes. Our understanding of the signaling network is incomplete because the kinase for the majority of phosphorylation substrates remains unknown. To elucidate the underlying topology of signaling cascades from high-throughput data, we need to be able to predict kinase substrate relationships. Currently, prediction algorithms ignore the crucial biological context of kinase substrate relationships.

Methods
To address this we consider temporal behaviours: given that phosphorylation events occur in a coordinated way, with some kinases being active before others, we predict site-specific kinase substrate relationships from large scale in vivo experiments. First, we used available phosphorylation databases to construct a publicly-
accessible repository of experimentally-predicted kinase substrate relationships. Next, based on the substrates reported for each kinase in this database, we identify how the kinase activities change over time in temporal datasets.

**Results**

Applying this to an insulin-stimulated phosphorylation screen we were able to distinguish between the substrates of AKT and RPS6KB1, two kinases with the same consensus motif, and identified IRS-1-S270 as a novel putative AKT site. We subsequently used our ssKSR-LIVE algorithm to predict novel substrates for the kinases driving insulin signaling, shedding light on their role in driving insulin-stimulated biological processes. This algorithm can be applied to other high-throughput screens of signal transduction, and thus can be used to improve our understanding of complex diseases caused by dysregulated signalling, including cancer and type 2 diabetes.