Program of the Tuesday, September 9

Tuesday, September 9

7:45	Registration opens		
8:45	**Keynote 4: Nada LAVRAČ.** ***Advances in data mining for biomedical research.***		Auditorium ERASME
9:35	Distribution in the two parallel sessions		
	Auditorium ERASME	Room SCHUMANN	
	Session Tue1: Computational Systems Biology (1)	**Session Tue4: Biological Knowledge Discovery from Data**	
9:40	*PP20 - Stronger findings for metabolomics through Bayesian modeling of multiple peaks and compound correlations*. Tommi Suvitaival	*PP28 - Unveiling new biological relationships using shared hits of chemical screening assay pairs.* Monica Campillos	
10:05	*PP21 - Causal network inference using biochemical kinetics*. Chris Oates	*PP29 - Identification of structural features in chemicals associated with cancer drug response: A systematic data-driven analysis*. Suleiman Ali Khan	
10:30	**Highlight Talk:** *HP05 - High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling*. Sabine Hug	**Highlight Talk:** *HP06 - Shaping the interaction landscape of bioactive molecules*. David Gfeller	
10:55	Coffee Break		Main Floor & 1st Floor
	Session Tue2: Computational Systems Biology (2)	**Session Tue5: Gene Expression (1)**	
11:15	*PP22 - Effects of small particle numbers on long-term behaviour in discrete biochemical systems*. Peter Dittrich	*PP30 - Estimating the activity of transcription factors by the effect on their target genes*. Rainer Koenig	
11:40	*PP23 - TEMPI: Probabilistic modeling time-evolving differential PPI networks with multiple information.* Yongsoo Kim	*PP31 - Modeling DNA methylation dynamics with approaches from phylogenetics*. Dennis Kostka	
12:05	*PP24 - Experimental design schemes for learning Boolean network models.* Nir Atias	**Highlight Talk:** *HP07 - Key regulators control distinct transcriptional programmes in blood and progenitor and mast cells*. Felicia Ng	
12:30	LUNCH		Dining Room Contades
13:30	**Industrial and Demo Track**		See corresponding pages
14:35	**Keynote 5: Ewan Birney.** ***Big Data in Biology.***		Auditorium ERASME
15:25	Distribution in two parallel sessions		
	Auditorium ERASMUS	Room SCHUMANN	
	Session Tue3: Bioinformatics of Health and Disease (1)	**Session Tue6: Gene Expression (2)**	
15:30	*PP25 - OncodriveROLE classifies cancer driver genes in Loss of Function and Activating mode of action.* Michael Philipp Schroeder	*PP32 - Two-dimensional segmentation for analyzing HiC data*. Celine Levy-Leduc	
15:55	*PP26 - ContrastRank: a new method for ranking putative cancer driver genes and classification of tumor samples.* Emidio Capriotti	*PP33 - Broad-Enrich: Functional interpretation of large sets of broad genomic regions*. Raymond Cavalcante	
16:20	*PP27 - Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask*	**Highlight Talk:** *HP08 - Chromatin position effects quantified from thousands of reporters*	

	learning. Mehmet Gönen	*integrated in parallel*. Lodewyk Wessels	
16:45	Coffee Break - Poster Session begins		Main Foor & 1st Floor
17:30	1st departure to river boat trip (then every ¼ hour)		Main Floor
18:45	End of Poster Session		1st Floor
19:15	Last departure to river boat trip or mini-train		Main Floor
From 19:00	Gala Evening at the Council of Europe or at l'Ancienne Douane		

Tue2 (Area D): Computational Systems Biology (2)

Chairs: Oliver Kohlbacher, Anne Siegel

PP22 - Effects of small particle numbers on long-term behaviour in discrete biochemical systems

Peter Kreyssig[1], Christian Wozar[1], Stephan Peter[1], Tomas Veloz[2,3,4], Bashar Ibrahim[1,5,6] and Peter Dittrich[1]

[1]Bio Systems Analysis Group, Department of Mathematics and Computer Science and Jena Centre for Bioinformatics, Friedrich Schiller University Jena, 07743 Jena, Germany. [2]Mathematics Department, University of British Columbia, Kelowna, BC V1V 1V7, Canada. [3]Instituto de Filosofía y Ciencias de la Complejidad - IFICC, Los Alerces 3024 Ñuñoa, Santiago, Chile. [4]Center Leo Apostel, Vrije Universiteit Brussel, Krijgskundestraat 33, B-1160 Brussels, Belgium. [5]Umm Al-Qura University, 1109 Makkah Al-Mukarramah, Kingdom of Saudi Arabia. [6]Al-Qunfudah Center for Scientific Research (QCSR), 21912 Al-Qunfudah, Kingdom of Saudi Arabia.

ABSTRACT

Motivation: The functioning of many biological processes depends on the appearance of only a small number of a single molecular species. Additionally, the observation of molecular crowding leads to the insight that even a high number of copies of species does not guarantee their interaction. How single particles contribute to stabilising biological systems is not well understood yet. Hence we aim at determining the influence of single molecules on the long-term behaviour of biological systems, *i.e.* whether they can reach a steady state or not.

Results: We provide theoretical considerations and a tool to analyse SBML models for the possibility to stabilise due to the described effects. The theory is an extension of chemical organisation theory which we called discrete chemical organisation theory. Furthermore we scanned the BioModels Database for the occurrence of discrete chemical organisations. To exemplify our method we describe an application to the Template model of the mitotic spindle assembly checkpoint mechanism.

Availability: http://www.biosys.uni-jena.de/Services.html

Contact: bashar.ibrahim@uni-jena.de , dittrich@minet.uni-jena.de

Supplementary Information: Supplementary data are available at Bioinformatics online.

PP23 - TEMPI: Probabilistic modeling time-evolving differential PPI networks with multiple information

Yongsoo Kim[1], Jin-Hyeok Jang[1], Seungjin Choi[2] and Daehee Hwang[1,3]

[1]School of Interdisciplinary Bioscience and Bioengineering and [2]Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea, [3]Center for Systems Biology of Plant Senescence and Life History, Institute for Basic Science, Daegu Gyeongbuk Institute of Science and Technology, Daegu 711-873, Korea.

ABSTRACT

Motivation: Time-evolving differential protein-protein interaction (PPI) networks are essential to understand serial activation of differentially regulated (up- or down-regulated) cellular processes (DRPs) and their interplays over time. Despite developments in the network inference, current methods are still limited in identifying temporal transition of structures of PPI networks, DRPs associated with the structural transition, and the interplays among the DRPs over time.

Results: Here, we present a probabilistic model for estimating Time-Evolving differential PPI networks with MultiPle Information (TEMPI). This model describes probabilistic relationships among network structures, time-course gene expression data, and Gene Ontology biological processes (GOBPs). By maximizing the likelihood of the probabilistic model, TEMPI estimates jointly the time-evolving differential PPI networks (TDNs) describing temporal transition of PPI network structures together with serial activation of DRPs associated with transiting networks. This joint estimation enables us to interpret the TDNs in terms of temporal transition of the DRPs. To demonstrate the utility of TEMPI, we applied it to two time-course datasets. TEMPI identified the TDNs that correctly delineated temporal transition of DRPs and time-dependent associations between the DRPs. These TDNs provide hypotheses for mechanisms underlying serial activation of key DRPs and their temporal associations.

Availability: Source code and sample data files are available at http://sbm.postech.ac.kr/tempi/sources.zip .

Contact: seungjin@postech.ac.kr or dhwnag@dgist.ac.kr

PP24 - Experimental design schemes for learning Boolean network models

Nir Atias, Michal Gershenzon, Katia Labazin and Roded Sharan

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel.

ABSTRACT
Motivation: A holy grail of biological research is a working model of the cell. Current modeling frameworks, especially in the protein-protein interaction domain, are mostly topological in nature, calling for stronger and more expressive network models. One promising alternative is logic-based, or Boolean network modeling, which was successfully applied to model signaling regulatory circuits in human. Learning such models requires observing the system under a sufficient number of different conditions. To date, the amount of measured data is the main
bottleneck in learning informative Boolean models, underscoring the need for efficient experimental design strategies.
Results: We developed novel design approaches that greedily select an experiment to be performed so as to maximize the difference or the entropy in the results it induces with respect to current best-fit models. Unique to our maximum difference approach is the ability to account for all (possibly exponential number of) Boolean models displaying high fit to the available data. We applied both approaches to simulated and real data from the EFGR and IL1 signaling systems in human. We demonstrate the utility of the developed strategies in substantially improving on a random selection approach. Our design schemes highlight the redundancy in these data sets, leading up to 11-fold savings in the number of experiments to be performed.
Availability: Source code will be made available upon acceptance of the manuscript.
Contact: roded@post.tau.ac.il

Tue3 (Area G): Bioinformatics of Health and Disease (1)

Chairs: To be announced

PP25 - OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action

Michael P Schroeder[1], Carlota Rubio-Perez[1], David Tamborero[1], Abel Gonzalez-Perez[1,*] and Nuria Lopez-Bigas [1,2]

[1] Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona, Spain. [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, Barcelona, Spain.

ABSTRACT
Motivation: Several computational methods have been developed to identify cancer drivers genes – genes responsible for cancer development upon specific alterations. These alterations can cause the loss of function of the gene product, for instance in tumor suppressors, or increase or change its activity or function, if it is an oncogene. Distinguishing between these two classes is important to understand tumorigenesis in patients and has implications for therapy decision making. Here, we assess the capacity of multiple gene features related to the pattern of genomic alterations across tumors to distinguish between activating and loss of function cancer genes and we present an automated approach to aid the classification of novel cancer drivers according to their role.
Result: OncodriveROLE is a machine learning-based approach that classifies driver genes according to their role, using several properties related to the pattern of alterations across tumors. The method shows an accuracy of 0.93 and Matthew's Correlation Coefficient of 0.84 classifying genes in the Cancer Gene Census. The OncodriveROLE classifier, its results when applied to two list of predicted cancer drivers and TCGA-derived mutation and copy number features used by the classifier are available at http://bg.upf.edu/oncodrive-role.
Contact: abel.gonzalez@upf.edu , nuria.lopez@upf.edu

PP26 - ContrastRank: a new method for ranking putative cancer driver genes and classification of tumor samples

Rui Tian[1], Malay Basu[1,2] and Emidio Capriotti[1,2,3]

[1]Division of Informatics, Department of Pathology, University of Alabama at Birmingham, 619 19th St. South, 35249 Birmingham, AL, USA. [2]Department of Clinical and Diagnostic Sciences, University of Alabama at Birmingham, 1705 University Boulevard, 35249 Birmingham, AL, USA. [3]Department of Biomedical Engineering, University of Alabama at Birmingham, 1075 13th Street South, 35249 Birmingham, AL, USA.

ABSTRACT
Motivation: The recent advance in high-throughput sequencing technologies is generating a huge amount of data that are becoming an important resource for deciphering the genotype underlying a given phenotype. Genome sequencing has been extensively applied to the study of the cancer genomes. Although a few methods have been already proposed for the detection of cancer-related genes, their automatic identification is still a challenging task. Using the genomic data made available by The Cancer Genome Atlas Consortium

(TCGA), we propose a new prioritization approach based on the analysis of the distribution of putative deleterious variants in a large cohort of cancer samples.
Results: In this paper, we present ContastRank, a new method for the prioritization of putative impaired genes in cancer. The method is based on the comparison of the putative defective rate of each gene in tumor versus normal and 1000 genome samples. We show that the method is able to provide a ranked list of putative impaired genes for colon, lung and prostate adenocarcinomas. The list significantly overlaps with the list of known cancer driver genes previously published. More importantly, by using our scoring approach, we can successfully discriminate between TCGA normal and tumor samples. A binary classifier based on ContrastRank score reaches an overall accuracy higher than 90% and the Area Under the Curve (AUC) of Receiver Operating Characteristics (ROC) higher than 0.95 for all the three types of adenocarcinoma analysed in this paper. In addition, using ContrastRank score we are able to discriminate the three tumor types with a minimum overall accuracy of 77% and AUC of 0.83.
Conclusions: We describe ContrastRank, a method for prioritizing putative impaired genes in cancer. The method is based on the comparison of exome sequencing data from different cohorts and can detect putative cancer driver genes. ContrastRank can also be used to estimate a global score for an individual genome about the risk of adenocarcinoma based on the genetic variants information from a whole-exome VCF (Variant Calling Format) file. We believe that the application of ContrastRank can be an important step in genomic medicine to enable genome-based diagnosis.
Availability: The lists of ContrastRank scores of all genes in each tumor type are available as supplementary materials. A webserver for evaluating the risk of the three studied adenocarcinomas starting from whole-exome VCF file is under development.
Contact: emidio@uab.edu

PP27 - Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning

Mehmet Gönen and Adam A. Margolin

Sage Bionetworks, 1100 Fairview Avenue North, Seattle, WA 98109, USA. Present address: Department of Biomedical Engineering, Oregon Health & Science University, 3303 SW Bond Avenue, Portland, OR 97239, USA.

ABSTRACT
Motivation: Human immunodeficiency virus (HIV) and cancer require personalized therapies due to their inherent heterogeneous nature. For both diseases, large-scale pharmacogenomic screens of molecularly characterized samples have been generated with the hope of identifying genetic predictors of drug susceptibility. Thus, computational algorithms capable of inferring robust predictors of drug responses from genomic information are of great practical importance. Most of the existing computational studies that consider drug susceptibility prediction against a panel of drugs formulate a separate learning problem for each drug, which cannot make use of commonalities between subsets of drugs.
Results: In this study, we propose to solve the problem of drug susceptibility prediction against a panel of drugs in a multi-task learning framework by formulating a novel Bayesian algorithm that combines kernel-based nonlinear dimensionality reduction and binary classification (or regression). The main novelty of our method is the joint Bayesian formulation of projecting data points into a shared subspace and learning predictive models for all drugs in this subspace, which helps us to eliminate off-target effects and drug-specific experimental noise. Another novelty of our method is the ability of handling missing phenotype values due to experimental conditions and quality control reasons. We demonstrate the performance of our algorithm via cross-validation experiments on two benchmark drug susceptibility datasets of HIV and cancer. Our method obtains statistically significantly better predictive performance on most of the drugs compared to baseline single-task algorithms that learn drug-specific models. These results show that predicting drug susceptibility against a panel of drugs simultaneously within a multi-task learning framework improves overall predictive performance over single-task learning approaches.
Availability: Our Matlab implementations for binary classification and regression are available at https://github.com/mehmetgonen/kbmtl .
Contact: mehmet.gonen@sagebase.org

Tue4 (Area H): Biological Knowledge Discovery from data

Chairs: To be announced

PP28 - Unveiling new biological relationships using shared hits of chemical screening assay pairs

Xueping Liu[1,2] and Monica Campillos[1,2]

[1]Institute of Bioinformatics and Systems Biology and [2]German Center for Diabetes Research, Helmholtz Center Munich, 85764, Neuherberg, Germany.

ABSTRACT
Motivation: Although the integration and analysis of the activity of small molecules across multiple chemical screens is a common approach to determine the specificity and toxicity of hits, the suitability of these approaches to reveal novel biological information is less explored. Here, we test the hypothesis that assays sharing selective hits are biologically related.
Results: We annotated the biological activities (i.e. biological processes or molecular activities) measured in assays and constructed chemical hit profiles with sets of compounds differing on their selectivity level for 1,640 assays of ChemBank repository. We compared the similarity of chemical hit profiles of pairs of assays with their biological relationships and observed that assay pairs sharing non promiscuous chemical hits tend to be biologically related. A detailed analysis of a network containing assay pairs with the highest hit similarity confirmed biological meaningful relationships. Furthermore, the biological roles of predicted molecular targets of the shared hits reinforced the biological associations between assay pairs.
Contact: monica.campillos@helmholtz-muenchen.de

PP29 - Identification of structural features in chemicals associated with cancer drug response: A systematic data-driven analysis

Suleiman Ali Khan[1], Seppo Virtanen[1], Olli Kallioniemi[2], Krister Wennerberg[2], Antti Poso[2,3] and Samuel Kaski[1,4]

[1]Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, 00076 Espoo, Finland. [2]Institute for Molecular Medicine Finland FIMM, University of Helsinki, 00014 Helsinki, Finland. [3]School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211 Kuopio, Finland. [4]Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland.

ABSTRACT
Motivation: Analysis of relationships of drug structure to biological response is key to understanding off-target and unexpected drug effects, and for developing hypotheses on how to tailor drug therapies. New methods are required for integrated analyses of a large number of chemical features of drugs against the corresponding genome-wide responses of multiple cell models.
Results: In this paper, we present the first comprehensive multi-set analysis on how the chemical structure of drugs impacts on genome-wide gene expression across several cancer cell lines (CMap database). The task is formulated as searching for drug response components across multiple cancers to reveal shared effects of drugs and the chemical features that may be responsible. The components can be computed with an extension of a very recent approach called Group Factor Analysis (GFA). We identify 11 components that link the structural descriptors of drugs with specific gene expression responses observed in the three cell lines, and identify structural groups that may be responsible for the responses. Our method quantitatively outperforms the limited earlier methods on CMap and identifies both the previously reported associations and several interesting novel findings, by taking into account multiple cell lines and advanced 3D structural descriptors. The novel observations
include: previously unknown similarities in the effects induced by 15-delta prostaglandin J2 and HSP90 inhibitors, which are linked to the 3D descriptors of the drugs; and the induction by simvastatin of leukemia-specific response, resembling the effects of corticosteroids.
Availability: Code http://research.ics.aalto.fi/mi/software/GFAsparse
Contact: suleiman.khan@aalto.fi, samuel.kaski@aalto.fi
Supplementary Information: Available at *Bioinformatics* online

Highlight Talk: *HP06 - Shaping the interaction landscape of bioactive molecules*

David Gfeller, Aurelien Grosdidier, Matthias Wirth, Antoine Daina, Olivier Michielin and Vincent Zoete

Swiss Institute of Bioinformatics, Lausanne, Switzerland.

ABSTRACT
Bioactive small molecules, such as drugs or metabolites, interact with proteins targets to modulate their activity, which in turn results in the observed phenotypic effects. However, for most bioactive compounds the list of targets is only partially known. Therefore computational predictions of bioactive molecule targets are powerful to narrow down the number of potential targets and to rationalize side effects of known molecules. Here, we introduce a new computational approach to accurately predict the targets of bioactive small molecules based on a combination of 2D and 3D similarity measures with known ligands. The method is trained on a large dataset of 280,381 small molecules interacting with 2686 targets from the ChEMBL database. Predictions can be carried out in five different organisms, and mapping predictions by homology within and between different species is enabled for close paralogs and orthologs. The method is accessible free of charge at http://www.swisstargetprediction.ch.

Publication:
Gfeller D, Michielin O, Zoete V. Shaping the interaction landscape of bioactive molecules. Bioinformatics. 2013 Dec 1;29(23):3073-9.

Contact: david.gfeller@gmail.com

Tue5 (Area B): Gene Expression (1)

Chair: Morgane Thomas-Chollier

PP30 - Estimating the activity of transcription factors by the effect on their target genes

Theresa Schacht[1,2,3], Marcus Oswald[1,2], Roland Eils[3,4], Stefan Eichmüller[5] and Rainer Koenig[1,2,3]

[1]Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, D-07747 Jena, Erlanger Allee 101, Germany. [2]Network Modeling, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute Jena, Beutenbergstrasse 11a, 07745 Jena, Germany. [3]Theoretical Bioinformatics, German Cancer Research Center, INF 580, 69121 Heidelberg, Germany. [4]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, and Bioquant, University of Heidelberg, Im Neuenheimer Feld 267, Germany. [5]Translational Immunology, German Cancer Research Center (DKFZ), INF 280, 69120 Heidelberg, Germany.

ABSTRACT

Motivation: Understanding regulation of transcription is central for elucidating cellular regulation. Several statistical and mechanistic models have come up the last couple of years explaining gene transcription levels using information of potential transcriptional regulators as transcription factors (TFs) and information from epigenetic modifications. The activity of TFs is often inferred by their transcription levels, promoter binding and epigenetic effects. However, in principle, these methods do not take hard-to-measure influences such as post-transcriptional modifications into account.

Results: For TFs, we present a novel concept circumventing this problem. We estimate the regulatory activity of TFs using their cumulative effects on their target genes. We established our model using expression data of 59 cell lines from the National Cancer Institute. The trained model was applied to an independent expression dataset of melanoma cells yielding excellent expression predictions and elucidated regulation of melanogenesis.

Implementation: Using mixed integer linear programming (MILP), we implemented a switch like optimization enabling a constrained but optimal selection of TFs and optimal model selection estimating their effects. The method is generic and can also be applied to further regulators of transcription.

Contact: Rainer.Koenig@uni-jena.de

PP31 - Modeling DNA methylation dynamics with approaches from phylogenetics

John A. Capra[1] and Dennis Kostka[2]

[1]Center for Human Genetics Research and Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, 37232, USA. [2]Departments of Developmental Biology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA, 15201, USA.

ABSTRACT

Motivation: Methylation of CpG dinucleotides is a prevalent epigenetic modification that is required for proper development in vertebrates. Genome-wide DNA methylation assays have become increasingly common, and this has enabled characterization of DNA methylation in distinct stages across differentiating cellular lineages. Changes in CpG methylation are essential to cellular differentiation; however, current methods for modeling methylation dynamics do not account for the dependency structure between precursor and dependent cell types.

Results: We developed a continuous-time Markov chain approach, based on the observation that changes in methylation state over tissue differentiation can be modeled similarly to DNA nucleotide changes over evolutionary time. This model explicitly takes precursor to descendant relationships into account and enables inference of CpG methylation dynamics. To illustrate our method, we analyzed a high-resolution methylation map of the differentiation of mouse stem cells into several blood cell types. Our model can successfully infer unobserved CpG methylation states from observations at the same sites in related cell types (90% correct), and this approach more accurately reconstructs missing data than imputation based on neighboring CpGs (84% correct). Additionally, the single CpG resolution of our methylation dynamics estimates enabled us to show that DNA sequence context of CpG sites is informative about methylation dynamics across tissue differentiation. Finally, we identified genomic regions with clusters of highly dynamic CpGs and present a likely functional example. Our work establishes a framework for inference and modeling that is well-suited to DNA methylation data, and our success suggests that other methods for analyzing DNA nucleotide substitutions will also translate to the modeling of epigenetic phenomena.

Availability: Source code is available at www.kostkalab.net/software .

Contact: tony.capra@vanderbilt.edu , kostka@pitt.edu

Highlight Talk: *HP07 - Key regulators control distinct transcriptional programmes in blood and progenitor and mast cells*

Felicia Ng, Fernando Calero-Nieto, Nicola Wilson, Rebecca Hannah, Victoria Moignard, Ana Leal-Cervantes, Isabel Jimenez-Madrid, Evangelia Diamanti, Lorenz Wernisch and Berthold Gottgens

Cambridge Institute for Medical Research, United Kingdom.

ABSTRACT

Despite major advances in the generation of genome-wide binding maps, the mechanisms by which transcription factors (TFs) regulate cell type identity have remained largely obscure. Through comparative analysis of 10 key haematopoietic TFs in both mast cells and blood progenitors, we demonstrate that the largely cell type-specific binding profiles are not opportunistic, but instead contribute to cell type-specific transcriptional control, because (i) mathematical modelling of differential binding of shared TFs can explain differential gene expression, (ii) consensus binding sites are important for cell type-specific binding and (iii) knock-down of blood stem cell regulators in mast cells reveals mast cell-specific genes as direct targets. Finally, we show that the known mast cell regulators Mitf and c-fos likely contribute to the global reorganisation of TF binding profiles. Taken together therefore, our study elucidates how key regulatory TFs contribute to transcriptional programmes in several distinct mammalian cell types.

Publication:

Calero-Nieto FJ, Ng FS, Wilson NK, Hannah R, Moignard V, Leal-Cervantes AI, Jimenez-Madrid I, Diamanti E, Wernisch L, Göttgens B. Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. EMBO J. 2014 Jun 2;33(11):1212-26.

Contact: fn231@cam.ac.uk

Tue6 (Area F): Gene expression (2)

Chairs: To be announced

PP32 - Two-dimensional segmentation for analyzing HiC data

Celine Levy-Leduc[1], Maud Delattre[1], Tristan Mary-Huard[1,2] and Stephane Robin[1]

[1]AgroParisTech/INRA MIA 518, 16 rue Claude Bernard, 75005 Paris, France. [2]UMR de Génétique Végétale, INRA/Université Paris-Sud/CNRS, 91190 Gif-sur-Yvette, France.

ABSTRACT

Motivation: The spatial conformation of the chromosome has a deep influence on gene regulation and expression. HiC technology allows the evaluation of the spatial proximity between any pair of loci along the genome. It results in a data matrix where blocks corresponding to (self-)interacting regions appear. The delimitation of such blocks is critical to better understand the spatial organization of the chromatin. From a computational point of view, it results in a 2D-segmentation problem.

Results: We focus on the detection of cis-interacting regions, which appear to be prominent in observed data. We define a block-wise segmentation model for the detection of such regions. We prove that the maximization of the likelihood with respect to the block boundaries can be rephrased in terms of a 1D-segmentation problem, for which the standard dynamic programming applies. The performance of the proposed methods are assessed by a simulation study on both synthetic and re-sampled data. A comparative study on public data shows good concordance with biologically confirmed regions.

Availability: The HiCseg R package is available from the Comprehensive R Archive Network (CRAN) and from the web page of the corresponding author.

Contact: celine.levy-leduc@agroparistech.fr

PP33 - Broad-Enrich: Functional interpretation of large sets of broad genomic regions

Raymond Cavalcante[1], Chee Lee[1], Ryan Welch[1,2], Snehal Patil[3], Terry Weymouth[3], Laura Scott[2] and Maureen Sartor[1,2,3]

[1]Department of Computational Medicine and Bioinformatics, [2]Department of Biostatistics, and [3]Center of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

ABSTRACT

Motivation: Functional enrichment testing facilitates the interpretation of ChIP-seq data in terms of pathways and other biological contexts. Previous methods developed and used to test for key gene sets affected in ChIP-seq experiments treat peaks as points, and are based on the number of peaks associated with a gene or a binary score for each gene. These approaches work well for transcription factors, but histone modifications often occur over broad domains, and across multiple genes.

Results: To incorporate the unique properties of broad domains into functional enrichment testing, we developed Broad-Enrich, a method that uses the proportion of each gene's locus covered by a peak. We show that our method has a well-calibrated false positive rate, performing well with ChIP-seq data having broad domains compared to alternative approaches. We illustrate Broad-Enrich with 55 ENCODE ChIP-seq datasets using different methods to define gene loci. Broad-Enrich can also be applied to other datasets consisting of broad genomic domains such as copy number variations.
Availability: http://broad-enrich.med.umich.edu for web version and R package.
Contact: sartorma@umich.edu
Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Highlight Talk: *HP08 - Chromatin position effects quantified from thousands of reporters integrated in parallel*

Lodewyk Wessels[1], Waseem Akhtar[1], Johann de Jong[1], Alex Pindyurin[2], Ludo Pagie[1], Wouter Meuleman[3], Jeroen de Ridder[4], Anton Berns[1], Maarten van Lohuizen[1] and Bas van Steensel[1]

[1]Netherlands Cancer Institute, Amsterdam, The Netherlands. [2]Institute of Molecular and Cellular Biology, Novosibirsk, Russian Federation. [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Delft University of Technology, The Netherlands.

ABSTRACT
Reporter genes integrated into the genome are a powerful tool to reveal effects of regulatory elements and local chromatin context on gene expression. However, such assays have been low throughput. Here, we describe an approach to monitor transcriptional activity of thousands of randomly integrated reporters. Computational analyses of more than 27,000 distinct reporter integrations in mouse embryonic stem cells reveal the following. First, lamina associated domains act as attenuators of transcription, likely by reducing access of transcription factors to binding sites. Second, chromatin compaction as derived from HiC data is predictive of reporter activity. Third, we find evidence of cross-talk between neighbouring genes and estimate that enhancers can influence gene expression on average over ~20 kb. Most importantly, the richness and size of the datasets opens up the opportunity for additional extensive and robust computational analyses. We will showcase the utility with recent analyses shedding new light on gene regulation.
Publication:
Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF, van Lohuizen M, van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell. 2013 Aug 15;154(4):914-27.

Contact: l.wessels@nki.nl