ECCB'14

HIGHLIGHT PAPERS ABSTRACTS

HP01: *Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles.*

Enrica Calura, Paolo Martini, Gabriele Sales and Chiara Romualdi.
Department of Biology, University of Padova, via U. Bassi 58/B, 35121 Padova, Italy.

ABSTRACT
The production rate of gene expression data is nothing less than astounding. However, with the benefit of hindsight we can assert that, since we completely ignored the non-coding part of the transcriptome, we spent the last decade to study cell mechanisms having few data in our hands. In this scenario, microRNAs, which are key post-trascriptional regulators, deserve special attention. Currently, miRNA and gene circuits are identified through the combination of binding prediction and expression correlation analyses, MAGIA, the web tool we developed, is an example to feel this aim (Sales et al NAR 2010, Bisognin et al NAR 2012). Although effective in many cases the simple correlation does not imply a causal relationship and a lot of false positive miRNA-mRNA interactions are still found. Moreover, miRNA and target genes are characterized by many-to-many relationships and they should be considered as part of a much more complex system of cellular interactions. Recently, to analyze the cellular circuits we developed a new web tool dedicated to topological pathway analyses called Graphite Web (Sales et al NAR 1013). Given the state of knowledge about the biogenesis of miRNAs, their mechanisms of action and the numerous experimentally validated target genes, miRNAs are also gradually appearing in the formal pathway representations such as KEGG and Reactome maps. However, the number of miRNAs annotated in pathway maps is very small and pathway analyses exploiting this new regulatory layer are still lacking. To fill these gaps, we developed micrographite a new pipeline to perform topological pathway analysis integrating gene and miRNA expression profiles. Micrographite analysis of gene and miRNA integrated transcriptome is used to study and dissect the epithelial ovarian cancer gene complexity and miRNA transcriptome defining and validating a new regulatory circuits.

Publications :

Calura E, Fruscio R, Paracchini L, Bignotti E, Ravaggi A, Martini P, Sales G, Beltrame L, Clivio L, Ceppi L, Di Marino M, Fuso Nerini I, Zanotti L, Cavalieri D, Cattoretti G, Perego P, Milani R, Katsaros D, Tognon G, Sartori E, Pecorelli S, Mangioni C, D'Incalci M, Romualdi C, Marchini S. MiRNA landscape in stage I epithelial ovarian cancer defines the histotype specificities. Clin Cancer Res. 2013 Aug 1;19(15):4114-23.
Calura E, Martini P, Sales G, Beltrame L, Chiorino G, D'Incalci M, Marchini S, Romualdi C. Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles. Nucleic Acids Res. 2014;42(11):e96.

Contact: enrica.calura@unipd.it

HP02: *Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data.*

Ségolène Caboche[1], Christophe Audebert[2], Yves Lemoine[3] and David Hot[2,3].
[1]FRE 3642 Molecular and Cellular Medecine, CNRS, Institut Pasteur de Lille and University of Lille Nord de France. [2]Genes Diffusion, Douai, France. [3]Transcriptomics and Applied Genomics, Center for Infection and Immunity of Lille, Inserm U1019, Lille, France.

ABSTRACT
A fundamental step in High-throughput sequencing (HTS) data analysis is the mapping of reads onto reference sequences. Choosing a suitable mapper is a subtle task because of the difficulty of evaluating mapping algorithms. We present a benchmark procedure to compare mappers using both real and simulated datasets and considering computational resource and time requirements, robustness of mapping, ability to report positions for reads in repetitive regions, and ability to retrieve true genetic variation positions. To measure robustness, a new definition for a correctly mapped read was introduced. We developed CuReSim, a read simulator, and CuReSimEval, a tool to evaluate the mapping quality of the simulated reads. The benchmark procedure was applied to evaluate mappers in the context of whole genome sequencing of small genomes with Ion Torrent data. These results were used to develop a pipeline to quickly and automatically characterize pathogens during an episode of infection.

Publication:

Caboche S, Audebert C, Lemoine Y, Hot D. Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. BMC Genomics. 2014 Apr 5;15:264.

Contact: segolene.caboche@pasteur-lille.fr

HP03: ***Patterns of positive selection in seven ant genomes.***

Julien Roux[1], Eyal Privman[2], Sébastien Moretti[1], Josephine Daub[3], Marc Robinson-Rechavi[1] and Laurent Keller[1].
[1]University of Lausanne, Switzerland. [2]University of Haifa, Israel.[3]University of Bern, Switzerland.

ABSTRACT
The evolution of ants is marked by remarkable adaptations that allowed the development of very complex social systems. To identify how ant-specific adaptations are associated with patterns of molecular evolution, we searched for signs of positive selection on amino-acid changes in proteins. We identified 24 functional categories of genes which were enriched for positively selected genes in the ant lineage. We also reanalyzed genome-wide datasets in bees and flies with the same methodology, to check whether positive selection was specific to ants or also present in other insects. Notably, genes implicated in immunity were enriched for positively selected genes in the three lineages, ruling out the hypothesis that the evolution of hygienic behaviors in social insects caused a major relaxation of selective pressure on immune genes. Our scan also indicated that genes implicated in neurogenesis and olfaction started to undergo increased positive selection before the evolution of sociality in Hymenoptera. Finally, the comparison between these three lineages allowed us to pinpoint molecular evolution patterns that were specific to the ant lineage. In particular, there was ant-specific recurrent positive selection on genes with mitochondrial functions, suggesting that mitochondrial activity was improved during the evolution of this lineage. This might have been an important step toward the evolution of extreme lifespan that is a hallmark of ants.

Publication:
Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. Patterns of positive selection in seven ant genomes. Mol Biol Evol. 2014 Jul;31(7):1661-85.

Contact: julien.roux@unil.ch

HP04: ***Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes.***

Kęstutis Timinskas[1], Monika Balvočiūtė[2], Albertas Timinskas[1] and Česlovas Venclovas[1].
[1]Institute of Biotechnology, Vilnius University, Lithuania. [2]University of Otago, New Zealand.

ABSTRACT
Bacteria, unlike archaea and eukaryotes, use distinct C-family DNA polymerases for genome replication. Unfortunately, except for a few species, bacterial genome replication is poorly characterized. It is not known whether all bacteria use C-family DNA polymerases for DNA replication, how many distinct C-family groups are there, and how many different replication systems they form. In order to address these questions, we performed extensive computational analysis of C-family polymerases in nearly 2000 complete bacterial genomes. We found that all the genomes without exception encode at least one C-family polymerase implying the universal use of this polymerase family for bacterial DNA replication. Our analysis revealed four distinct groups of C-family polymerases. Based on their properties and distribution in genomes we discovered a novel, so far experimentally uncharacterized, replication system in Clostridia. Computational results also indicated that one of the C-family groups might be responsible for shaping genomic G+C content.

Publication:
Timinskas K, Balvočiūtė M, Timinskas A, Venclovas Č. Comprehensive analysis of DNA polymerase III α subunits and their homologs in bacterial genomes. Nucleic Acids Res. 2014 Feb;42(3):1393-413.

Contact: venclovas@ibt.lt

HP05: ***High-dimensional Bayesian parameter estimation: Case study for a model of JAK2/STAT5 signaling.***

Sabine Hug[1], Andreas Raue[2], Jan Hasenauer[1], Julie Bachmann[3], Ursula Klingmüller[3], Jens Timmer[4] and Fabian J. Theis[1].
[1]Institute of Computational Biology, Helmholtz Zentrum München, Germany. [2]Merrimack Pharmaceuticals, Inc., Cambridge, MA, USA. [3]Systems Biology of Signal Transduction, DKFZ-ZMBH Alliance, German Cancer Research Center, Heidelberg, Germany. [4]Institute for Physics, University of Freiburg, Germany.

ABSTRACT
Mechanistic dynamical models are nowadays commonly used for the analysis for complex datasets. Dynamical models depend however on many unknown parameters which have to be inferred from experimental data. The statistical inference in a high-dimensional parameter space is however conceptually and computationally challenging. In this paper we provide a proof of principle that the rigorous statistical analysis is also feasible in these demanding situations.
To ensure rigorous assessment of model and prediction uncertainties we take advantage of both a profile posterior approach and Markov chain Monte Carlo sampling.
We analyzed a dynamical model of the JAK2/STAT5 signaling pathway containing more than hundred parameters. The profile posterior reveals that the corresponding posterior distribution is bimodal. To nevertheless guarantee efficient mixing we applied a multi-chain sampling approach. The Bayesian parameter

estimation enables the assessment of prediction uncertainties and the design of additional experiments enhancing the explanatory power of the model.

Publication:

Hug S, Raue A, Hasenauer J, Bachmann J, Klingmüller U, Timmer J, Theis FJ. High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. Math Biosci. 2013 Dec;246(2):293-304.

Contact: sabine.hug@helmholtz-muenchen.de

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

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

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

HP09: *Novel Developments in computational clinical breath analysis and biomarker detection.*

Anne-Christin Hauschild[1], Jörg Ingo Baumbach[2] and Jan Baumbach[3].

[1]Max Planck Institute for Informatics, Saarbrücken, Germany. [2]Faculty Applied Chemistry, Reutlingen University, Germany. [3]University of Southern Denmark, Denmark.

ABSTRACT

The volatolom is the sum of volatile organic compounds that are emitted by all living cells and tissues. We seek to non-invasively "sniff" biomarker molecules that are predictive for the biomedical fate of individual patients. This promises great hope to move the therapeutic windows to earlier stages of disease progression. While portable devices for breathomics measurement exist, we face the traditional biomarker research barrier: a lack of robustness hinders translation to the world outside laboratories. To move from biomarker discovery to validation, from separability to predictability, we have developed several bioinformatics methods for computational breath analysis, which have the potential to redefine non-invasive biomedical decision making by rapid and cheap matching of decisive medical patterns in exhaled air. We aim to provide a supplementary diagnostic tool complementing classic urine, blood and tissue samples. The presentation will review the state of the art, highlight existing challenges and introduce new data mining methods for identifying breathomics biomarkers.

Publications:

Hauschild AC, Kopczynski D, D'Addario M, Baumbach JI, Rahmann S, Baumbach J. Peak detection method evaluation for ion mobility spectrometry by using machine learning approaches. Metabolites. 2013 Apr 16;3(2):277-93.

Eckel SP, Baumbach J, Hauschild AC. On the importance of statistics in breath analysis--hope or curse? J Breath Res. 2014 Mar;8(1):012001.

Maurer F, Hauschild AC, Eisinger K, Baumbach J, Mayor A and Baumbach JI. MIMA - a software for analyte identification in MCC/IMS chromatograms by mapping accompanying GC/MS measurements. Int. J. Ion Mobil. Spec. 2014 Apr:17:95–101.

Smolinska A, Hauschild AC, Fijten RR, Dallinga JW, Baumbach J, van Schooten FJ. Current breathomics-a review on data pre-processing techniques and machine learning in metabolomics breath analysis. J Breath Res. 2014 Jun;8(2):027105.

Contact: jan.baumbach@imada.sdu.dk

HP10: *Text mining technologies for database curation.*

Fabio Rinaldi[1], Simon Clematide[1], Simon Hafner[1], Gerold Schneider[1], Gintare Grigonyte[2], Martin Romacker[3] and Therese Vachon[4].

[1]University of Zurich, Switzerland. [2]University of Stockholm, Sweden. [3]F. Hoffmann-LaRoche, Switzerland. [4]Novartis, Switzerland.

ABSTRACT

Although human curation for life science databases offers the best guarantee of high quality annotations, it suffers from severe bottlenecks which have long been recognized in the curation community. The most pressing problem is that of efficiency of the process: it is impossible for human curators to keep up with the growing pace of publication. Text mining technologies, coupled with advanced user interfaces, offer the potential to partially alleviate this bottleneck. We survey the results of several recent competitive evaluations of text mining technologies, discuss how text mining systems can be integrated in a curation workflow, and illustrate our approach to assisted curation, which has been tested in collaboration with major databases.

Publication :

Rinaldi F, Clematide S, Hafner S, Schneider G, Grigonyte G, Romacker M, Vachon T. Using the OntoGene pipeline for the triage task of BioCreative 2012. Database (Oxford). 2013 Feb 9;2013:bas053.

Contact: fabio@ontogene.org