

MEETING REPORT

Bioinformatics: living on the edge

Evangelos Bellos¹, Lachlan JM Coin^{2,3*} and Myrsini Kaforou⁴

Abstract

A report on the 11th European Conference on Computational Biology (ECCB), Basel, Switzerland, September 9-12, 2012.

Keywords bioinformatics, genomics, proteomics, transcriptomics, interactions, networks

Meeting report

Can music be decomposed to single notes on a piece of paper? Can life be fully understood by the function of genes in isolation? These are some of the questions that Denis Noble (University of Oxford, UK) posed in one of the more philosophical keynote addresses of this year's European Conference on Computational Biology (ECCB). Noble, a famed critic of genetic reductionism, in concert with renowned guitarist Christoph Denoth, set out to explore how synergy begets complexity both in art and in science. And in doing so, he captured a central theme of the conference: holism.

From gene prediction to comparative genomics to protein structure prediction, the early days of post-genomic bioinformatics research were arguably characterized by a reductionist approach to understanding life. Systems biology, on the other hand, started on the opposite end of the spectrum, attempting to directly model complex behavior arising from biomolecular interplay, largely abstracted from the individual components. In this post-ENCODE era, in which tremendous progress has been made in understanding each biomolecule individually, the focus now seems to have shifted toward a middle ground: leveraging the treasure trove of information about the nodes of the network to elucidate the intricate web of interactions represented by the network edges.

Given the sustained development of next-generation sequencing technologies and their essential role in the ascendancy of genomics, it is perhaps surprising that talks on this subject were conspicuously scarce at ECCB. However, this is not to say that genomics was not pervasive throughout the conference; it was, from keynotes to posters, just indirectly so. Sequencing technologies appear to be maturing into stable tools that are ubiquitously used to obtain deep knowledge of biomolecules themselves, as well as intermolecular relationships. As a silent partner, genomics is providing both the starting point and the means for investigating networks, but without stealing the spotlight.

Genetic interactions: within genomes

The systematic identification and characterization of genome-wide genetic interactions is one area that has already been transformed by the availability of sequence data for multiple species. Colm Ryan (University of California, San Francisco, USA) compared experimentally derived epistasis maps of the fission yeast *Schizosaccharomyces pombe* and the budding yeast *Saccharomyces cerevisiae* to show that the overall level of genetic cross-talk between different biological processes tends to be more conserved than the underlying interactions. He further demonstrated how epistasis maps can be used to annotate uncharacterized genes, drawing on the observation that genes with similar genetic interaction profiles are often involved in the same functional pathway. Similar approaches may be used as a model for genetic interaction analysis in more complex organisms, in an effort to understand not only biological functions, but also the mechanisms that underlie disease phenotypes. Philipp Bucher (École Polytechnique Fédérale de Lausanne, Switzerland) exploited lineage-specific whole-genome duplication events in teleost fish (including zebrafish and pufferfish lineages) to investigate the post-duplication fate of ultraconserved non-coding elements. These elements tend to be organized in large clusters around developmental genes, and were found mostly to be retained together in only one of the two genome copies. This winner-takes-all scenario suggests that ultraconserved non-coding elements operate as part of a dense cooperativity network, which could be responsible for their high levels of conservation.

*Correspondence: l.coin@imperial.ac.uk

²Department of Genomics of Common Disease, School of Public Health, Imperial College London, The Hammersmith Hospital, Du Cane road, London, W12 0NN, UK

Full list of author information is available at the end of the article

Genetic interactions: between genomes

In a *tour de force* keynote presentation focusing on the social organization of fire ants (*Solenopsis invicta*), Laurent Keller (University of Lausanne, Switzerland) demonstrated that an ant's behavior is determined not only by its own genome but also by other genomes in the population. By way of example, Keller described what might be called a selfish allele in a 13.9 Mbp non-recombining region, encoding the odorant protein Gp-9, that regulates its own frequency by mediating altruistic and aggressive instincts. Michal Linial (The Hebrew University of Jerusalem, Israel) explored interactions between virus and host genomes from an evolutionary perspective. Linial demonstrated how both amino acid and codon usage of viral proteomes have adapted to maximize translational efficiency in specific host environments. Linial also showed that host-acquired genes in virus genomes tend to have fewer protein domains and shorter linker regions than their host counterparts. Understanding these evolutionary adaptations may prove essential to the effective management of viral infections.

Protein-RNA interactions

The recently published results of the ENCODE project established that at least three quarters of the human genome is capable of being transcribed. One approach to understanding the functional role of this pervasive transcription may be the identification of new RNA-binding proteins, as presented by Cristoph Dieterich (Max Delbrück Center for Molecular Medicine, Germany). By purifying proteins that have been cross-linked to polyadenylated mRNA, followed by quantitative mass spectrometry, Dieterich identified 797 mRNA-bound proteins, one third of which were previously unknown and likely participate in post-transcriptional gene-regulation networks. Moreover, Dieterich used PAR-CLIP to sequence the mRNA-binding sites of these proteins. Mihaela Zavolan (Biozentrum, Switzerland) also reported a novel application of PAR-CLIP, in which the Argonaute component of the RISC complex was used as bait to discover non-canonical microRNA (miRNA)-target interactions. Intriguingly, Zavolan showed how a single miRNA molecule (miR-294) can target components of the chromatin-remodeling complex and prevent differentiation in embryonic stem cells, thus concluding that a handful of miRNAs may be sufficient to determine cell fate.

Protein-protein interactions

Although protein-protein interaction (PPI) networks have been extremely valuable as a starting point for modeling cellular processes, it is well known that PPI databases have a high false-positive rate, while omitting many genuine interactions. The gold standard for the

field, as described by Barry Honig (Columbia University, USA), is a three-dimensional structure of a complex that incorporates the interacting proteins, but this level of confirmation is available for less than 0.5% of all recorded PPIs. Honig presented a computationally efficient method for superimposing homologous proteins onto known structural interactions in order to generate interaction models, which are subsequently combined with co-expression and functional similarity information. This approach makes clever use of homology to known structures; however, as Chris Sander (Memorial Sloan-Kettering Cancer Center, USA) noted, only half of the well-characterized protein families in the Pfam database have a known three-dimensional structure for any of their members. To address this limitation, Sander identified co-evolving residues from multiple sequence alignments in order to predict residue-residue proximity in folded protein structures. Philip Kim (University of Toronto, Canada) illustrated the tissue-specific nature of PPI networks by monitoring how alternative splicing events impact upon them. Kim presented evidence that as much as a third of all tissue-specific alternatively spliced exons lead to differential PPI through both the creation of new and the destruction of existing interactions.

Protein-drug interactions

Chemogenomics is an emerging research area that is focused on identifying novel drug-protein interactions by screening libraries of small molecules against the proteome, through the computational analysis of high-throughput data. Yoshihiro Yamanishi (Kyushu University, Japan) presented a method, based on sparsity-induced binary classifiers, that incorporates data on the chemical structures of drugs, genomic information about target proteins and all known protein-ligand interactions across different protein families into a predictive model. The model can predict underlying interactions between drug chemical substructures and protein functional sites, which are involved in drug-target interaction networks. With a view to understanding and minimizing drug side effects in patients, Sayaka Mizutani (Kyoto University, Japan) presented an approach for integrating drug side effect data with drug-protein binding information, using sparse canonical correlation. This technique was used to derive highly correlated sets of drug-targeted proteins and side effects, along with the drugs that drive their correlation. Such methods can play an important role in predicting potential side-effects and their mediating pathways even at the drug design phase.

Conclusions

The opening keynote lecture of ECCB was delivered by Nobel laureate Aaron Ciechanover (Technion - Israel Institute of Technology, Israel) who described his

magnum opus: unraveling the previously unknown cellular mechanism of protein degradation. Ciechanover outlined many aspects of ubiquitin-mediated protein degradation, including the role of E3 ubiquitin ligases in ensuring high target specificity. He further highlighted the link between dysregulation of the ubiquitin-proteasome pathway and disease. Various levels of this pathway have been identified as drug targets to treat or suspend the progression of diseases such as myeloma, leukemia and viral encephalitis. While this talk was inspiring, the fact that these discoveries were made without the assistance of high-throughput technologies or bioinformatics was sobering, and raises the question of whether computational analyses of high-throughput data can lead to the elucidation of still unknown, fundamental cellular mechanisms. In other words, when will bioinformatics have its Nobel Prize moment? The talks presented at this conference suggest that by transforming information on complex biomolecular interactions, obtained at high cellular and temporal resolution, into sequence data, bioinformaticians may be well on the way to achieving such a feat.

Abbreviations

ECCB, European Conference on Computational Biology; ENCODE, Encyclopedia of DNA Elements; Mbp, megabase pair; miRNA, microRNA; PAR-CLIP, photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation; PPI, protein-protein interaction; RISC, RNA-induced silencing complex.

Competing Interests

The authors declare that they have no competing interests.

Author details

¹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Hospital, Norfolk Place, London, W2 1PG, UK. ²Department of Genomics of Common Disease, School of Public Health, Imperial College London, The Hammersmith Hospital, Du Cane road, London, W12 0NN, UK. ³BGI-Shenzhen, Beishan Road, Shenzhen, 518083, China. ⁴Section of Paediatrics and Wellcome Trust Centre for Clinical Tropical Medicine, Division of Infectious Diseases, Department of Medicine, Imperial College London, St Mary's Hospital, Norfolk Place, London, W2 1PG, UK.

Published: 26 October 2012

doi:10.1186/gb-2012-13-10-321

Cite this article as: Bellos E, *et al.*: Bioinformatics: living on the edge. *Genome Biology* 2012, **13**:321.