EMBnet.journal
Bioinformatics in Action

EMBnet Conference 2020

Bioinformatics Approaches to Precision Research

23-24 September 2020

Zoom Platform

26
2020
Supplement A

EMBnet.journal
Bioinformatics in Action

Contents

Editorial

The articles published by the EMBnet.journal in this Supplement consist of an introductory article by the EMBnet Executive Board and abstracts of research works presented at the EMBnet Conference 2020 on "Bioinformatics Approaches to Precision Research". The Conference was held on 23th-24th September 2020 using the web video conferencing platform Zoom and in conjunction with the EMBnet Annual General Meeting (AGM). Since more than 30 years ago, the EMBnet AGM is the occasion that all EMBnet members wait with great delight because of the pleasure of staying together physically in an enthusiastic and friendly atmosphere that is the driving force of our community. In 2020, the restrictions to move imposed by the COVID-19 pandemic prevented the organisation of the meeting in person. Nevertheless, as Executive Board, we thought that organising a conference where our members could share and discuss their research activities and scientific achievements was the best way to give them the possibility to enjoy still the pleasure of meeting each other and stimulating the cross-fertilisation of ideas and collaborations. The decision to collect in this Supplement of the EMBnet Journal the contributions presented at the Conference was taken to offer to researchers inside and outside EMBnet the opportunity to get an almost complete overview of the multifaceted research activities carried out by our community for potential future collaborations. Presentations spanned many cutting-edge topics of modern research, from new applications of Bioinformatics and Artificial Intelligence for the optimisation and automation of processes for sustainable farming activities to advanced approaches for the development of precision therapeutic approaches, to investigate mechanisms of epigenetic inheritance, to investigate how vegetables in the human diet can aid to keep health status and prevent many diseases.

Enjoy reading, and do not hesitate to contact us for any further information on the research works presented or to get the contacts of the related research teams.

Domenica D'Elia
Chair of the EMBnet Executive Board
domenica.delia@ba.itb.cnr.it

http://dx.doi.org/10.14806/ej.26.A.1008

EMBnet.journal
Bioinformatics in Action

D'Elia et al. (2021) EMBnet.journal 26(Suppl A), e1006
http://dx.doi.org/10.14806/ej.26.A.1006

Reports

The EMBnet International Conference 2020: Bioinformatics Approaches to Precision Research

Domenica D'Elia[1]✉, Cesar Bonavides-Martinez[2], Erik Bongcam-Rudloff[3], Emiliano Barreto-Hernandez[4], Aspasia Efthimiadou[5], Eleni Papakonstantinou[6], Dimitrios Vlachakis[6,7,8], Lubos Klucar[9]

[1]Institute for Biomedical Technologies, National Research Council, Bari, Italy

[2]Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

[3]Swedish University of Agricultural Sciences, Uppsala, Sweden

[4]Bioinformatics Center, Biotechnology Institute, Universidad Nacional de Colombia, Bogota, Colombia

[5]Hellenic Agricultural Organization-Demeter, Institute of Soil and Water Resources, Department of Soil Science of Athens, Lycovrisi, Greece

[6]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[7]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, Athens, Greece

[8]Division of Endocrinology and Metabolism, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[9]Institute of Molecular Biology, SAS, Bratislava, Slovakia

Competing interests: DD none; CBM none; EBR none; EBH none; AE none; EP none; DV none; LK none

Abstract

In this paper, we first summarise the 2020 International Conference of EMBnet - The Global Bioinformatics Network on "Bioinformatics Approaches to Precision Research", held on 23th-24th September 2020, and then briefly introduce the main topics and contributions presented at the Conference and published in this supplement issue. The Conference's main aim was to share the knowledge and scientific achievements of EMBnet members to prompt the cross-fertilisation of ideas and collaborations. The title reflects the type of contributions presented (20 in total), covering a wide range of cutting-edge research topics for which advanced bioinformatics methods and precision approaches are essential, and training in Bioinformatics. As for many other organisations, 2020 was the first year that the EMBnet Annual General Meeting was not in presence but ran on a virtual meeting platform because of the pandemic's travel restrictions to moves. If this situation represented a limitation to our social activities, it allowed us to offer the possibility to attend the Conference to a larger number of researchers (i.e., 190 from 38 different countries). Among all registered people, 148 were following the conference presentations on both days. Under the authors' permission, presentations were recorded and are available at the EMBnet Conference 2020 playlist on YouTube, whereas the programme is available at the EMBnet Conference 2020 Google Drive folder.

EMBnet Conferences in a nutshell

EMBnet – The Global Bioinformatics Network[1] is the first and oldest network established in Europe by researchers in Bioinformatics. It was constituted in 1988 as the "European Molecular Biology Network" by a handful of European countries to distribute and provide European research institutes with access to sequence data and bioinformatics tools, to connect molecular biologists and bioinformaticians within Europe and to provide education and training in Bioinformatics.

In 2008 EMBnet celebrated its 20th anniversary with an international conference on "Leading Applications and Technologies in Bioinformatics". This Conference was the first meeting organised by the network that was open to the international scientific community outside EMBnet (D'Elia *et al.*, 2009).

Since then, each year, the Annual General Meeting (AGM) of the Network has always been accompanied by an open conference, a workshop or a school on cutting-edge challenges in the bioinformatics field and usually organised in collaboration with other network organisations or societies. With the Iberoamerican

[1]https://www.embnet.org/wp/about/

Society for Bioinformatics (SoIBio)[2], EMBnet organised a big conference in Cancun, Mexico, in 2009 (Lopez-Bojorquez *et al.*, 2010) and in 2018 in Viña del Mar (Chile), celebrating the 30th anniversary of EMBnet, also with the collaboration of the International Society for Computational Biology (ISCB)[3]. Other partners were the project EMBRACE (Gisel and Bongcam-Rudloff, 2010), STATEGRA[4], AllBio[5], the COST Action SeqAhed (Attwood *et al.,* 2011), the COST Action CHARME (CA15110) and NETTAB[6], the Network Tools and Applications in Biology workshops series in Bioinformatics. In 2019 the EMBnet AGM was organised in the Republic of North Macedonia in conjunction with the ICT Innovation Conference 2019[7]. On this occasion, EMBnet organised as a satellite event of the main Conference a workshop on "Big Data Analysis and Reproducibility: new challenges and needs of the post-genomic era" (Efthimiadou *et al.*, 2019).

Plans for the EMBnet AGM 2020 were in preparation when the SARS-COV-2 invaded Europe and immediately after the American and African continents as a tsunami. There were no clues on how the pandemics could have been evolving in the upcoming months. The decision was taken in July 2020 to involve as much as possible the EMBnet members in an open and virtual scientific event where each member could present new ideas and projects, research activities and achievements. The main aim was to offer the members the possibility to share their work inside and outside the EMBnet community to prompt the cross-fertilisation of ideas and collaborations. The Conference attracted many researchers from all over the world, most of them being young researchers and students who actively participated in the scientific discussion of the research presented. The enthusiasm was contagious, and the atmosphere the one of a community that share common interest and passion for its work with optimism and willingness to do although the COVID-19 pandemic. As organisers of the EMBnet Conference 2020, we thank the speakers and attendants for their contribution to its success.

EMBnet Conference 2020: numbers

The Conference was an entirely digital event, and for this reason, we did not need to ask for a registration fee. Although the short anticipation of the conference programme, the number of registration was 190. Among all registered people, 148 followed the conference presentations on both days from 38 different countries inside and outside of Europe, such as Algeria, Brazil, Canada, Chile, China, Colombia, Costa Rica, Ecuador, India, Malaysia, Mexico, Nepal, Nigeria, Pakistan, Russia, Saudi Arabia, Serbia, Sri Lanka, Taiwan, United Kingdom, United States of America.

The Conference daily sessions were organised to start at 3 pm and end at 8 pm (CEST) and using the Zoom Platform[8]. The time slot was fixed taking into consideration the time zone on diverse continents. Nevertheless, we recorded all presentations and made them available in the form of a playlist on YouTube[9] to allow access at any time for people who could not follow live the Conference. These video presentations are included also in the digital version of this EMBnet. journal issue.

The programme and presentations files were deposited in a Google Drive folder dedicated to the Conference and open to participants under presenters' permission. A Slack Channel was also created and made available to registered people for timely communications with the programme committee members and speakers before, during and after the Conference.

The authors submitted contributions to the Programme Committee evaluation as short abstracts. Selected Works, in total 20, are published in this EMBnet Journal Supplement as Proceedings of the Conference.

The EMBnet Conference 2020 scientific programme

The Conference was open by Domenica D'Elia, the Chair of the EMBnet Executive Board, who welcomed participants and briefly presented EMBnet, its activities, mission and products[10].

The first session was introduced and chaired by Professor Erik Bongcam-Rudloff, Chair of the EMBnet Project Committee on Education & Training in Bioinformatics. Presentations of the first conference session gave a multifaceted overview of how bioinformatics precision approaches can be applied to various domains for sustainable development, in agriculture and animal farm, and to human health for more effective therapeutic strategies.

The session was opened by a presentation of Professor Dimitrios Vlachakis, member of the EMBnet Executive Board, on a holistic approach of systems biology and bioinformatic analysis on breast milk exosome. He demonstrated how breast milk molecular fingerprinting could pave the way to shed light on the underlying genetics and epigenetics that a mother can transmit to her children.

Dr Tomas Klingström presented the development of the Gigacow infrastructure[11] for precision dairy farming and the research opportunities presented to researchers at SLU and international collaborators.

[2]http://www.soibio.org/
[3]https://www.iscb.org/iscb-latinamerica2018
[4]http://www.stategra.eu/
[5]http://www.allbioinformatics.eu/doku.php
[6]http://www.igst.it/nettab/2016/
[7]http://www.ictinnovations.org/11th-international-conference-ict-innovations-2019
[8]https://zoom.us/
[9]https://www.youtube.com/playlist?list=PLVP2771lOozCHAOgVXBdqPW3oclEmzNIA
[10]https://drive.google.com/drive/folders/1VX2Lf5ho5gYGIHhDKlR_O_F20l-mfEDo
[11]https://www.slu.se/institutioner/husdjursgenetik/forskning/gigacow/

D'Elia et al. (2021) EMBnet.journal 26(Suppl A), e1006
http://dx.doi.org/10.14806/ej.26.A.1006

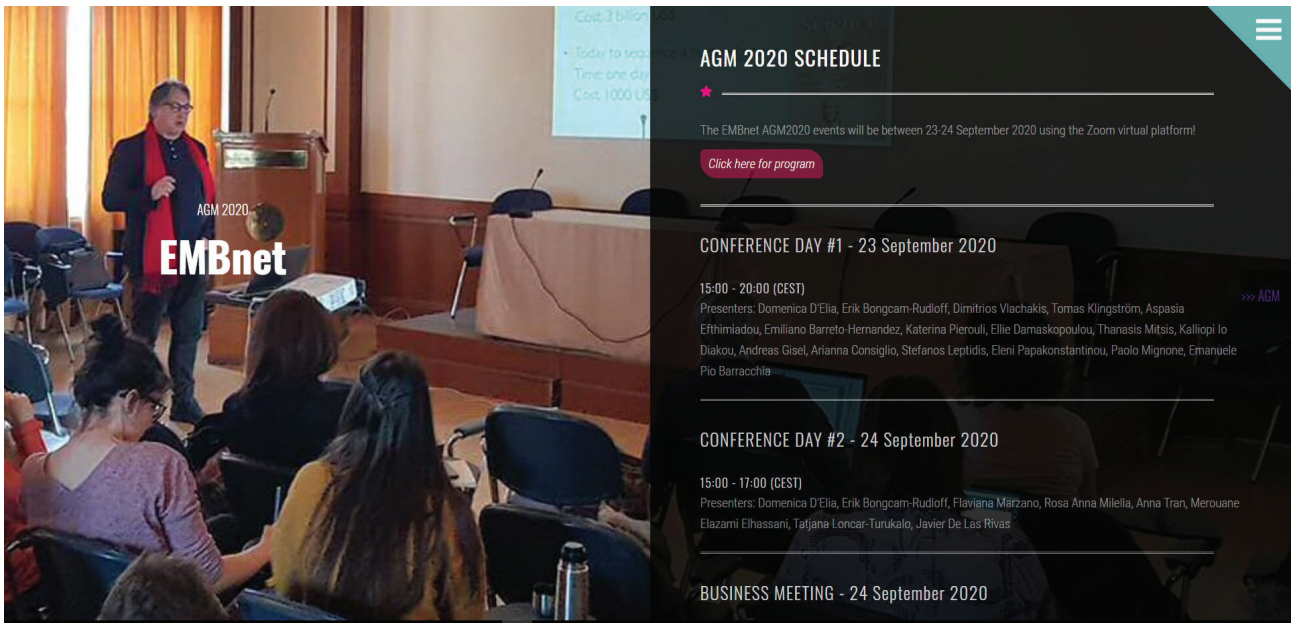

Figure 1. The EMBnet Conference 2020 web site.

Dr Aspasia Efthimiadou, Chair of the Publicity & Public Relations Project Committee of EMBnet, exposed a case study on cotton cultivation to demonstrate how Neural Network-based Decision Support Systems, combined with data analysis, can be considered the future of sustainable agriculture.

Professor Emiliano Barreto-Hernandez, Treasurer and member of the EMBnet Executive Board, presented SGIG, the Genomic Information Management System, developed to integrate clinical, epidemiological, laboratory, and genomic data for precision epidemiology of multi-drug resistance bacteria in Colombia.

Finally, Dr Stefanos Leptidis, presented a methodology for identifying cell types using their single-cell micro RNA (miRNA) profile coupled to their predicted targets from various miRNA target prediction algorithms. This method is helpful to elucidate the intricate cellular interactions and regulatory pathways involved in organ-specific pathologies for the development of precision therapeutic approaches.

The second session on day one was chaired by Prof. Dimitrios Vlachakis. Themes treated spanned from genome regulation by long non-coding RNAs (ncRNAs) presented by Katerina Pierouli, to how the combination of Genome-Wide Association Studies (GWAS), Single Nucleotide Polymorphisms (SNPs), and methylation profiles analysis will provide the possibility to investigate mechanisms of epigenetic inheritance in children following exposure to abuse (presenter Ellie Damaskopoulou). On another side, Thanasis Mitsis illustrated how GWAS could be used to get insights into the complex interplay of nuclear receptor transcriptional networks to elucidate their contribution to the maintenance of cell homeostasis. Andreas Gisel, a member of the Education & Training Project Committee of EMBnet, talked about the 3'Tag-Seq technology for transcriptomics studies, underlining how this technology gives not only precise information about gene expression but also for alternative annotation in the 3'UTR of genes. Arianna Consiglio, illustrated a study on the expression profile of non-coding RNAs in coronaviruses that demonstrates a possible action of RNA interfering on the human immune system response. Eleni Papakonstantinou, exposed a computational drug design strategy developed to discover the most potent molecules with an inhibitory effect on the helicase function and the viral replication cycle of the Yellow Fever Virus.

The third session, chaired by Prof. Emiliano Barreto-Hernandez, included two presentations illustrating applications of machine learning techniques to Big Data analysis for i) gene regulatory network reconstruction, presented by Paolo Mignone and using transfer learning techniques; and for ii) the prediction of relationships between ncRNAs and human diseases, presented by Emanuele Pio Barracchia, exploiting multi-type hierarchical clustering techniques.

The second day of the Conference was chaired by Lubos Klucar, Secretary of the EMBnet Executive Board, and included scientific presentations and a space dedicated to education and training.

The first presentation was by Flaviana Marzano, exposing a combined approach (in silico and lab experimental validation) demonstrating that plant micro RNAs can interfere with human lncRNAs to control cancer genes expression. On another side, Rosa Anna Milella exposed the results of a nutrigenomics study underlining the effects of grape intake on human gene expression, cell signalling pathways, and many ncRNAs. Anna Tran talked about in silico characterisation of the gene repertoires of immunoglobulins (IGs) and T cell receptors (TRs) of various inbred laboratory strains of M. musculus to design and develop or adapt high-performance software tools and a methodology to carry

Reports

out the annotation of the loci IG and TR of the mouse strains with a "Gold standard" quality (equivalent to the manual annotation). Merouane Elazami Elhassani presented an approach based on deep neural network-based models trained in a supervised manner, which automatically learns features from annotated IG and TR genes to predict the L-PART1 exon (the first exon of IG and TR variable V-GENE).

Education & Training in Bioinformatics session

Invited speakers were Tatjana Loncar-Turukalo, Coordinator of the Short-Term Scientific Mission (STSM) programme of the COST Action CA18131 - ML4Microbiome[12] (Statistical and machine learning techniques in human microbiome studies); Dr Javier De Las Rivas, President of SoIBio, the Iberoamerican Society for Bioinformatics; and Prof. David Coornaert of the University College "Haute Ecole en Hainaut - Campus Technique", in Mons (BE).

Dr Tatjana Loncar-Turukalo exposed the aims of the COST Action ML4Microbiome[13] with a focus on training and education activities and possible opportunities for participation and collaboration. Dr Javier De Las Rivas presented "Advancement of Bioinformatics and Computational Biology in Latin America: SoIBio and other scientific networks" (De Las Rivas *et al.*, 2019), making an overview of the history and current state of research in Bioinformatics in Latin America and underlining the role of SoIBio as a leading forum to join efforts of many scientists from LA to accelerate the data-driven biology research in LA and also a sustainable capacity-building programme based on education and training in Bioinformatics in collaboration with the CABANA project[14]. Prof. David Coornaert was invited along with two students of the University College Haute Ecole en Hainaut in Mons to present the programme of collaboration with EMBnet to train bachelor students inside the ERASMUS programme. His presentation was concluded by a nice presentation of the two students, Cyril Radermecker and Ahmed Kanfoud, about their ERASMUS stage in 2020 across the EMBnet Node in Greece under the supervision of Prof. Dimitrios Vlachakis.

The Conference ended at 5 pm. The EMBnet Executive Board thanked the speakers and present people for their contributions and participation with the promise to keep in touch for upcoming events or potential collaborations.

After the Conference, EMBnet members continued with the programme of the Annual General Meeting, including the Executive Board and Project Committees

[12]https://www.cost.eu/actions/CA18131/#tabs|Name:overview
[13]https://www.ml4microbiome.eu/
[14]https://www.cabana.online/

Figure 2. The survivors toasting at the success of the EMBnet Conference and of the AGM 2020.

activity reports, the presentation of the latest achievements and plans for the EMBnet Journal and the discussion on priorities and plans for 2020/2021. A virtual toast among some of the Operational Board's members (Fig. 2), survived at the intense last day activities, the success of the Conference and the pleasure to stay all together again, although geographically separate, with the wish that in 2021, EMBnet members can meet again in person to exchange true hugs that have been missed so much to all us.

References

1. Attwood T, Bongcam-Rudloff E, Gisel A (2011) SEQAHEAD - COST Action BM1006: Next Generation Sequencing Data Analysis Network. EMBnet.journal, **17**.A:7-9. http://dx.doi.org/10.14806/ej.17.1.218
2. D'Elia D, Gisel A, Eriksson NE, Kossida S, Mattila K *et al.* (2009) The 20th anniversary of EMBnet: 20 years of bioinformatics for the Life Sciences community. BMC Bioinformatics **10**, Suppl 6:S1 doi: http://dx.doi.org/10.1186/1471-2105-10-S6-S1
3. De Las Rivas J, Bonavides-Martínez C, Campos-Laborie FJ (2019) Bioinformatics in Latin America and SoIBio impact, a tale of spin-off and expansion around genomes and protein structures. Briefings in Bioinformatics, **20**(2):390–397. http://dx.doi.org/10.1093/bib/bbx064
4. Efthimiadou A, Papakonstantinou E, Vlachakis D, D'Elia D, Bongcam-Rudloff E (2019) Big Data analysis and reproducibility: new challenges and needs of the post-genomics era. In Springer Communications in Computer and Information Science book series. ICT Innovations 2019 Web proceedings: 253-257. ISSN 1857-7288
5. Gisel A, Bongcam-Rudloff E (2010) EMBRACE workshop "NEXT GENERATION SEQUENCING II". EMBnet.journal, **16**.A:5-7. http://dx.doi.org/10.14806/ej.16.1.176
6. Lopez-Bojorquez Lucia *et al.* (2010) Report on the International Conference & Meetings EMBnet-RIBio 2009. EMBnet.journal, **16**.A:3-5. http://dx.doi.org/10.14806/ej.16.A.155

EMBnet Conference 2020

Bioinformatics Approaches to Precision Research

23-24 September 2020

Zoom Platform - Time 3 pm to 8 pm (CEST)

www.embnet.org www.embnet.eu EMBnet in a nutshell

Programme

1st Day: 23 September 2020

15:00-15:20 **Welcome & Programme Presentation**

Domenica D'Elia & Erik Bongcam-Rudloff

Chair *Erik Bongcam-Rudloff*

15:20-15:40 Exosomes in breast milk; a beneficial genetic trojan horse from mother to child

Dimitrios Vlachakis, Aspasia Efthimiadou, Flora Bacopoulou, Elias Eliopoulos, George P. Chrousos

15:40-16:00 Precision dairy farming – A Phenomenal opportunity

Tomas Klingström

16:00-16:20 Climate mitigation. Is it Possible via DSS for agriculture? A case study: Reducing Environmental Footprint of cotton cultivation

Dimitrios Leonidakis, Nikolaos Katsenios, Panagiotis Sparangis, Christoforos Nikitas Kasimatis, Dimitrios Vlachakis, **Aspasia Efthimiadou**

16:20-16:40 Precision Epidemiology of Multi-drug resistance bacteria: bioinformatics tools
J.Donato, L. Lugo, H. Perez, H. Ballen, D. Talero, S. Prada, F.Brion, V. Rincon, L. Falquet, M.T. Reguero, **E. Barreto-Hernandez**

16:40-17:00 Single-cell mapping of microRNA expression during cardiac development
Stefanos Leptidis, E. Papakonstantinou, K. Pierouli, A. Mitsis, S. Chlamydas, A. Efthimiadou, G.P. Chrousos, E. Eliopoulos, E. Hansson, D. Vlachakis

17:00-17:10 Break

Chair *Dimitrios Vlachakis*

17:10-17:25 Genome regulation by long non-coding RNAs
Katerina Pierouli, George N. Goulielmos, Elias Eliopoulos, Dimitrios Vlachakis

17:25-17:40 Mechanisms of epigenetic inheritance in children following exposure to abuse
Ellie Damaskopoulou, George P. Chrousos, Elias Eliopoulos, Dimitrios Vlachakis

17:40-17:55 Genome-wide association studies (GWAS) in an effort to provide insights into the complex interplay of nuclear receptor transcriptional networks and their contribution to the maintenance of homeostasis
Thanasis Mitsis, G.P. Chrousos, E. Eliopoulos, D.Vlachakis

17:55-18:10 Computer aided drug design and pharmacophore modelling towards the discovery of novel anti-ebola agents
Kalliopi Io Diakou, George P. Chrousos, Elias Eliopoulos, Dimitrios Vlachakis

18:10-18:25 3′-Tag RNA-sequencing
Andreas Gisel

18:25-18:40 Expression profiling of non-coding RNA in coronaviruses provides clues for virus RNA interference with immune system response
Arianna Consiglio, V.F. Licciulli, D. Catalano, G. Grillo, D. D'Elia

18:40-18:55 A computational drug design strategy against the Yellow Fever Virus helicase
Eleni Papakonstantinou, Katerina Pierouli, George Goulielmos, Elias Eliopoulos

18:55-19:10 Break

Chair *Emiliano Barreto-Hernandez*

19:10-19:25 Big Data analytics for knowledge transfer among organisms while reconstructing Gene Regulatory Networks
Paolo Mignone, G. Pio, D. D'Elia, M. Ceci

19:25-19:40 LP-HCLUS: a novel tool for the prediction of relationships between ncRNAs and human diseases
Emanuele Pio Barracchia, G.Pio, D. D'Elia, M. Ceci

19:40-20:00 Concluding Remarks

2nd Day: 24 September 2020

15:00-15:15 Welcome & Programme Presentation

Dimitrios Vlachakis & Lubos Klucar

Chair *Lubos Klucar*

15:15-15:30 Plant micro RNAs can control cancer genes expression through a sequence-specific targeting mechanism: the case of MALAT1 and NEAT1
Flaviana Marzano, M.F. Caratozzolo, A. Consiglio, F. Licciulli, S. Liuni, E. Sbisà, D. D'Elia, A. Tullo, D. Catalano

15:30-15:45 Benefits of a grape-rich diet on human health: a nutrigenomics study underlining the potential role of non-coding RNAs
Rosa Anna Milella, M. Gasparro, F. Alagna, M.F.Cardone, S. Rotunno, C.T. Ammollo, F. Semeraro, A. Tullo, F. Marzano, D. Catalano, M.Colucci, D. D'Elia

15:45-16:00 In silico characterization of the gene repertoires of immunoglobulins and T cell receptors of the various inbred laboratory strains of Mus musculus
Anna Tran, Géraldine Folch, Véronique Giudicelli and Sofia Kossida

16:00-16:15 Predicting L-PART1 exon using deep learning
Merouane Elazami Elhassani

16:15-16:30 Ml4microbiome: Statistical and machine learning techniques in human microbiome studies (CA 18131)
Tatjana Loncar-Turukalo, D. D'Elia **(https://www.ml4microbiome.eu/)**

16:30-16:45 Advancement of Bioinformatics and Computational Biology in Latin America: SoIBio and other scientific networks
Javier De Las Rivas, President of SoIBio (http://www.soibio.org/)

16:45-17:00 An ERASMUS experience in Greece
by David Coornaert and students **Cyril Radermecker & Ahmed Kanfoud**

17:00- 20:00 EMBnet Annual Business Meeting

Only EMBnet members

Programme Committee

Domenica D'Elia	Italy
Erik Bongcam-Rudloff	Sweden
Dimitrios Vlachakis	Greece
Lubos Klucar	Slovakia
Emiliano H. Barreto	Colombia
Aspasia Efthimiadou	Greece

EMBnet.journal
Bioinformatics in Action

Vlachakis et al. (2021) EMBnet.journal 26(Suppl A), e993
http://dx.doi.org/10.14806/ej.26.A.993

Exosomes in breast milk: a genetic trojan horse from mother to child

Dimitrios Vlachakis[1,2,3]✉, Aspasia Efthimiadou[4], Flora Bacopoulou[5], Elias Eliopoulos[1], George P. Chrousos[5]

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[2]Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[3]School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, United Kingdom

[4]Hellenic Agricultural Organization-Demeter, Institute of Soil and Water Resources, Department of Soil Science of Athens, Lycovrisi, Greece

[5]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece

Competing interests: DV none; AE none; FB none; EE none; GPC none

Breast milk is the ideal food for premature and mature babies and has undoubtedly immediate and ultimate benefits (Haschke *et al.*, 2016). Among other things, it protects against infections, reduces the risk of necrotising enterocolitis and retinopathy of premature babies, improves neurodevelopmental outcome, and reduces the risk of obesity and metabolic syndrome later in life (Oddy 2002). In the present study, breast milk is being studied with all the available omics technologies available. More specifically, functional genomics, comparative genomics, transcriptomics, sequencing, proteomics, and metabolomics are applied. The above results and this multidimensional information are coordinated under the framework of a holistic approach of systems biology and bioinformatic analysis. Important lncRNAs and protein molecules are being validated as candidate biomarkers in exosomes of a larger group of breast milk and blood/serum samples. Validated ncRNAs / proteins are being analysed in exudates of breast milk, bovine, goat, and sheep milk to explore new ways to understand the genetics underlying breast milk. The expression of ncRNAs, unlike mRNAs, is a direct indicator of their functional presence. The information generated in this study is analysed by data mining and data combining techniques and algorithms. The benefits of breast milk are attributed to its various components, including nutrients, hormones, growth factors, immune cells, antibodies, cytokines, antimicrobial peptides, and extracellular vesicles (O'Reilly *et al.*, 2021). Breast milk molecular fingerprinting will pave the way to shed light on the underlying genetics and epigenetics that a mother offers to her child.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code:T2EDK-02222).

References

1. Haschke F, Haiden N and Thakkar SK (2016) Nutritive and Bioactive Proteins in Breastmilk. Ann Nutr Metab **69 Suppl 2**, 17-26. http://dx.doi.org/10.1159/000452820
2. O'Reilly D, Dorodnykh D, Avdeenko NV, Nekliudov NA, Garssen J *et al.* (2021) Perspective: The Role of Human Breast-Milk Extracellular Vesicles in Child Health and Disease. Adv Nutr **12**(1), 59-70. http://dx.doi.org/10.1093/advances/nmaa094
3. Oddy WH (2002) The impact of breastmilk on infant and child health. Breastfeed Rev **10**(3), 5-18.

EMBnet.journal
Bioinformatics in Action

Klingström et al. (2021) EMBnet.journal 26(Suppl A), e966
http://dx.doi.org/10.14806/ej.26.A.966

Oral Presentations

Precision dairy farming – A Phenomenal opportunity with SLU Gigacow

Tomas Klingström[1]✉, Karl-Johan Petersson[1], Natalie von der Lehr[2], Hans Persson[3], Dirk-Jan de Koning[1]

[1] Department of Animal Breeding and Genetics, Quantitative genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] Department of Animal Breeding and Genetics. Swedish University of Agricultural Sciences, Uppsala, Sweden
[3] Department of Animal Breeding and Genetics, Interbull Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden
Competing interests: TK none; KJP none; NVDL none; HP none; DJDK none

In a modern dairy farm, machines and sensors will automatically collect data on milk composition, activity, and cows' feeding behaviour to ensure good animal health and high productivity. Genotyping of calves for the estimation of genetic breeding values is now also a routine procedure at many farms, and genotyping with a 45K SNP-chip is available as a commercial service[1] at the cost of approximately 35 € per calf. In a farm, each cow's pedigree is known, and the herd structure ensures that between 60 and several hundred individuals will be exposed to the same environment, which means that collaboration with dairy farmers or research farms can yield extensive data on phenotypes and genotypes for a large number of animals living in the same environment.

SLU Gigacow[2] is an infrastructure by the Swedish University of Agricultural Sciences (SLU) to collect large-scale data collection from dairy farms. With the SLU Gigacow infrastructure, we aim to collect as many measurable characteristics as possible from cows on farms participating in the network. We then supplement these measurements with genetic information from each animal using a commercial 45K SNP-chip and other information about the farms to create a state-of-the-art infrastructure for genotyping and phenotyping. We currently collaborate with 17 commercial dairy farms and two research farms to provide daily data collection from over 5000 dairy cows in the network. The data is available to researchers at SLU and by agreement with external stakeholders working to improve the productivity, profitability and sustainability of the dairy farming industry.

For fundamental research, the unique production characteristics of dairy farms offer many opportunities yet to be fully exploited. With SLU Gigacow, we can combine genomic information with detailed phenotyping through automated or semi-automated monitoring systems to create large cohorts of individuals in a shared environment to study phenotypes, complex social behaviour as well as host-microbiome studies.

The presentation will cover the development of the Gigacow infrastructure and the research opportunities presented to researchers at SLU and international collaborators.

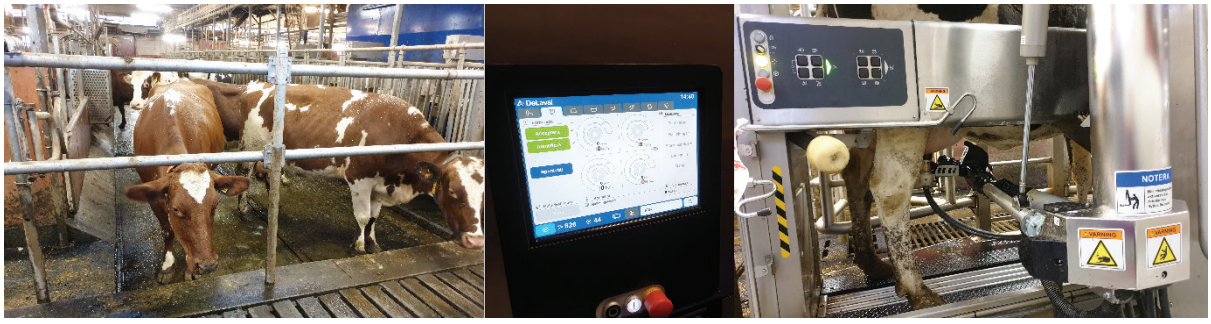

Figure 1. Swedish Red and White Cattle, the information dashboard mounted on the DeLaval Voluntary Milking System and the milking robot itself.

[1] https://www.eurogenomics.com/actualites/eurogenomics-new-eurog-md-beadchip.html
[2] https://www.slu.se/institutioner/husdjursgenetik/forskning/gigacow/

EMBnet.journal
Bioinformatics in Action

Leonidakis et al. (2021) EMBnet.journal 26(Suppl A), e996
http://dx.doi.org/10.14806/ej.26.A.996

Climate mitigation. Is it possible via DSS for agriculture? A case study: Reducing the Environmental Footprint of cotton cultivation

Dimitrios Leonidakis[1], Nikolaos Katsenios[2], Panagiotis Sparangis[2], Christoforos Nikitas Kasimatis[2], Dimitrios Vlachakis[3,4,5], Aspasia Efthimiadou[2]✉

[1]Farmacon G.P., K. Therimioti 25, Giannouli, Larisa, 41500 Thessaly, Greece

[2]Department of Soil Science of Athens, Institute of Soil and Water Resources, Hellenic Agricultural Organization-Demeter, Sofokli Venizelou 1, Lycovrissi, 14123 Attica, Greece

[3]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, 11855 Athens, Greece

[4]Lab of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, 11527 Athens, Greece

[5]Department of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, United Kingdom

Competing interests: DL none; NK none; PS none; CNK none; DV none; AE none

Cotton is a very important industrial cultivation in Greece and all over the world. The main issue regarding the cultivation of cotton is irrigation because it is strictly water-dependent (Dağdelen *et al.*, 2006). This means that the costs and the environmental footprint are high.
However, recent research indicated that the amounts of water that are used for irrigation could be reduced. The technological evolution has made Neural Network based Decision Support Systems, combined with data analysis, to be considered the future of sustainable agriculture. Cotton growers need to adopt new technologies not only to increase production but also to reduce water needs. The development of an innovative cotton production support system was conducted, consisting of five different types of measurements. Data from IoT sensors, weather stations, remote sensing data (Sentinel 2 images), soil analysis and on-site measurements (yield and EM38) derived from five experimental fields in Greece, creating a dataset of thirteen different inputs. A total of thirteen different algorithms were put into the test and evaluated to find the best one in terms of time and efficiency. In this research, we implemented a Decision Support System to assess the true water need of cotton cultivation (Salinari *et al.*, 2014).

References

1. Dağdelen N, Yılmaz E, Sezgin F, Gürbüz T (2006) Water-yield relation and water use efficiency of cotton (Gossypium hirsutum L.) and second crop corn (Zea mays L.) in western Turkey. Agricultural Water Management **82**(1):63–85. http://dx.doi.org/10.1016/j.agwat.2005.05.006
2. Salinari F, Mariani L, Poni S, Cola G, Bettati T *et al.* (2014) Development of a water stress alert system embedded in a DSS for Integrated Vineyard Management. Acta Horticulturae **1038**:565–572. http://dx.doi.org/10.17660/ActaHortic.2014.1038.71

EMBnet.journal
Bioinformatics in Action

Donato et al. (2021) EMBnet.journal 26(Suppl A), e960
http://dx.doi.org/10.14806/ej.26.A.960

Oral Presentations

Precision Epidemiology of Multi-drug resistant bacteria: bioinformatics tools

John Donato[1]✉, Luis Lugo[1], Hermes Perez[1], Harold Ballen[1], Diego Talero[1], Sebastian Prada[1], François Brion[2], Veronica Rincon[1], Laurent Falquet[3], Maria Teresa Reguero[1], Emiliano Barreto-Hernandez[1]

[1] Bioinformatics Center, Biotechnology Institute, Universidad Nacional de Colombia, Bogota, Colombia
[2] Haute Ecole en Hainaut – Department of Sciences and technologies, Mons, Belgium
[3] Biochemistry/Bioinformatics Unit, Université de Fribourg, and Swiss Institute of Bioinformatics, Fribourg, Switzerland
Competing interests: JD none; LL none; HP none; HB none; DT none; SP none; FB none; VR none; LF none; MTR none; EBH none

Precision epidemiology is a tool that allows researchers and the health community to understand, following, and controlling the infections, overall when the microorganisms are antibiotic-resistant. Next-Generation sequencing (NGS) techniques have become an important tool for the precise identification and genomic characterisation of these microorganisms, facilitating their accurate epidemiological monitoring. The World Health Organization (WHO) has prioritised the development of tools such as the Global Antimicrobial Resistance Surveillance System (GLASS) (WHO, 2015) to collect clinical, epidemiological, and laboratory data as a global system for antimicrobial resistance surveillance and to serve as a repository for accurate epidemiological monitoring.

Considering the WHO guidelines, we are developing the Genomic Information Management System (SGIG) (Donato, 2018) to integrate clinical, epidemiological, laboratory, and genomic data obtained by NGS. It consists of several modules: 1) the module enabling the entry of patient, microbiological, and molecular data; 2) the module for processing of NGS data, from its quality assessment to obtaining the annotation and comparison of the assembled genomes, using tools such as FastQC, Spades, Prokka, RGI CARDdb, and Roary; 3) the module for the identification and typification of bacteria, that uses a bidirectional recurrent neural network architecture (Lugo, 2018), along with the inclusion of standard methods such as rMLST (Jolley *et al.*, 2012); 4) the module that uses a script in Python for the prediction of the antibiotic resistance profile, which searches for resistance genes in each genome stored in the local database and relates them with the antibiotic resistance data associated to each gene, through the presence-absence rules of genomic determinants obtained and cured of the information reported in the literature (Perez, 2017). And module 5, which creates reports where it is possible to have access to the clinical and demographic statistics, isolate phenotypic resistance profiles, isolate resistance profile predictions based on the genomics data and isolate comparative genomic results.

The system is built using the Groovy and Grails v3.1.9 frameworks. The system backend consumes services built with Biopython v1.70 and BioPerl v5.8. The database is implemented in MySQL v5.6.30. The system is installed on a DELL server with the operating system openSUSE Leap 42.1.4. Funded by Univesridad Nacional de Colombia and Colciencias (project Código:66234).

References

1. Donato JJ (2018) Sistema de gestión de información de genomas completos de cepas de Acinetobacter baumannii para la identificación, tipificación y seguimiento de resistencia a antibióticos. Master thesis, Universidad Nacional de Colombia - Sede Bogotá. https://repositorio.unal.edu.co/handle/unal/62957 (accessed 02 February 2021)
2. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C *et al.* (2012) Ribosomal multilocus sequence typing universal characterisation of bacteria from domain to strain. Microbiology **158**(Pt 4):1005-1015. https://doi.org/10.1099/mic.0.055459-0
3. Lugo LE (2018) A Recurrent Neural Network approach for whole-genome bacteria classification. Master thesis, Universidad Nacional de Colombia - Sede Bogotá. https://repositorio.unal.edu.co/handle/unal/68663 (accessed 05 February 2021)
4. Perez H. (2017) Modelo bioinformático para predecir la resistencia a antibióticos a partir del genoma de una Bacteria. Master thesis, Universidad Nacional de Colombia - Sede Bogotá. https://repositorio.unal.edu.co/handle/unal/62924 (accessed 05 February 2021)
5. WHO - World Health Organization (2015) Global Antimicrobial Resistance Surveillance System (GLASS). Retrieved from: https://www.who.int/glass/en/ (accessed 02 February 2021)

EMBnet.journal
Bioinformatics in Action

Leptidis et al. (2021) EMBnet.journal 26(Suppl A), e958
http://dx.doi.org/10.14806/ej.26.A.958

Single-cell mapping of microRNA expression during cardiac development

Stefanos Leptidis[1], Eleni Papakonstantinou[1], Katerina Pierouli[1], Athanasios Mitsis[1], Sarantis Chlamydas[2], Aspasia Efthimiadou[3], George P. Chrousos[4,5], Elias Eliopoulos[1], Emil Hansson[6], Dimitrios Vlachakis[1,7,8]✉

[1] Laboratory of Genetics, Department of Biotechnology, School of Food, Biotechnology and Development, Agricultural University of Athens, Athens, Greece

[2] Active Motif, Office park Nysdam, Avenue Reine Astrid 92, La Hulpe, Belgium

[3] Hellenic Agricultural Organization-Demeter, Institute of Soil and Water Resources, Department of Soil Science, Lycovrisi, Greece

[4] Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[5] Center for Adolescent Medicine and UNESCO Chair on Adolescent Health Care, First Department of Pediatrics, Medical School, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece

[6] Karolinska Institutet/AstraZeneca Integrated Cardio Metabolic Centre, Huddinge, Sweden

[7] Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[8] School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, United Kingdom

Competing interests: SL none; EP none; KP none; AM none; SC none; AE none; GPC none; EE none; DV none

The heart is an exceptionally complex tissue comprised of a variety of different cell types. Understanding physiological cardiac development and its relationship to the development of pathological cardiac diseases require the careful investigation of their related developmental pathways. A highly significant regulatory layer during cellular differentiation is the post-transcriptional regulation via non-coding RNAs and, more specifically, microRNAs (Liu *et al.*, 2010). Previous microRNA transcriptomic studies in the heart lacked in the identification of their differential expression per cell-type (Leptidis *et al.*, 2013). Since microRNAs can target many mRNAs, identifying their cell-type-specific expression is necessary to elucidate the intricate cellular interactions and regulatory pathways and the development of targeted therapeutic approaches.

This study uses data from single-cell small RNA sequencing (small-seq) (Faridani *et al.*, 2016) from early embryonic cardiac progenitor murine cells. We aim to identify the transcriptional profile of small RNAs, mainly microRNAs, during cardiac development. Unlike single-cell RNA sequencing (scRNAseq), there are no established cell-type markers nor data analysis methods in the case of small-seq. Thus, we develop a methodology for identifying cell-types using their microRNA profile, coupled to their predicted targets stemming from various miRNA target prediction algorithms. These data are then cross-referenced with preliminary scRNAseq data in the same tissue, with established cell-types. Deciphering the transcriptomic landscape of microRNAs during cardiac development, along with identifying cell-types based on the relationship between their RNA and microRNA fingerprint, enables the in-depth study of the intricate regulatory interactions between cells, cell-types and different embryonic days.

Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Reinforcement of Postdoctoral Researchers - 2nd Cycle" (MIS-5033021), implemented by the State Scholarships Foundation (ΙΚΥ).

References

1. Wang J, Chen Y, Hao S, Feng W, Shen Z (2017) Balanced Faridani OR et al. (2016) Single-cell sequencing of the small-RNA transcriptome. Nat Biotechnol **34**:1264-1266. http://dx.doi.org/10.1038/nbt.3701
2. Leptidis S *et al.* (2013) A deep sequencing approach to uncover the miRNOME in the human heart. PLoS One **8**:e57800. http://dx.doi.org/10.1371/journal.pone.0057800
3. Liu N, Olson EN (2010) MicroRNA regulatory networks in cardiovascular development. Dev Cell **18**:510-525. http://dx.doi.org/10.1016/j.devcel.2010.03.010

Ευρωπαϊκή Ένωση
European Social Fund

Operational Programme
Human Resources Development,
Education and Lifelong Learning
Co-financed by Greece and the European Union

ΕΣΠΑ 2014-2020

EMBnet.journal
Bioinformatics in Action

Pierouli et al. (2021) EMBnet.journal 26(Suppl A), e962
http://dx.doi.org/10.14806/ej.26.A.962

Oral Presentations

Genome regulation by long non-coding RNAs

Katerina Pierouli[1]✉, George N. Goulielmos[2], Elias Eliopoulos[1], Dimitrios Vlachakis[1,3,4]

[1] Laboratory of Genetics, Department of Biotechnology, School of Food, Biotechnology and Development, Agricultural University of Athens, Athens, Greece

[2] Section of Molecular Pathology and Human Genetics, Department of Internal Medicine, School of Medicine, University of Crete, Heraklion, Greece

[3] Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[4] Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London, United Kingdom

Competing interests: KP none; GNG none; EE none; DV none

Regarding the human genome, only 2-3% of it translates into proteins, whereas 97-98% of the genome is comprised by sequences that are not translated, and thus defined as "dark DNA" that appear as non-coding RNAs (ncRNAs) and are interrelated to the regulation of gene expression. Long ncRNAs (lncRNAs) consist of more than 200 nucleotides and are derived from various regions in the genome (Mercer, Dinger, & Mattick, 2009; Wilusz, Sunwoo, & Spector, 2009). Several lncRNAs create RNA–protein, RNA–DNA and RNA–RNA complexes that are associated with chromatin modifications and lead the transcription factors to specific genomic DNA targets. Another function of lncRNAs is the regulation of the mRNA translation levels by interfering with miRNAs. For that reason, they are associated with various diseases, such as cancer, myocardial infractions, and Alzheimer's disease (Jarroux, Morillon, & Pinskaya, 2017; Ma, Bajic, & Zhang, 2013). The basic functions of ncRNAs are imprinted in the process of: 1) translation; 2) splicing; 3) replication; and 4) gene regulation (Mattick & Makunin, 2006). Remarkably, alternative splicing allows the human genome to direct the synthesis of more proteins than expected from the 20,000 genes encoding proteins (Black, 2003). For that reason, the analysis of lncRNAs functions in gene expression and genome regulation is pivotal. Using big data processing, recording potential repeated motifs and epigenetic modifications in splicing sites, and examining the interactions of lncRNAs with the chromatin remodelling complexes could lead to the discovery of lncRNAs that constitute potential drug targets and leading to more specialised and personalised treatment.

References

1. Black DL (2003) Mechanisms of alternative pre-messenger RNA splicingAnnu Rev Biochem, **72**:291-336. http://dx.doi.org/10.1146/annurev.biochem.72.121801.161720
2. Jarroux J, Morillon A, Pinskaya M (2017) History, Discovery, and Classification of lncRNAs. Adv Exp Med Biol, **1008**:1-46. http://dx.doi.org/10.1007/978-981-10-5203-3_1
3. Ma L, Bajic VB, Zhang Z (2013) On the classification of long non-coding RNA. sRNA Biol, **10**(6):925-933. http://dx.doi.org/10.4161/rna.24604
4. Mattick JS, Makunin IV (2006) Non-coding RNA. Human Molecular Genetics, **15**(suppl_1):R17-R29. http://dx.doi.org/10.1093/hmg/ddl046
5. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet, **10**(3):155-159. http://dx.doi.org/10.1038/nrg2521
6. Wilusz JE, Sunwoo H, Spector DL (2009) Long non-coding RNAs: functional surprises from the RNA world. Genes Dev, **23**(13):1494-1504. http://dx.doi.org/10.1101/gad.1800909

EMBnet.journal
Bioinformatics in Action

Damaskopoulou et al. (2021) EMBnet.journal 26(Suppl A), e995
http://dx.doi.org/10.14806/ej.26.A.995

Mechanisms of epigenetic inheritance in children following exposure to abuse

Elissavet Damaskopoulou[1]✉, George P. Chrousos[2], Elias Eliopoulos[3], Dimitrios Vlachakis[3,4,5]

[1]Agricultural University of Athens, Athens, Greece

[2]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece

[3]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[4]Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[5]School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, United Kingdom

Competing interests: ED none; GPC none; EE none; DV none

Child abuse refers to any kind of bad treatment of a minor under 18, of a physical and/or emotional nature, sexual abuse, neglect and negligent conduct and exposure to danger.

Abuse, in any form, is known to have devastating effects on the later course of children's lives. What is not known yet, is whether the genes expressions of abused children raised in an institution is affected by the family and then the institutional environment in which they live the first years of their lives. In recent years, studies have shown that adults who have been exposed in their childhood to neglect, abuse and stress due to adverse environmental conditions, are more likely to develop chronic biological (organic) and psychological diseases such as depression, hypertension, diabetes, various cardiorespiratory and autoimmune diseases and this now seem to have a biological background. Similar studies have shown that as children grow older, they experience a variety of emotional, social and behavioural problems, including hyperactivity, learning difficulties and stress. DNA methylation is one of the mechanisms through which the epigenetic process occurs and the differentiation of DNA expression in these individuals.

Most studies in humans and animals have focused on epigenetic variations in genes called adversity genes. The most common of these is the glucocorticoid receptor (GR) gene, also known as NR3C1, which shows a differentiation in its expression. Other genes that studies have shown to be related to child abuse are FOXP1 and FOXP2 and FKBP5 and SLC6A4. The genetic and epigenetic interplay between the aforementioned genes is studied and presented herein in an effort to elucidate their role in the heritability of child abuse.

EMBnet.journal
Bioinformatics in Action

Mitsis et al. (2021) EMBnet.journal 26(Suppl A), e952
http://dx.doi.org/10.14806/ej.26.A.952

Oral Presentations

Genome-wide association studies (GWAS) to provide insights into the complex interplay of nuclear receptor transcriptional networks and their contribution to the maintenance of homeostasis

Thanasis Mitsis[1], Dimitrios Vlachakis[1,2,3]✉, George P. Chrousos[2,3], Elias Eliopoulos[1]

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology & Biotechnology, Agricultural University of Athens, Athens, Greece

[2]University Research Institute of Maternal and Child Health & Precision Medicine, and UNESCO Chair on Adolescent Health Care, National and Kapodistrian University of Athens, "Aghia Sophia" Children's Hospital, Athens, Greece

[3]Division of Endocrinology and Metabolism, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Competing interests: TM none; DV none; GPC none; EE none

Genome-wide association studies (GWAS) have become an essential tool in exploring the relationship between common sequence variation sites and specific traits (Tam *et al.*, 2019). A protein family that provides ample research material for GWAS are nuclear receptors. These receptors comprise one of the largest groups of transcriptional factors and regulate the activity of a wide range of biological processes (Weikum *et al.*, 2018). A biological system that seems to be heavily reliant on nuclear receptors' activity is homeostasis. Homeostasis can be described as the inner equilibrium, both physical and chemical, required for proper organism function. Homeostasis can be threatened by internal or external unforeseen stimuli called stressors, and as such, organisms have developed a complex mechanism that copes with such threats and acts to maintain homeostasis called the stress response system (Chrousos, 2009). This research uses the glucocorticoid receptor (GR), a vital mediator of the stress response system (Nicolaides *et al.*, 2015), and a heavily researched nuclear receptor as the basis of a GWAS research on the interplay between nuclear receptors and homeostasis. Specifically, a comprehensive list of epigenetic factors, receptor cofactors, and enzymes that interact with GR was constructed in an effort to create a concise network of the various biological functions this receptor partakes. This network, plus the remaining nuclear receptors found in humans was studied with a large genomic dataset. The results are expected to provide insight into the interplay of nuclear receptor transcriptional networks and their contribution to homeostasis maintenance.

References

1. Chrousos GP (2009) Stress and disorders of the stress system. Nature Reviews Endocrinology **5**(7), 374-81. http://dx.doi.org/10.1038/nrendo.2009.106
2. Nicolaides NC, Kyratzi E, Lamprokostopoulou A, Chrousos GP, Charmandari E (2015) Stress, the stress system and the role of glucocorticoids. Neuroimmunomodulation **22**(1-2), 6-19. http://dx.doi.org/10.1159/000362736
3. Tam V, Patel N, Turcotte M, Bossé Y, Paré G *et al.* (2019) Benefits and limitations of genome-wide association studies. Nature Reviews Genetics **20**(8), 467-484. http://dx.doi.org/10.1038/s41576-019-0127-1
4. Weikum ER, Liu X, Ortlund EA (2018) The nuclear receptor superfamily: A structural perspective. Protein Sci. **27**(11), 1876-92. http://dx.doi.org/10.1002/pro.3496

EMBnet.journal
Bioinformatics in Action

Diakou et al. (2021) EMBnet.journal 26(Suppl A), e961
http://dx.doi.org/10.14806/ej.26.A.961

Computer-aided drug design and pharmacophore modelling towards the discovery of novel anti-ebola agents

Kalliopi Io Diakou[1]✉, George P. Chrousos[2], Elias Eliopoulos[1], Dimitrios Vlachakis[1]

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology & Biotechnology, Agricultural University of Athens, Athens, Greece

[2]1st Department of Pediatrics, "Agia Sophia" Children's Hospital, Athens, Greece

Competing interests: KID none; GPC none; EE none; DV none

Ebolavirus is a genus of the Filoviridae viral family, containing six known species (de La Vega *et al.*, 2015). Four species of this viral family (Ebola, Sudan, Taï Forest, and Bundibugyo viruses) cause human disease in the form of viral hemorrhagic fevers, with frequent outbreaks that reach epidemic scale in the African continent, exhibiting high numbers of casualties (Rugarabamu *et al.*, 2020). The Ebola virus (EBOV) genome is a linear, single-stranded, non-segmented, negative-sense RNA containing seven genes, which code for structural and non-structural proteins (Mühlberger, 2007). Among these proteins is the viral glycoprotein (GP), the only virally expressed protein on the virion surface, critical for attachment to host cells and catalysis of membrane fusion (Lee and Saphire, 2009). The viral glycoprotein (GP) is produced through proteolytic cleavage of the precursor (pre-GP) and is comprised of two subunits (GP1 and GP2), connected by a disulfide bond (Ning *et al.*, 2017). As a result of its critical role in the virus life cycle and replication, the EBOV GP is a crucial component in vaccine development and an essential target in the research for neutralising antibodies and inhibitors of attachment and fusion (Hoenen *et al.*, 2019). In addition to standard approaches, the study of possible post-translational modifications concerning the EBOV GP can provide new insight into the efforts of developing new anti-ebola agents (Cook and Lee, 2013).

References

1. Cook JD and Lee JE (2013) The secret life of viral entry glycoproteins: moonlighting in immune evasion. PLoS Pathog **9**(5):e1003258-e1003258. http://dx.doi.org/10.1371/journal.ppat.1003258
2. de La Vega MA, Stein D and Kobinger GP (2015) Ebolavirus Evolution: Past and Present. PLoS Pathog **11**(11):e1005221-e1005221. http://dx.doi.org/10.1371/journal.ppat.1005221
3. Hoenen T, Groseth A and Feldmann H (2019) Therapeutic strategies to target the Ebola virus life cycle. Nat Rev Microbiol **17**(10):593-606. http://dx.doi.org/10.1038/s41579-019-0233-2
4. Lee JE and Saphire EO (2009) Ebolavirus glycoprotein structure and mechanism of entry. Future Virol **4**(6): 621-635. http://dx.doi.org/10.2217/fvl.09.56
5. Mühlberger E (2007) Filovirus replication and transcription. Future Virol **2**(2):205-215. http://dx.doi.org/10.2217/17460794.2.2.205
6. Ning YJ, Deng F, Hu Z and Wang H (2017) The roles of ebolavirus glycoproteins in viral pathogenesis. Virol Sin **32**(1):3-15. http://dx.doi.org/10.1007/s12250-016-3850-1
7. Rugarabamu S, Mboera L, Rweyemamu M, Mwanyika G, Lutwama J *et al.* (2020) Forty-two years of responding to Ebola virus outbreaks in Sub-Saharan Africa: a review. BMJ Glob Health **5**:e001955. http://dx.doi.org/10.1136/bmjgh-2019-001955

EMBnet.journal
Bioinformatics in Action

Olagunju et al. (2021) EMBnet.journal 26(Suppl A), e968
http://dx.doi.org/10.14806/ej.26.A.968

Oral Presentations

3'-Tag RNA-sequencing

Temitayo Adebanji Olagunju[1,2], Chisom Ezekannagha[1], Andreas Gisel[1,3]✉

[1] International Institute of Tropical Agriculture, Ibadan, Nigeria
[2] University of Ibadan, Ibadan, Nigeria
[3] CNR, Institute for Biomedical Technologies, Bari, Italy
Competing interests: TAO none; CE none; AG none

3'Tag-Seq is an approach to produce gene expression profiling data for small budgets. The main difference to traditional RNA-seq is that 3'Tag-seq produces one read per transcript and is not resequencing the whole mRNA. This procedure allows the user to request up to 10 times fewer reads for the same transcriptome analysis and therefore is confronted with close and NOT closing to 10 times cheaper sequencing costs. 3'Tag-seq provides high-quality expression data but lacks additional information on alternative splicing events provided by the regular RNA-seq (Ma *et al.*, 2019). Since most of the transcriptome analysis is focused on expression data and differential expression analysis, 3'Tag-seq is, for these cases, the approach of choice. In our work, we outlined the procedure on how to analyse such 3'Tag-seq data, evidence the problem of false-positive counts due to expressed regions with poly T regions which could mimic in the mature RNA polyA tails. Further, we highlighted that 3'Tag-seq can give information on the position of 3'UTR regions. In a test case of the expression analysis of cassava (Manihot esculenta) under our experimental conditions, we find read hits in only 10996 of the 26351 annotated 3'UTR (Prochnik *et al.*, 2012); note that additional 15031 genes do not have a 3'UTR annotation. However, most of these genes contain read hits in the downstream regions up to 1500nt after the gene end. This technology not only gives precise expression information but also data for alternative gene annotation in the region of the 3'UTR.

Acknowledgements

This work has been supported by BMGF (grant INV-008053 "Metabolic Engineering of Carbon Pathways to Enhance Yield of Root and Tuber Crops") and Genomix4Life, Salerno, Italy

References

1. Ma F, Fuqua BK, Hasin Y, Yukhtman C, Vulpe CD, Lusis AJ and Pellegrini M (2019) A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. BMC Genomics **20**:9. http://dx.doi.org/10.1186/s12864-018-5393-3
2. Prochnik S, Marri PR, Desany B, Rabinowicz PD *et al.* (2012) The Cassava Genome: Current Progress, Future Directions. Trop Plant Biol. 2012; **5**(1): 88–94. http://dx.doi.org/10.1007/s12042-011-9088-z

EMBnet.journal
Bioinformatics in Action

Consiglio et al. (2021) EMBnet.journal 26(Suppl A), e954
http://dx.doi.org/10.14806/ej.26.A.954

Expression profiling of non-coding RNA in coronaviruses provides clues for virus RNA interference with the human immune system response

Arianna Consiglio✉, **Flavio Licciulli, Domenico Catalano, Giorgio Grillo, Domenica D'Elia**

Institute for Biomedical Technologies, National Research Council, Bari, Italy
Competing interests: AC none; FL none; DC none; GG none; DD none

Due to viral infection, the human immune system activates a complex response whose magnitude also depends on the interplay between the virus and the host's immune response regulation. The most dangerous effects induced by the SARS-CoV-2 infection are an exacerbated inflammatory response and an extensive lung pathology. Related to the damages caused by the inflammatory response, an important aspect that deserves to be investigated is the host cell response at a very early stage of the virus infection. An extensive analysis of transcriptome profiles of infected cells is the most effective analysis approach to investigate to what extent and which gene signalling pathways are directly involved at this stage. To elucidate these aspects can immediately bring to the identification of biomarkers of infection and targets for new and more effective therapeutic approaches.

It is noteworthy that non-coding RNAs (ncRNAs) are essential regulators of human gene expression. Recent studies have demonstrated that viruses belonging to the family of SARS-CoV-2 can regulate the expression of small (sRNA) and long non-coding RNAs (lncRNA) (Morales *et al.*, 2017; Liu and Ding, 2017). We have investigated the potential of the SARS-CoV-2 genome transcription to produce fragments of RNAs that can interfere with the host regulatory non-coding RNAs (small non-coding RNAs) by using a large scale bioinformatics analysis on data available in public repositories.

Comparing the SARS-CoV-2 genome sequence (NC_045512.2) with RNA-Seq data of human lung cancer cells infected with MERS-CoV [GEO ID GSE139516], mouse lung cells affected by SARS-CoV [GEO ID GSM907704] and bronchial lavages and Peripheral Blood Mononuclear Cells (PBMC) from COVID-19 patients (Xiong *et al.*, 2020) we discovered that small fragments of ncRNAs from SARS-CoV-2 might interfere with the activity of endogenous miRNAs that target genes involved in the inflammatory response, in particular with the allergic asthmatic reaction (IL4-IL13 signalling pathway).

The analysis was carried out with a bioinformatic pipeline developed using Python, BASH and R, and the software BLAST, STAR and mirdeep2 for the comparative analyses.

These preliminary results open the way for more effective treatment of COVID-19 patients and defence from future coronavirus pandemics. For further information: http://bioinformatics.ba.itb.cnr.it/CoV-ncRNASig

References

1. Morales *et al.* (2017) SARS-CoV-Encoded Small RNAs Contribute to Infection-Associated Lung Pathology. Cell Host & Microbe **21**, 344–55. http://dx.doi.org/10.1016/j.chom.2017.01.015
2. Weiwei Liu and Chan Ding (2017) Roles of LncRNAs in Viral Infections. Front Cell Infect Microbiol. **7**, 205. http://dx.doi.org/10.3389/fcimb.2017.00205
3. Yong Xiong *et al.* (2020) Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerging Microbes & Infections **9**(1), 761-770. http://dx.doi.org/10.1080/22221751.2020.1747363

EMBnet.journal
Bioinformatics in Action

Papakonstantinou et al. (2021) EMBnet.journal 26(Suppl A), e994
http://dx.doi.org/10.14806/ej.26.A.994

Oral Presentations

A computational drug design strategy against the Yellow Fever Virus helicase

Eleni Papakonstantinou[1]✉, Katerina Pierouli[1], Dimitrios Vlachakis[1,2,3], George N. Goulielmos[4], Elias Eliopoulos[1]

[1]Laboratory of Genetics, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece

[2]Laboratory of Molecular Endocrinology, Center of Clinical, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

[3]School of Informatics, Faculty of Natural & Mathematical Sciences, King's College London, London, United Kingdom

[4]Section of Molecular Pathology and Human Genetics, Department of Internal Medicine, School of Medicine, University of Crete, Heraklion, Greece

Competing interests: EP none; KP none; DV none; GNC none; EE none

Yellow Fever is an acute viral hemorrhagic disease transmitted through infected Aedes species mosquitoes. It causes fever, bleeding, shock, heatstroke, liver, kidney and myocardial damage and unfortunately, has a high mortality rate. While the Yellow Fever Virus (YFV) affects mainly parts of South America and Africa, in recent years, cases of infection on both animals and humans have been reported in North America, Asia and Europe. A potential YFV infection could have a significant impact on the health of the population, the economy, and the well-being of a country. We present a computational strategy for developing novel antiviral inhibitors that target the enzymatic activity of the YFV helicase. We use a holistic bioinformatic approach to enhance our understanding of the YFV helicase enzyme mode and design a series of compounds as candidate drugs against the endemic YFV virus. Phylogenetic studies and structural analyses of the viral helicase and RNA-helicase complex are performed to design a 3D pharmacophore that will incorporate all physicochemical properties essential for interaction and will be used for high throughput virtual screening and the identification of lead compounds. The in silico pipeline will be evaluated and optimised after in vitro experimental studies aiming to indicate the optimal precursors and their respective moieties. In this way, the in silico pipeline will allow for the discovery of the most potent molecules with an inhibitory effect on the YFV helicase function and the viral replication cycle.

Acknowledgements

This work is funded by the ESPA Young Researchers Support, «Rational Drug Design of Novel Antiviral Agents against the Helicase Enzyme of the Yellow Fever Virus», MIS 5048546, NSRF 2014 – 2020

EMBnet.journal
Bioinformatics in Action

Mignone et al. (2021) EMBnet.journal 26(Suppl A), e956
http://dx.doi.org/10.14806/ej.26.A.956

Big Data analytics for knowledge transfer among organisms while reconstructing Gene Regulatory Networks

Paolo Mignone[1,2], Gianvito Pio[1,2]✉, Domenica D'Elia[3], Michelangelo Ceci[1,2,4]

[1]Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[2]National Interuniversity Consortium for Informatics (CINI), Rome, Italy
[3]Institute for Biomedical Technologies, National Research Council, Bari, Italy
[4]Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

Competing interests: PM none; GP none; DD none; MC none

The reconstruction of gene regulatory networks (GRNs) from gene expression data is pivotal for understanding gene regulatory mechanisms and processes. In this context, machine learning and big data analytics tools can be considered fundamental. However, most existing methods (i) produce poor results when the amount of labelled examples is limited or when no negative example is available and (ii) they are not able to exploit information extracted from GRNs of other (better studied) related organisms.

We overcome these limitations by proposing an innovative transfer learning method, called BioSfer (Mignone *et al.*, 2020), which can exploit the knowledge about the GRN of a source organism for the reconstruction of the GRN of the target organism. In the first stages, we identify two predictive models to discover unknown links for both the considered GRNs. In the final stage, we build a new geometrically-combined model, which can identify unknown links better. Moreover, the proposed method is natively able to work in the positive-unlabeled setting, where no negative example is available, by fruitfully exploiting a set of unlabeled examples. In our experiments, we reconstructed the human GRN by exploiting the knowledge of the GRN of M. musculus. The qualitative analysis showed that the proposed method is able to identify biologically plausible gene regulations that are not identified by other tools. Results showed that the proposed method outperforms state-of-the-art approaches (Zhang *et al.*, 2017; Wang *et al.*, 2017; Long *et al.*, 2014; Huynh-Thu *et al.*, 2010; Aibar *et al.*, 2017; Mignone *et al.*, 2018) and identifies previously unknown functional relationships among the analysed genes.

Availability of data and materials

The system, the adopted datasets and all the results are available at: http://www.di.uniba.it/~mignone/systems/biosfer/index.html

Acknowledgements

We acknowledge the EU Commission's support through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant number ICT-2013-612944) and of the National Research Council (CNR) Flagship Project InterOmics.

References

1. Aibar S, Bravo Gonzalez-Blas C, Moerman T, Huynh-Thu V, Imrichova H *et al.* (2017). SCENIC: Single-Cell Regulatory Network Inference And Clustering. Nature Methods **14**:1083-1086. http://dx.doi.org/10.1038/nmeth.4463
2. Huynh-Thu V, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS ONE **5**(9):e12776. http://dx.doi.org/10.1371/journal.pone.0012776
3. Long M, Wang J, Ding G, Sun J, Yu PS (2014) Transfer Joint Matching for Unsupervised Domain Adaptation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1410-1417. http://dx.doi.org/10.1109/CVPR.2014.183
4. Mignone P, Pio G. (2018) Positive Unlabeled Link Prediction via Transfer Learning for Gene Network Reconstruction. ISMIS 2018: The 24th International Symposium on Methodologies for Intelligent Systems. http://dx.doi.org/10.1007/978-3-030-01851-1_2, 2018
5. Mignone P, Pio G, D'Elia D, Ceci M (2020) Exploiting Transfer Learning for the Reconstruction of the Human Gene Regulatory Network. Bioinform. **36**(5):1553-1561. http://dx.doi.org/10.1093/bioinformatics/btz781
6. Wang J, Chen Y, Hao S, Feng W, Shen Z (2017) Balanced distribution adaptation for transfer learning. In IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 2017, pp. 1129-1134. http://dx.doi.org/10.1109/ICDM.2017.150
7. Zhang J, Li W, Ogunbona P (2017) Joint geometrical and statistical alignment for visual domain adaptation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5150-5158. http://dx.doi.org/10.1109/CVPR.2017.547

EMBnet.journal
Bioinformatics in Action

Barracchia et al. (2021) EMBnet.journal 26(Suppl A), e955
http://dx.doi.org/10.14806/ej.26.A.955

Oral Presentations

LP-HCLUS: a novel tool for the prediction of relationships between ncRNAs and human diseases

Emanuele Pio Barracchia[1,2]✉, Gianvito Pio[1,2], Domenica D'Elia[3], Michelangelo Ceci[1,2,4]

[1] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[2] National Interuniversity Consortium for Informatics (CINI), Rome, Italy
[3] Institute for Biomedical Technologies, National Research Council, Bari, Italy
[4] Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

Competing interests: EPB none; GP none; DD none; MC none

The discovery of a functional relationship between human diseases and non-coding RNAs (ncRNAs) is not new. In the last decade, it improved the elucidation of many diseases' mechanisms and the improvement of therapeutic approaches (Lekka and Hall, 2018; Wang *et al.*, 2016; Yang *et al.*, 2014). Nevertheless, the function of many ncRNAs is still unclear or completely unknown, and therefore, their role in human diseases is difficult, if not impossible, to be identified. We have developed a new system, called LP-HCLUS, that is able to predict previously unknown disease-ncRNA associations by exploiting multi-type hierarchical clustering techniques.

Differently from other approaches, LP-HCLUS is able to analyse and benefit from heterogeneous networks of interactions/relationships among multiple types of entities (*e.g.*, diseases, ncRNAs, target genes) and relationships between them. To this aim, the proposed method first estimates the strength of the disease-ncRNA associations, exploiting both direct and indirect relationships. It constructs a hierarchy of heterogeneous clusters based on known and estimated relationships between diseases and ncRNAs. Finally, LP-HCLUS uses the generated clusters to induce new relationships, associating each of them with a certainty score. We conducted several experiments, comparing the performances achieved by LP-HCLUS with those obtained by two different competitors: HOCCLUS2 (Pio *et al.*, 2013) and ncPred (Alaimo *et al.*, 2014). In particular, we analysed two different datasets: HMDD v3.0, which contains data about relationships between diseases and miRNAs, and a dataset constructed integrating different

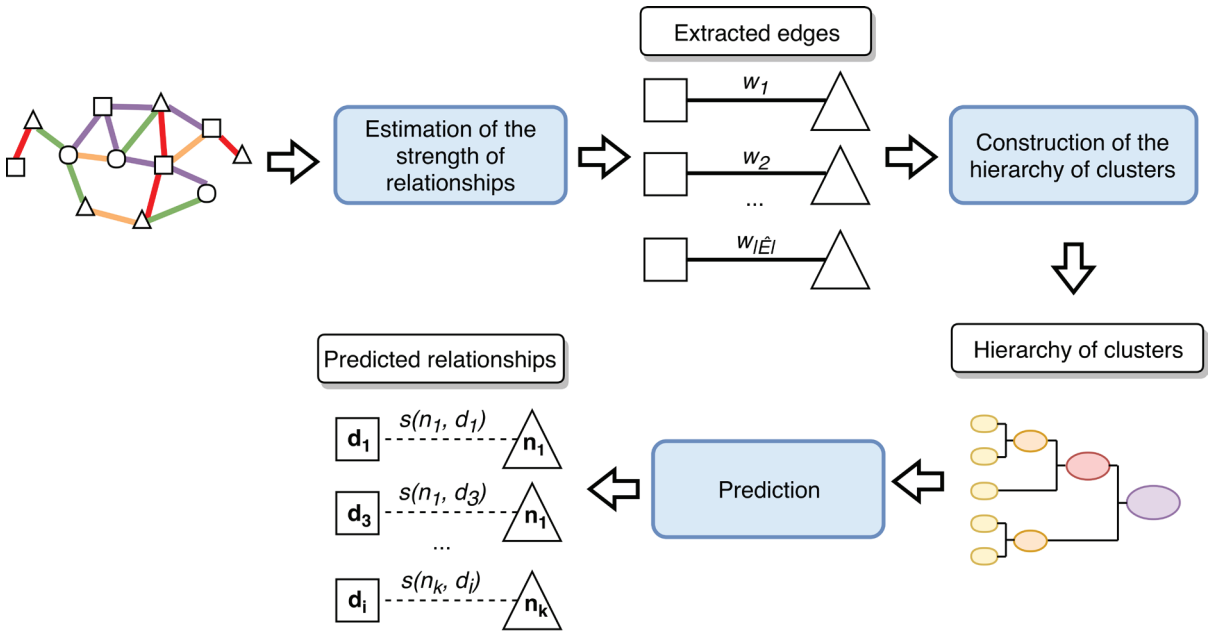

Figure 1. The workflow of the LP-HCLUS method.

state-of-the-art data sources (Chen *et al.*, 2013; Helwak *et al.*, 2013; Bauer-Mehren *et al.*, 2010; Jiang *et al.*, 2009).

The results show that our system is able to outperform its competitors, and it can help biologists to conduct more focused research. Such a conclusion is also confirmed by a qualitative analysis conducted on the predicted associations that showed that many associations predicted by LP-HCLUS with a high certainty score have been subsequently validated and introduced in a more recent version of HMDD dataset (v3.2). The importance of such a development is also in its easy transfer for applications in any biological study involving heterogeneous data from different sources and types (*e.g.*, different omics data, chemicals, biochemical and structural data, *etc.*).

Availability of data and materials

The system LP-HCLUS, the adopted datasets and all the results are available at: http://www.di.uniba.it/~gianvitopio/systems/lphclus/

Acknowledgements

We would like to acknowledge the financial support of the European Commission through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant Number ICT-2013-612944). We also acknowledge the financial support of the Ministry of Education, Universities and Research (MIUR) through the PON projects “Big Data Analytics” (AIM1852414 - Activity 1, Line 1) and TALIsMAn - Tecnologie di Assistenza personALizzata per il Miglioramento della quAlità della vitA (Grant N. ARS01_0111), and of Italian National Research Council (CNR) through the InterOmics Flagship project.

References

1. Alaimo S, Giugno R, Pulvirenti A (2014) ncPred: ncRNA-Disease Association Prediction through Tripartite Network-Based Inference. Frontiers in Bioengineering and Biotechnology **2**. http://dx.doi.org/10.3389/fbioe.2014.00071
2. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics **26** (22):2924–2926. http://dx.doi.org/10.1093/bioinformatics/btq538
3. Chen G, Wang Z, Wang D, Qiu C, Liu M *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Research **41** (Database issue):983–986. http://dx.doi.org/10.1093/nar/gks1099
4. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell **153** (3):654–665. http://dx.doi.org/10.1016/j.cell.2013.03.043
5. Jiang Q, Wang Y, Hao Y, Juan L, Teng M *et al.* (2009) miR2disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Research **37** (Database issue):98-104. http://dx.doi.org/10.1093/nar/gkn714
6. Lekka E, Hall J (2018) Noncoding RNAs in disease. FEBS Letters **592** (17):2884-2900. http://dx.doi.org/10.1002/1873-3468.13182
7. Pio G, Ceci M, D'Elia D, Loglisci C, Malerba D (2013) A Novel Biclustering Algorithm for the Discovery of Meaningful Biological Correlations between microRNAs and their Target Genes. BMC Bioinformatics **14** (Suppl 7):8. http://dx.doi.org/10.1186/1471-2105-14-S7-S8
8. Wang P, Guo Q, Gao Y, Zhi H, Zhang Y *et al.* (2016) Improved method for prioritization of disease associated lncRNAs based on ceRNA theory and functional genomics data. Oncotarget **8** (3):4642–4655. http://dx.doi.org/10.18632/oncotarget.13964
9. Yang X, Gao L, Guo X, Shi X, Wu H *et al.* (2014) A Network Based Method for Analysis of lncRNA-Disease Associations and Prediction of lncRNAs Implicated in Diseases. Plos One **9** (1):e87797. http://dx.doi.org/10.1371/journal.pone.0087797

EMBnet.journal
Bioinformatics in Action

Marzano et al. (2021) EMBnet.journal 26(Suppl A), e984
http://dx.doi.org/10.14806/ej.26.A.984

Oral Presentations

Plant micro RNAs can control cancer genes expression through a sequence-specific targeting mechanism: the case of MALAT1 and NEAT1

Flaviana Marzano[1], Mariano Francesco Caratozzolo[1], Arianna Consiglio[2], Flavio Licciulli[2], Sabino Liuni[2], Elisabetta Sbisà[2], Domenica D'Elia[2]✉, Apollonia Tullo[1], Domenico Catalano[2]

[1] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council, Bari, Italy
[2] Institute for Biomedical Technologies, National Research Council, Bari, Italy
Competing interests: FM none; MFC none; AC none; FL none; SL none; ES none; DD none; AT none; DC none

It is well known that some plant compounds, or phytochemicals, can positively affect human health by reducing inflammation and oxidative stress (Issa *et al.*, 2006; Dell'Agli *et al.*, 2013). Although the large number of works published, the molecular mechanisms through which plants can impact human health are still unclear. Recent studies suggest that plant nutrients control the expression of human genes by DNA methylation and histone modifications (Choi and Friso, 2010; Tollefsbol 2014). Moreover, some plant/food-derived microRNAs (miRNAs) accumulate in the sera and tissues of various animals and regulate their gene expression in a sequence-specific manner (Zhang *et al.*, 2012; García-Segura *et al.*, 2013).

We performed a study using a combined in silico and experimental approach to investigate the potential effects and elucidate the molecular mechanisms of edible plant miRNAs on the expression of human genes involved in cancer onset and progression (Marzano *et al.*, 2020). This study demonstrates that plant miRNAs can bind human transcripts in a sequence-specific manner and that their binding is functional. In particular, we have shown that the plant mtr-miR-5754 and gma-miR-4995 directly target the tumour-associated long non-coding RNAs MALAT1 and NEAT1, respectively, in a sequence-specific manner, thus reducing cancer cell proliferation. For the first time, we provide evidence that plant miRNAs can also target human regulatory ncRNAs. These findings open the way to new biotechnological applications in human nutrition and chronic diseases prevention.

Acknowledgements

This work was supported by the National Research Council (CNR) MIUR Flagship Project InterOmics.

References

1. Choi SW, Friso S (2010) Epigenetics: A New Bridge between Nutrition and Health, Advances in Nutrition **1**(1):8–16. http://dx.doi.org/10.3945/an.110.1004
2. Dell'Agli M *et al.* (2013) Plant food supplements with anti-inflammatory properties: a systematic review. Crit Rev Food Sci Nutr **53**(4):403-13. http://dx.doi.org/10.1080/10408398.2012.682123
3. García-Segura L, Pérez-Andrade M, Miranda-Ríos J. (2013) The Emerging Role of MicroRNAs in the Regulation of Gene Expression by Nutrients. J Nutrigenet Nutrigenomics **6**:16-31. http://dx.doi.org/10.1159/000345826
4. Issa AY, Volate SR, Wargovich MJ (2006) The role of phytochemicals in inhibition of cancer and inflammation: New directions and perspectives. Journal of Food Composition and Analysis **19**:405–419. http://dx.doi.org/10.1016/j.jfca.2006.02.009
5. Marzano F *et al.* (2020) Plant miRNAs reduce cancer cell proliferation by targeting MALAT1 and NEAT1: a beneficial cross-kingdom interaction. Frontiers in Genetics **11**:552490. http://dx.doi.org/10.3389/fgene.2020.552490
6. Tollefsbol TO (2014) Dietary Epigenetics in Cancer and Aging. Cancer Treat Res. **159**. http://dx.doi.org/10.1007/978-3-642-38007-5_15
7. Zhang L *et al.* (2012) Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. Cell Res. **22**(1):107-26. http://dx.doi.org/10.1038/cr.2011.158

EMBnet.journal
Bioinformatics in Action

Milella et al. (2021) EMBnet.journal 26(Suppl A), e989
http://dx.doi.org/10.14806/ej.26.A.989

Benefits of a grape-rich diet on human health: a nutrigenomics study underlining the potential role of non-coding RNAs

Rosa Anna Milella[1], Marica Gasparro[1], Fiammetta Alagna[1,2], Maria Francesca Cardone[1], Silvia Rotunno[1,3], Concetta Tiziana Ammollo[4], Fabrizio Semeraro[4], Apollonia Tullo[5], Flaviana Marzano[5], Domenico Catalano[6], Mario Colucci[4], Domenica D'Elia[6]✉

[1]Research Centre for Viticulture and Enology, Council for Agricultural Research and Economics, Turi, Bari, Italy

[2]ENEA Italian National Agency for New Technologies Energy and Sustainable Economic Development, Trisaia Research Center, Rotondella, Matera, Italy

[3]Institute for Sustainable Plant Protection, National Research Council, Torino, Italy

[4]Department of Biomedical Sciences and Human Oncology, University of Bari "Aldo Moro", Bari, Italy

[5]Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council, Bari

[6]Institute for Biomedical Technologies, National Research Council, Bari, Italy

Competing interests: RAM none; MG none; FA none; MFC none; SR none; CTA none; FS none; AT none; FM none; DC none; MC none; DD none

Grape (*Vitis vinifera L.*) is one of the most typical fruit of the Mediterranean diet, characterised by high polyphenols content with marked antioxidant and anti-inflammatory activities (Goszcz *et al.*, 2017). By using transcriptional profiling techniques, it has become increasingly clear that polyphenols can influence the expression of genes. Many of these genes are key elements of cell signalling cascades (Spencer 2009; Fraga and Oteiza 2011) and regulatory non-coding RNAs (ncRNAs) (Budisan *et al.*, 2017).

A previous study showed that table grape extracts exert a marked antithrombotic activity in vitro (Ammollo *et al.*, 2017). To shed light on the molecular basis of grape intake effects on human health and investigate the potential role of ncRNAs, we carried out a nutrigenomics study (Milella *et al.*, 2020a and 2020b). To this aim, 20 healthy subjects were enrolled to follow a grape-rich diet for 21 days. The gene expression profiles of peripheral blood mononuclear cells (PBMCs) extracted from six of these subjects, after 21 days of fresh table grape-rich diet and after an additional 28-day washout, were analysed. The results showed 930 genes differentially expressed. Among these genes, more than 200 are long ncRNAs (lncRNAs), almost all downregulated after the washout period when polyphenols' direct effect is supposed to be completely exhausted.

The functional analysis of differentially expressed genes revealed significant changes in processes critical for organismal and cell wellbeing such as inflammation and immunity, thrombosis, DNA and protein repair, autophagy and mitochondrial biogenesis. From our analysis of lncRNAs down-regulated in our study, we observed that many of them are over-expressed in many types of tumours, chronic diseases due to the persistency of inflammation and metabolic syndromes such as obesity. Altogether, these findings provide exciting clues for the crucial role of ncRNAs in grape intake's long-term effects on a series of biological processes. These lncRNAs would deserve to be further investigated for potential applications in the care of chronic diseases and cancer and their prevention.

Acknowledgements

This work was supported by a grant from MIUR (Ministero Italiano dell'Università e della Ricerca - project "PON02_00186_2937475 - PROALIFUN") and by the CNR (National Research Council) Flagship Project InterOmics.

References

1. Ammollo CT *et al.* (2017) Grape intake reduces thrombin generation and enhances plasma fibrinolysis. Potential role of circulating procoagulant microparticles. J Nutr Biochem **50**:66–73. http://dx.doi.org/10.1016/j.jnutbio.2017.08.012
2. Budisan L *et al.* (2017) Dietary Intervention by Phytochemicals and Their Role in Modulating Coding and Non-Coding Genes in Cancer. Int J Mol Sci **18**(6). http://dx.doi.org/10.3390/ijms18061178
3. Fraga CG, Oteiza PI (2011) Dietary flavonoids: Role of (-)-epicatechin and related procyanidins in cell signalling. Free Radic Biol Med **51**(4):813-23. http://dx.doi.org/10.1016/j.freeradbiomed.2011.06.002
4. Goszcz K, Duthie GG, Stewart D, Leslie SJ, Megson IL (2017) Bioactive polyphenols and cardiovascular disease: chemical antagonists, pharmacological agents or xenobiotics that drive an adaptive response? Br J Pharmacol **174**:1209–25. http://dx.doi.org/10.1111/bph.13708
5. Milella RA *et al.* (2020b) Gene expression signature induced by grape intake in healthy subjects reveals wide-spread beneficial effects on peripheral blood mononuclear cells. Journal of Functional Food, **64**:103705. http://dx.doi.org/10.1016/j.jff.2019.103705
6. Milella RA *et al.* (2020a) Microarray data and pathway analyses of peripheral blood mononuclear cells from healthy subjects after a three weeks grape-rich diet. Data in Brief **29**, 105278. http://dx.doi.org/10.1016/j.dib.2020.105278
7. Spencer JP (2009) Flavonoids and brain health: multiple effects underpinned by common mechanisms. Genes Nutr. **4**(4):243-50. http://dx.doi.org/10.1007/s12263-009-0136-3

EMBnet.journal
Bioinformatics in Action

Tran et al. (2021) EMBnet.journal 26(Suppl A), e953
http://dx.doi.org/10.14806/ej.26.A.953

Oral Presentations

In silico characterisation of the gene repertoires of immunoglobulins and T cell receptors of the various inbred laboratory strains of *Mus musculus*

Anna Tran✉, **Géraldine Folch, Véronique Giudicelli, Marie-Paule Lefranc, Sofia Kossida**

IMGT®, The International ImMunoGeneTics Information System®, Institute of Human Genetics (IGH), Scientific Research National Center (CNRS), University of Montpellier, Montpellier, France

Competing interests: AT none; GF none; VG none; MPL none; SK none

The laboratory mouse is the most widely used animal model in the life sciences for the study of disease and human development. Mouse strains are known for their differences in the adaptive immune response, but the genomic repertoires of genes that code for antigen receptors, immunoglobulins or antibodies (IG) and T cell receptors (TR) are far from having been fully and precisely sequenced and/or characterised in each strain despite the existence of Mouse Genome Informatics resources dedicated to the species.

IG (proteins composed of two heavy chains or IGH, and two light chains IGK or IGL) and TR (composed of chains alpha and beta, or chains gamma and delta) are encoded by four types of genes, variable (V), diversity (D), joining (J), constant (C) belonging to multigene families and are very polymorphic. The synthesis of these molecules results from complex mechanisms, including rearrangements of the V, D and J genes at the DNA level, the mechanisms of N-diversity and, for IGs, of somatic hypermutations. These mechanisms are at the origin of an extreme diversity of IG and TR (potentially more than 2.1012 IG and 2.1012 different TR per individual) and the effectiveness of the adaptive immune system.

Knowing and understanding the organisation of these repertoires in the different strains is therefore essential for understanding the reactions of the adaptive immune system and for the choice of mouse models in biology. For example, on IGH locus, the most widely used inbred strains C57BL/6 and BALB/c have only a few sequences in common, which means that their IGH locus are probably a mosaic of very disparate genes. It is highly probable that the same holds true for the loci of other inbred strains of mice. It is important to document this diversity to understand the variation within as well between strain models of antibody-mediated diseases, among other things.

IMGT®, the international ImMunoGeneTics information system®[1], is a unique source of knowledge in immunogenetics and immunoinformatics and is recognised as the international reference. IMGT® engages in the precise and detailed characterisation of the IG and TR loci by mouse strain according to IMGT® standards to establish their genomic repertoires and allow their comparison. The work carried out during thisthesis aims to design and develop and/or adapt high-performance software tools and a methodology which implement the standards and carry out the annotation of the loci IG and TR of the mouse strains with a "Gold standard" quality (equivalent to the manual annotation). This will allow enrichment of IMGT® databases and implementation of strain-specific research and analysis in IMGT® software tools.

[1]http://www.imgt.org

Acknowledgements

We thank all members of the IMGT® team for their expertise and constant motivation. IMGT® was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037), 5th PCRDT Quality of Life and Management of Living Resources (QLG2 2000 01287), and 6th PCRDT Information Science and Technology (ImmunoGrid, FP6 IST 028069) programmes of the European Union (EU). IMGT® received financial support from the GIS IBiSA, the Agence Nationale de la Recherche (ANR) Labex MabImprove (ANR 10 LABX 53 01), the Région Occitanie Languedoc Roussillon (Grand Plateau Technique pour la Recherche (GPTR), BioCampus Montpellier. IMGT® is currently supported by the Centre National de la Recherche Scientifique (CNRS), the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), the University of Montpellier, and the French Infrastructure Institut Français de Bioinformatique (IFB) ANR 11 INBS 0013. IMGT® is a registered trademark of CNRS. IMGT® is member of the International Medical Informatics Association (IMIA) and a member of the Global Alliance for Genomics and Health (GA4GH). This work was granted access to the High Performance Computing (HPC) resources of Meso@LR and of Centre Informatique National de l'Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) under the allocation 036029 (2010 2021) made by GENCI (Grand Equipement National de Calcul Intensif).

Tran et al. (2021) EMBnet.journal 26(Suppl A), e953
http://dx.doi.org/10.14806/ej.26.A.953

References

1. RS Sellers, CB Clifford, PM Treuting, and Cory Brayton. Immunological variation between inbred laboratory mouse strains: points to consider in phenotyping genetically immunomodiffed mice. Veterinary pathology, **49**(1):32-43, 2012. http://dx.doi.org/10.1177%2F0300985811429314
2. Cynthia L Smith, Judith A Blake, James A Kadin, Joel E Richardson, Carol J Bult, and Mouse Genome Database Group. Mouse genome database (mgd)-2018: knowledgebase for the laboratory mouse. Nucleic acids research, **46**(D1):D836{D842, 2018. http://dx.doi.org/10.1093/nar/gkx1006
3. Corey T Watson, Justin T Kos, William S Gibson, Leah Newman, Gintaras Deikus, Christian E Busse, Melissa L Smith, Katherine JL Jackson, and Andrew M Collins. A comparison of immunoglobulin ighv, ighd and ighj genes in wild-derived and classical inbred mouse strains. Immunology and cell biology, **97**(10):888-901, 2019. http://dx.doi.org/10.1111/imcb.12288
4. Marie-Paule Lefranc, Véronique Giudicelli, Patrice Duroux, Joumana Jabado-Michaloud, Géraldine Folch, Safa Aouinti, Emilie Carillon, Hugo Duvergey, Amélie Houles, Typhaine Paysan-Lafosse, *et al.* Imgt®, the international immunogenetics information system® 25 years on. Nucleic acids research, **43**(D1):D413–D422,2015 http://dx.doi.org/10.1093/nar/gku1056

EMBnet.journal
Bioinformatics in Action

Elhassani et al. (2021) EMBnet.journal 26(Suppl A), e957
http://dx.doi.org/10.14806/ej.26.A.957

Oral Presentations

Predicting L-PART1 exon using deep learning

Merouane Elazami Elhassani[1,2]✉, Loic Maisonnasse[2], Antoine Olgiati[2], Jérôme Rey[2], Veronique Giudicelli[1], Patrice Duroux[1], Sofia Kossida[1]

[1]IMGT®, The International ImMunoGeneTics Information System®, Institute of Human Genetics (IGH), Scientific Research National Center (CNRS), University of Montpellier, Montpellier, France

[2]ATOS Montpellier, immeuble Archimède, Montpellier, France

Competing interests: MEE none; LM none; AO none; JR none; VG none; PD none; SK none

Identifying and annotating Immunoglobulins (IG), T cell receptors (TR) genes of jawed vertebrate species effectively and precisely is still an arduous task, essentially due to the continuous avalanche of large genomic sequences produced by NGS technologies. In this study, to predict the L-PART1 exon (the first exon of IG and TR variable V-GENE), different Deep Neural Network-based models (DNN, CNN, RNN, CNN-RNN) were trained in a supervised manner, which automatically learn features from annotated IG and TR genes. Those models will then be used to predict the L-PART1 exon within newly given sequences. Correct detection of this component would dramatically increase the chances to find the subsequent components constituting the V-GENE of the IG and TR. Sequence data were extracted from IMGT/LIGM-DB, the IMGT® (Lefranc *et al.*, 2015) nucleotides database: around 2756 L-PART1 of 354 species tagged as positive samples, negative samples were generated with different noise strategies. The final training datasets were formed by shuffling a portion of positive and negative samples. The majority of the trained models showed promising results. However, further studies are needed to investigate the possibility of predicting and locating all the components within the V-GENE of IG and TR.

Acknowledgements

We thank all members of the IMGT® team for their expertise and constant motivation. IMGT® was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037), 5th PCRDT Quality of Life and Management of Living Resources (QLG2 2000 01287), and 6th PCRDT Information Science and Technology (ImmunoGrid, FP6 IST 028069) programmes of the European Union (EU). IMGT® received financial support from the GIS IBiSA, the Agence Nationale de la Recherche (ANR) Labex MabImprove (ANR 10 LABX 53 01), the Région Occitanie Languedoc Roussillon (Grand Plateau Technique pour la Recherche (GPTR), BioCampus Montpellier. IMGT® is currently supported by the Centre National de la Recherche Scientifique (CNRS), the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), the University of Montpellier, and the French Infrastructure Institut Français de Bioinformatique (IFB) ANR 11 INBS 0013. IMGT® is a registered trademark of CNRS. IMGT® is member of the International Medical Informatics Association (IMIA) and a member of the Global Alliance for Genomics and Health (GA4GH). This work was granted access to the High Performance Computing (HPC) resources of Meso@LR and of Centre Informatique National de l'Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) under the allocation 036029 (2010 2021) made by GENCI (Grand Equipement National de Calcul Intensif).

References

1. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G *et al.* (2015) IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res., **43**(Database issue):D413-22. http://dx.doi.org/10.1093/nar/gku1056

EMBnet.journal
Bioinformatics in Action

Loncar-Turukalo et al. (2021) EMBnet.journal 26(Suppl A), e997
http://dx.doi.org/10.14806/ej.26.A.997

Towards the optimisation and standardisation of Machine Learning techniques for human microbiome research: the ML4Microbiome COST Action (CA 18131)

Tatjana Loncar-Turukalo[1], Marcus J. Claesson[2], Randi J. Bertelsen[3], Aldert Zomer[4], Domenica D'Elia[5]✉

[1]Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
[2]APC Microbiome Ireland, University College Cork, Cork, Ireland
[3]University of Bergen, Bergen, Norway
[4]Utrecht University, Utrecht, Netherlands
[5]Institute for Biomedical Technologies, National Research Council, Bari, Italy
Competing interests: TLT none; MjC none; RJB none; AZ none; DD none

The analysis of data generated by metagenome projects in different human body sites has unveiled relevant differences in the microbiome in health and disease. For example, it has been demonstrated that the gut microbiome is crucial in many intestinal and non-intestinal diseases or pathophysiological conditions such as obesity, diabetes, development and functionality of the immune system, cardiovascular diseases, etc. (Lynch and Pedersen, 2016).

The ML4Microbiome COST Action[1] (CA18131) is a network of bioinformaticians, computer scientists and biologists, from 34 different European countries, working together to evaluate and optimise application of machine learning (ML) algorithms on microbiome data and develop standardised data processing pipelines for the analysis and interpretation of microbiome sequencing data. The need to establish this kind of network is related to the complexity of metagenomics data. Microbiota are dynamic ecosystems with active host regulation (Schnorr, 2018). Metagenome data and their influencing factors are complex to analyse and interpret. They are indeed inherently convoluted, noisy and highly variable. Due to the data's compositional nature, non-standard statistical methodologies and ML methods are required to unlock its clinical and scientific potential. While a range of statistical modelling and ML methods are now available, sub-optimal implementation often leads to errors, over-fitting and misleading results due to a lack of suitable analytical practices and ML expertise in the microbiome community. For these reasons, the field requires innovative approaches, specifically adapted to the properties of microbiome data and training to create ML expertise in the microbiome scientific community.

The ML4Microbiome working plan relays on the coordinated and integrated efforts of four working groups (WGs). WG1 aims to evaluate and constantly update the state-of-the-art of ML technologies and methods applied in the field to define priority areas of intervention. WG2 is committed to establishing benchmark datasets for testing ML methods. In particular, based on available public and private data, WG2 selects the data types and creates public benchmark repositories to propose a DREAM Challenge to foster ML analysis of microbiome data. WG3 applies information and data made available by WG1 and 2, to optimise and standardise the use of ML existing methods to microbiome data, also investigating automation opportunities (*i.e.*, pipelines). Finally, WG4 is devoted to disseminating results of the Action, organising training courses and maintaining the website, newsletters and social media.

At this conference, we will present the ML4Microbiome aims and working plans focusing on upcoming training activities and COST tools the Action makes available to the scientific community, such as the possibility to apply for Short Term Scientific Missions[2]. STSMs are exchange visits aimed at supporting individual mobility, strengthening existing networks and fostering collaboration between researchers. We also provide the possibility to apply for grants dedicated to Inclusiveness Target Countries[3] (ITC). ITC grants support early carrier investigators (ECI) and PhD researchers from ITC to present their work at international conferences on the COST Action topic.

Availability

Information on how to join and collaborate is available at the ML4Microbiome website: https://www.ml4microbiome.eu/

[1]https://www.cost.eu/actions/CA18131

[2]https://www.ml4microbiome.eu/stsm/
[3]https://whocaresineurope.eu/inclusiveness-target-countries-itc

Acknowledgements

ML4Microbiome is funded by COST (European Cooperation in Science and Technology) under the Grant Agreement CA18131. www.cost.eu.
The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

References

1. Lynch SV, Pedersen O (2016) The human intestinal microbiome in health and disease. N. Engl. J. Med. **375**:2369–2379. http://dx.doi.org/10.1056/NEJMra1600266
2. Schnorr SL (2018) Meanings, measurements, and musings on the significance of patterns in human microbiome variation. Current Opinion in Genetics & Development **53**:43-52. http://dx.doi.org/10.1016/j.gde.2018.06.014

EMBnet.journal
Bioinformatics in Action

Coornaert et al. (2021) EMBnet.journal 26(Suppl A), e1003
http://dx.doi.org/10.14806/ej.26.A.1003

An ERASMUS student exchange involving nodes from Belgium, Sweden, Colombia, Slovakia and Greece was organized during 2020

David R.A. Coornaert✉, **Cyril Radermecker**, **Ahmed Kanfoud**

Haute Ecole en Hainaut, Mons, Belgium

Competing interests: DRAC none, CR none; AK none

Haute Ecole en Hainaut (HEH) is a university college in the Walloon region in Belgium. It is constituted of four different departments: Law & Economy, Education, Social Sciences and, in concern here, "Science and Technology". The latest department organises four professionalising bachelor levels (3 years) and six engineering levels (5 years).

In September 2008, the HEH inaugurated a bachelor level in biotechnology, with Bioinformatics as finality. The first graduated students emerged in June 2011. During their last year, students legally have to accomplish a 13 weeks professional internship within companies (industrial or academic) as long as a personal work known as "bachelor thesis".

Soon, difficulties arose as to find internship hosting companies in such a specific field in Belgium. The idea emerged then to contact SLU's renowned former EMBnet president, Professor Erik Bongcam-Rudloff, to host two Erasmus-wannabes students. We agreed on a mixture project, where the subject of work given to the students would partly serve as an internship (view, observe, imitate) and partly as a bachelor thesis (build, dig, develop).

The initiative that started in 2012 in Sweden is a complete reciprocal success. Many students followed up to Sweden and, thanks to Erik's recommendations, to Italy and Finland.

The year 2019-2020 saw six students leave Belgium toward EMBnet nodes in Sweden, Greece, Slovakia and Colombia.

During the EMBNet AGM 2020, it climaxed with a short (but intimidating) presentation of their work in Greece by two of these students.

We're very proud that, although initially HEH's bachelor in Bioinformatics, two former Erasmus students recently got their PhD (Axel Thieffry and Hadrien Gourlé), and that a third one is on the way (Renaud Vandamme).

This year, the HEH inaugurated an engineering degree in "Life Data Technology", whose first graduations are expected in June 2024. We're hence very demanding of future collaborations for these new students.

EMBnet.journal
Bioinformatics in Action

ISSN 1023-4144

DEAR READER,

If you have any comments or suggestions regarding this journal we would be very glad to hear from you. If you have a tip you feel we can publish then please let us know. Before submitting your contribution read the "Authors guidelines"[1] and send your manuscript and supplementary files using our on-line submission system[2].

Past issues are available as PDF files from the web archive[3].

Visit EMBnet website for more information: www.journal.embnet.org

EMBNET.JOURNAL EXECUTIVE EDITORIAL BOARD

[1]http://journal.embnet.org/index.php/embnetjournal/about/submissions#authorGuidelines
[2]http://journal.embnet.org/index.php/embnetjournal/author/submit
[3]http://journal.embnet.org/index.php/embnetjournal/issue/archive

PUBLISHER

EMBnet Stichting p/a
CMBI Radboud University
Nijmegen Medical Centre
6581 GB Nijmegen
Th e Netherlands

Email: erik.bongcam@slu.se
Tel: +46-18-67 21 21