EMBnet.journal
Bioinformatics in Action

SOILomics; where Microbiome Genetics meets Precision Agriculture

Ten Simples Rules on How to Organise a Bioinformatics Hackathon

Exosomics

AND MORE...

26
2022

EMBnet.journal
Bioinformatics in Action

Contents

Editorial 2

Letters to Editor

SOILomics; where Microbiome Genetics meets Precision Agriculture
Dimitrios Vlachakis, Aspasia Efthimiadou 3

Technical Notes

KEGG2Net: Deducing gene interaction networks and acyclic graphs from KEGG pathways
Sree K. Chanumolu, Mustafa Albahrani, Handan Can Hasan H. Otu 5

Ten Simples Rules on How to Organise a Bioinformatics Hackathon
Susanne Hollmann, Babette Regierer, Teresa K Attwood, Andreas Gisel, Jacques Van Helden, Gregoire Rossier, Paul J Kersey, Eija Korpelainen, Gert Vriend, Erik Bongcam-Rudloff 9

Reviews

Exosomics
Thanasis Mitsis, Katerina Pierouli, Kalliopi Io Diakou, Eleni Papakonstantinou, Flora Bacopoulou, George P. Chrousos, Dimitrios Vlachakis 16

Protein Spotlight

Protein Spotlight 223 24

Protein Spotlight 225 26

Protein Spotlight 227 28

Protein Spotlight 230 30

Cover picture: Red Passion flower (*Passiflora racemosa*), Uppsala, Sweden, 2019, © Erik Bongcam-Rudloff

Editorial

Volume 26 of EMBnet.journal was produced during the difficult conditions created by the COVID19 pandemic that surprised the world at the beginning of 2020.
Most researchers had to change their usual way of working and adapt to more virtual digitalised. Conferences and meetings were held using tools that were unknown to most. At the end of 2020, online meetings systems were used by all in society and not anymore just by academics.

EMBnet had to organise its Annual General Meeting (AGM) virtually using online meeting platforms and tools. 190 researchers from 38 different countries participated in the event. The journal produced a Supplement to document the successful event. Supplement A "EMBnet Conference 2020: Bioinformatics Approaches to Precision Research".

Despite the pandemics, EMBnet.journal was also able to attract authors to submit articles and produce a special issue on "Biostatistics and Mathematical Modelling in Bioinformatics" that gathered many highly interesting articles.

The year 2020 will go down in history as this was the year we were forced to transform our work habits and had to adapt to unprecedented work conditions. It has undoubtedly been a different year for everyone and EMBnet.journal's Editorial board would like to thank all people that worked so hard during the year and helped the journal team to produce this Volume.

Erik Bongcam-Rudloff
Editor-in-Chief
erik.bongcam@slu.se

http://dx.doi.org/10.14806/ej.26.0.1019

EMBnet.journal
Bioinformatics in Action

Vlachakis et al. (2021) EMBnet.journal 26, e940
http://dx.doi.org/10.14806/ej.26.0.940

SOILomics; where Microbiome Genetics meets Precision Agriculture

Dimitrios Vlachakis[1,2,3], **Aspasia Efthimiadou**[4]✉

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

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

[3]Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, Strand, London WC2R 2LS, United Kingdom

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

Competing interests: DV none; AE none

Abstract

Advances in genetics, soil biochemistry and microbiome analysis are opening up a new era in Precision Agriculture. In this direction, new techniques bring groundbreaking changes in land management practices through direct or indirect management of soil microbial communities. There is huge demand for the protection and enhancement of soil health and climate change resilience of crops. The increase in population, food consumption and fast approaching climate change pose a new threat to mankind that only by being proactive and highly prepared to deploy all novel and innovative stratagems in state-of-the-art soil microbiome precision agriculture can be avoided.

Introduction

Agriculture is one of the most exposed sectors to increased climate variability. Soil degradation, loss of biodiversity, increased air and water pollution affect the viability of the crop and the agricultural productivity. Nowadays, new and pioneering strategies utilizing the genetic flexibility and biochemical power of the microbes are being formulated. Those approaches attempt to coordinate and thus transform the microbial content of the field into a natural molecular plant to optimize and shield the crop and at the same time protect the environment mainly by reducing chemical interventions. In this direction, multidimensional approaches combining science and know-how from different fields aim for the capture of soil microbiome using homogeneous technologies, the traditional recording of soil and atmospheric conditions, and the creation of intelligent systems and algorithms for the management of resources and inputs based on microbiome analysis within a sophisticated Precision Agriculture System. The methodology that is being developed is primarily based on a very economical instrument, the lysimeter, which can be placed in the field easily and at low cost, and through an initial dense sampling of soil and organic technologies, the microbial profile of the crop will be attributed. The creation and establishment of a decision support system will minimize the required quantity of samples and at a lower cost will ensure the maximum reliability of the method. At the same time, the correlation of the results of the microbial analysis with the corresponding conditions and yields of crops through machine learning techniques will lead us to a sustainable approach of the agricultural systems.

Furthermore, major changes in the environment and socio-economic factors, combined with the steady increase in the earth's population, are impeding the food industry. The importance of soil microbiome in establishing the stability of the ecosystem along with its high involvement in sustainable agriculture is recognized in the past decade. Soil microbiome affects crop growth and soil functions, especially biological soil activity and fertility. Soil microorganisms contribute to soil health by cycling nutrients (such as nitrogen and phosphate) essential for plant growth and global biogeochemical cycles, they improve soil structure by increasing organic matter content, enhance the resilience of the plants by responding to environmental stresses, and confer disease resistance to crops by out-competing pathogenic microbes. Exploiting and modulating the metabolic capabilities of the soil microbiome by

Article history
Received: 14 August 2020
Accepted: 05 October 2020
Published: 20 October 2021

Vlachakis et al. (2021) EMBnet.journal 26, e940
http://dx.doi.org/10.14806/ej.26.0.940

genetically manipulating the beneficial native microbes that promote plant health and quality are ultimate practices for achieving stable yield and reduced impact on the agroecosystem. On a different note, microbiome members may also be associated with diseases and pathogenic effects on plants. The application of pesticides and agrochemicals against phytopathogens in intensive farming leads to impairment of soil functions and long-term crop yield losses, along with negative effects on beneficial soil microbiota. Thus, the management of the microbial communities and the study of their diversity, connectivity and impact on the soil health is a key component to maintain agricultural productivity and protect the environment. Climate change favors the spread of new or migratory microorganisms and alters the microbiome distribution in the soil, affecting not only its relative abundance, but also its function. Moreover, global warming directly changes respiration rates of soil microorganisms as the processes they mediate are temperature sensitive. Soil microorganisms are mainly responsible for cycling of soil organic carbon and other nutrients, and are key players in climate feedback through the production or consumption of greenhouse gases such as CO_2, CH_4 and N_2O. A better understanding of soil microbiome connectivity, physiology and metabolism and how it is impacted by climate change is required for sustainable soil management.

All in all, the advances in genetics, soil biochemistry and microbiome analysis are opening up a new era in precision agriculture. In this direction new techniques push to bring groundbreaking changes in land management practices through direct or indirect manipulation of soil microbial communities. There is huge demand and great need for the protection and enhancement of soil health and climate change resilience of crops. The increase in population, consumption of foods and fast approaching climate change pose a new threat to mankind that only by being proactive and highly prepared to deploy all novel and innovative stratagems in state-of-the-art soil microbiome precision agriculture can be avoided.

EMBnet.journal
Bioinformatics in Action

Chanumolu et al. (2021) EMBnet.journal 26, e949
http://dx.doi.org/10.14806/ej.26.0.949

Technical Notes

KEGG2Net: Deducing gene interaction networks and acyclic graphs from KEGG pathways

Sree K. Chanumolu[1], **Mustafa Albahrani**[1], **Handan Can**[1], **Hasan H. Otu**[1]✉

[1]Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE, United States
Competing interests: SKC none; MA none; HC none; HHO none

Abstract

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database provides a manual curation of biological pathways that involve genes (or gene products), metabolites, chemical compounds, maps, and other entries. However, most applications and datasets involved in omics are gene or protein-centric requiring pathway representations that include direct and indirect interactions only between genes. Furthermore, special methodologies, such as Bayesian networks, require acyclic representations of graphs. We developed KEGG2Net, a web resource that generates a network involving only the genes represented on a KEGG pathway with all of the direct and indirect gene-gene interactions deduced from the pathway. KEGG2Net offers four different methods to remove cycles from the resulting gene interaction network, converting them into directed acyclic graphs (DAGs). We generated synthetic gene expression data using the gene interaction networks deduced from the KEGG pathways and performed a comparative analysis of different cycle removal methods by testing the fitness of their DAGs to the data and by the number of edges they eliminate. Our results indicate that an ensemble method for cycle removal performs as the best approach to convert the gene interaction networks into DAGs. Resulting gene interaction networks and DAGs are represented in multiple user-friendly formats that can be used in other applications, and as images for quick and easy visualisation. The KEGG2Net web portal converts KEGG maps for any organism into gene-gene interaction networks and corresponding DAGs representing all of the direct and indirect interactions among the genes.

Introduction

The KEGG pathway database provides hundreds of manually curated maps that involve molecular interactions between gene products, compounds, maps, DNA, RNA, and other molecules (Kanehisa and Goto, 2000). The maps are categorised under seven groups, such as "Metabolism" or "Environmental Information Processing," which are stored in proprietary files in XML format, called KGML. There exist numerous approaches that process the KEGG pathway maps, such as KEGGtranslator (Wrzodek *et al.*, 2011), KEGGParser (Arakelyan and Nersisyan, 2013), CyKEGGParser (Nersisyan *et al.*, 2014), KEGGgraph (Zhang and Wiemann, 2009), KEGGconverter (Moutselos *et al.*, 2009), and graphite (Sales *et al.*, 2012) among others (Wrzodek *et al.*, 2013). These approaches convert KGML files to other formats (*e.g.*, SBML, BioPAX) to be used in applications for data visualisation (*e.g.*, Cytoscape) or graph-theoretic analysis (*e.g.*, MATLAB®, Bioconductor).

Although these approaches are extremely useful, they do not provide a representation that only includes genes deduced from the KEGG pathways considering all of the direct and indirect gene interactions. However, when analysing experimental datasets that only involve the genes (or gene products) in the context of KEGG pathways (*e.g.*, transcriptomic or proteomic data), an interaction network that only involves these molecules is required. Among the existing tools, KEGGgraph provides a *"genesOnly"* parameter that results in a gene-oriented graph; but that approach only deduces the direct gene interactions provided in the maps. graphite also represents gene-only networks, but it does not take into account all of the compounds between the genes to obtain the exhaustive set of indirect interactions – some compounds based on their identity or localisation are ignored. Indeed, there is a need for an approach that recovers all of the direct and indirect gene-gene interactions from a KEGG pathway that can be used in

Article history
Received: 14 September 2020
Accepted: 13 December 2020
Published: 05 March 2021

Chanumolu et al. (2021) EMBnet.journal 26, e949
http://dx.doi.org/10.14806/ej.26.0.949

downstream analysis involving data coming only from these molecules.

KEGG pathways include cycles that may be problematic in analysis approaches, such as Bayesian networks (BNs), which use directed acyclic graphs (DAGs) (Friedman *et al.*, 2000; Isci *et al.*, 2014; Isci *et al.*, 2011; Korucuoglu *et al.*, 2014; Liu *et al.*, 2016). None of the existing approaches that process KEGG pathways provide DAGs as their output. Furthermore, there has been no study that compares different cycle removal methods in the context of the fitness of biological data to the resulting DAGs.

In light of these observations and perceived needs, we developed KEGG2Net, a web resource that converts KEGG pathways into gene interaction networks involving all of the direct and indirect relations between the genes that can be deduced from the map. KEGG2Net offers four alternative methods to convert the resulting gene interaction networks into DAGs. This paper also provides a comparative assessment of the cycle removal methods via their fitness to the data obtained from the original gene interaction network deduced from KEGG.

Implementation

Given the KGML file for a pathway map, the engine parses the file to obtain an adjacency matrix that represents all of the interactions (relation, reaction, *etc.*) between all of the node types (compound, map, gene, *etc.*) defined in the file. In this graph, if there exists a path between two genes that contain non-gene nodes only, then an indirect relation between the two genes is established by placing an edge between them. Next, nodes that are not genes are removed from the adjacency matrix. This way, the resulting graph represents all of the direct and indirect interactions between the genes that can be deduced from the KEGG pathway map.

The resulting gene network may contain cycles that are removed using four different methods: a depth-first search (DFS) (Suominen and Mader, 2014), a greedy local heuristic to the minimum feedback arc set (MFAS) problem (Eades *et al.*, 1993), and two graph-hierarchy-based methods where the hierarchy is inferred either through PageRank (PR) (Page *et al.*, 1999) or an ensemble method (EN) (Sun *et al.*, 2017) based on the TrueSkillTM (Herbrich *et al.*, 2007) and social agony (Gupte *et al.*, 2011) metrics. The DFS-based approaches use fast, simple heuristics to remove back edges, MAFS-based approaches try to minimise the number of edges removed, and hierarchy-based methods define a hierarchy in the graph first and then devise an edge removal strategy that prioritises the maintenance of the defined hierarchy as much as possible.

The input to KEGG2Net is the KGML files for the pathways that belong to the organism selected by the user. The output of KEGG2Net consists of the gene interaction networks deduced from the pathways and four DAGs per network where the cycles are removed by the aforementioned four algorithms. The networks and DAGs are represented as adjacency matrices and simple interaction files (a.k.a. SIF or .sif format) for use in the downstream analysis by other software and for visualisation purposes.

In order to compare the accuracy of different cycle removal methods and to provide a sample output using graph images for visualisation purposes, we applied the KEGG2Net approach on the 335 available human KEGG pathways. The KEGG2Net workflow adopted in this paper is shown in Figure 1.

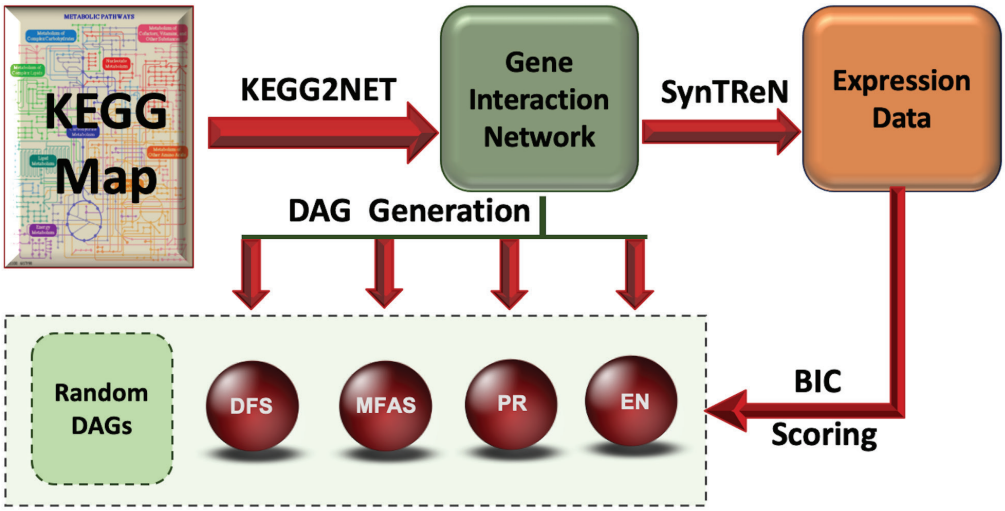


Figure 1. KEGG2Net workflow and directed acyclic graph (DAG) generation.
Each KEGG pathway is converted into a gene interaction network where only the genes in the pathway are represented, and all direct and indirect interactions among the genes are preserved. For each network, four DAGs (using four alternative cycle removal methods depth-first search (DFS), minimum feedback arc set (MFAS), PageRank (PR), and ensemble (EN)) and 1,000 random DAGs for each of the four DAGs are generated. The random DAGs follow the node and edge statistics of their corresponding DAG. For each gene interaction network, synthetic gene expression data is generated using SynTRen and the fitness of the four DAGs (and the corresponding 4x1,000 = 4,000 random DAGs) is assessed using Bayesian Information Criterion (BIC) scoring.

Chanumolu et al. (2021) EMBnet.journal 26, e949
http://dx.doi.org/10.14806/ej.26.0.949

Out of the 335 pathways, we considered only the networks with six or more non-isolated nodes, which left us with 280 networks. Of these networks, 150 did not have any cycles. For each of the remaining 130 that were cyclic, a synthetic gene expression data fitting the graph topology was generated using SynTReN (v. 1.2) (Van den Bulcke *et al.*, 2006). The expression data was processed to be used by BN scoring methods as previously described (Isci *et al.*, 2011). The four cycle removal methods were applied to the 130 networks with cycles generating four DAGs per network, which were scored based on the Bayesian Information Criterion (BIC) in the *bnlearn* R package (Scutari, 2010) using the processed synthetic expression data. For each of the four DAGs per network, we generated 1,000 random DAGs (with the same number of edges and nodes as the original DAG) and obtained their BIC scores based on the same processed synthetic expression data to assess the goodness of the original DAG's score.

Results

Our web portal provides the gene interaction networks obtained for all of the 335 human KEGG pathways. For the networks that have cycles, we also list the DAGs obtained using the four methods. The gene interaction networks and the DAGs obtained from them are represented by adjacency matrices, SIF format files, and graph images. KEGG2Net can be used for all of the organisms listed in KEGG where the user can download the relevant network and DAG files via our web portal.

One direct way to assess the performance of different cycle removal methods is to compare the number of edges removed by each method. However, this approach does not infer the degree to which the topology and the dependency structure among the nodes of the network are preserved in the DAGs. For this purpose, we first generated synthetic gene expression data that follow the regulatory dynamics explained by the gene interaction network deduced by KEGG2Net. We then scored each of the four DAGs with this expression data using BIC scoring, where a higher score indicated a better fit. Finally, for each of the four DAGs that result from the given network, we generated 1,000 random DAGs (i.e., 4,000 random DAGs per network) where the random DAGs had the same number of nodes and links as the DAG they were associated with. This exercise was repeated for all of the 130 networks, and the DAG statistics were compared.

The complete set of results for our simulations are given in the Supplementary Data[1], which involves the 130 KEGG2Net gene interaction networks with six or more non-isolated nodes and a cycle. We provide the original pathway's ID, name, number of edges, number of nodes, the number of edges removed by the four cycle removal methods, the rank of each method for each pathway based on the edges removed, and the rank of the score for each method (among themselves and their 1,000 random DAGs).

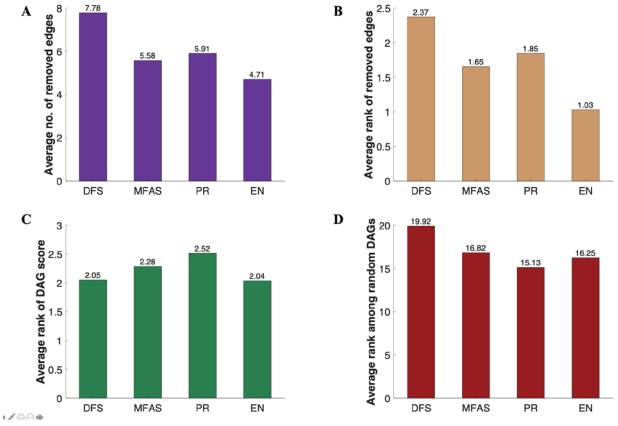


Figure 2. Directed acyclic graph (DAG) statistics. Based on the four cycle removal methods, depth-first search (DFS), minimum feedback arc set (MFAS), PageRank (PR), and ensemble (EN), applied on 130 gene interaction networks deduced from pathways using KEGG2Net, **A** average number of edges removed; **B** average rank of the method (per network) based on the number of edges removed (ascending); **C** average rank of the method's DAG score (per network) among the four methods (descending); **D** average rank of the method's DAG score (per network) among the 1,000 random DAGs that has the same numbers of edges and nodes as the DAG (descending). For parts **C** and **D**, Bayesian Information Criterion (BIC) scoring is used to assess the fitness of the DAGs to the synthetic data generated based on the gene interaction network (obtained by KEGG2Net).

These results, summarised in Figure 2, showed that on average, the EN method required the least number of edges to be removed and provided the DAG with the highest BIC score.

On average, the EN method removed 4.71 edges per pathway; and it was the method that required the least number of edges to be removed in 127 out of 130 networks. The EN method accomplished an average rank of 1.03 in all networks, where 1 represented the rank that removed the minimum number of edges.

The average rank of the EN score among the four DAGs was also the highest, where it attained an average rank of 2.04. In other words, on average, the EN DAG provided the best topology that fit the synthetic expression data. The only category where the EN method was outperformed was its average rank among the 1,000 random DAGs. On average, 16.25 random DAGs for a given network performed better than the EN DAG, whereas this number was 15.13 for the PR DAG, the only method that beat the EN approach in this category. Given the comprehensive evaluation summarised in Figure 2, we recommend EN as the method of choice for DAG generation.

Discussion

In this work, we provide a web resource that converts KEGG pathways into gene interaction networks

[1]http://journal.embnet.org/index.php/embnetjournal/article/downloadSuppFile/949/949_supp_1

representing all of the direct and indirect interactions between the genes. We also provide four alternative ways of converting the resulting graph into a DAG. Our results showed that the EN method finds the best DAG that explains the underlying hierarchy and dependency structure defined in the interaction network with the minimum number of edge removals. Our web portal lists the graph structure and the four DAGs for each of the KEGG human pathways. The KEGG2Net web resource can be used to obtain the networks and DAGs for any organism listed in the KEGG database.

Key Points

- Deduces only the gene-gene interactions from a KEGG pathway considering all direct and indirect interactions.
- Provides networks that are ready to be applied to transcriptomic or proteomic data for system-level analysis.
- Using multiple alternative methods, converts networks to directed acyclic graphs (DAGs) that can be used in methodologies such as Bayesian networks, which require DAGs.
- Generates multiple output formats that can be directly used in different network visualisation or analysis software.
- Reports a comparative analysis of cycle removal methods for biological pathways.

Funding

The research reported in this publication was supported by the National Library of Medicine (NLM) of the National Institutes of Health (NIH) under award number **R21LM012759**. The funding body was not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

References

1. Arakelyan A and Nersisyan L (2013) KEGGParser: parsing and editing KEGG pathway maps in Matlab. Bioinformatics **29**(4): 518-519 http://dx.doi.org/10.1093/bioinformatics/bts730
2. Eades P, Lin X and Smyth WF (1993) A fast and effective heuristic for the feedback arc set problem. Inform Process Lett **47**(6): 319–323 http://dx.doi.org/10.1016/0020-0190(93)90079-O
3. Friedman N, Linial M, Nachman I and Pe'er D (2000) Using Bayesian networks to analyze expression data. J Comput Biol **7**(3-4): 601-620 http://dx.doi.org/10.1089/106652700750050961
4. Gupte M, Shankar P, Li J, Muthukrishnan S and Iftode L (2011) Finding Hierarchy in Directed Online Social Networks. Proceedings of the 20th International Conference on World Wide Web (WWW '11): 557–566 http://dx.doi.org/10.1145/1963405.1963484
5. Herbrich R, Minka T and Graepel T (2007) TrueSkillTM: A Bayesian Skill Rating System. In: Advances in Neural Information Processing Systems (NIPS), Scholkopf P.B., Platt J.C., Hoffman T. (eds.) pp. 569–576.
6. Isci S, Dogan H, Ozturk C and Otu HH (2014) Bayesian network prior: network analysis of biological data using external knowledge. Bioinformatics **30**(6): 860-867 http://dx.doi.org/10.1093/bioinformatics/btt643
7. Isci S, Ozturk C, Jones J and Otu HH (2011) Pathway analysis of high-throughput biological data within a Bayesian network framework. Bioinformatics **27**(12): 1667-1674 http://dx.doi.org/10.1093/bioinformatics/btr269
8. Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28**(1): 27-30 http://dx.doi.org/10.1093/nar/28.1.27
9. Korucuoglu M, Isci S, Ozgur A and Otu HH (2014) Bayesian pathway analysis of cancer microarray data. Plos One **9**(7): e102803 http://dx.doi.org/10.1371/journal.pone.0102803
10. Liu F, Zhang SW, Guo WF, Wei ZG and Chen L (2016) Inference of Gene Regulatory Network Based on Local Bayesian Networks. PLoS Comput Biol **12**(8): e1005024 http://dx.doi.org/10.1371/journal.pcbi.1005024
11. Moutselos K, Kanaris I, Chatziioannou A, Maglogiannis I and Kolisis FN (2009) KEGGconverter: a tool for the in-silico modelling of metabolic networks of the KEGG Pathways database. BMC Bioinformatics **10**(324) http://dx.doi.org/10.1186/1471-2105-10-324
12. Nersisyan L, Samsonyan R and Arakelyan A (2014) CyKEGGParser: tailoring KEGG pathways to fit into systems biology analysis workflows. F1000Res **3**(145) http://dx.doi.org/10.12688/f1000research.4410.2
13. Page L, Brin S, Motwani R and Winograd T, (1999). "The PageRank citation ranking: Bringing order to the web", Technical Report Stanford InfoLab.
14. Sales G, Calura E, Cavalieri D and Romualdi C (2012) graphite - a Bioconductor package to convert pathway topology to gene network. BMC Bioinformatics **13**(20) http://dx.doi.org/10.1186/1471-2105-13-20
15. Scutari M (2010) Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software **35**(3): 1-22 http://dx.doi.org/10.18637/jss.v035.i03
16. Sun J, Ajwani D, Nicholson PK, Sala A and Parthasarathy S (2017) Breaking Cycles In Noisy Hierarchies. Proceedings of the 2017 ACM on Web Science Conference: 151-160 http://dx.doi.org/10.1145/3091478.3091495
17. Suominen O and Mader C (2014) Assessing and Improving the Quality of SKOS Vocabularies. Journal on Data Semantics **3**(1): 47–73 http://dx.doi.org/10.1007/s13740-013-0026-0
18. Van den Bulcke T, Van Leemput K, Naudts B, van Remortel P, Ma H *et al.* (2006) SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. BMC Bioinformatics **7**(43) http://dx.doi.org/10.1186/1471-2105-7-43
19. Wrzodek C, Buchel F, Ruff M, Drager A and Zell A (2013) Precise generation of systems biology models from KEGG pathways. BMC Syst Biol **7**(15) http://dx.doi.org/10.1186/1752-0509-7-15
20. Wrzodek C, Drager A and Zell A (2011) KEGGtranslator: visualizing and converting the KEGG PATHWAY database to various formats. Bioinformatics **27**(16): 2314-2315 http://dx.doi.org/10.1093/bioinformatics/btr377
21. Zhang JD and Wiemann S (2009) KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics **25**(11): 1470-1471 http://dx.doi.org/10.1093/bioinformatics/btp167

EMBnet.journal
Bioinformatics in Action

Hollmann et al. (2021) EMBnet.journal 26, e983
http://dx.doi.org/10.14806/ej.26.0.983

Ten Simples Rules on How to Organise a Bioinformatics Hackathon

Susanne Hollmann[1,2], Babette Regierer[2], Teresa K Attwood[3], Andreas Gisel[4,5], Jacques Van Helden[6], Gregoire Rossier[7], Paul J Kersey[8], Eija Korpelainen[9], Gert Vriend[10], Erik Bongcam-Rudloff[11] ✉

[1]Focus Area Plant Genomics and Systems Biology, Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany
[2]SB Science Management UG (haftungsbeschränkt), Berlin, Germany
[3]The University of Manchester, Manchester, United Kingdom
[4]CNR Institute for Biomedical Technologies, Bari, Italy
[5]International Institute for Tropical Agriculture, Ibadan, Nigeria
[6]Department of Theory and Approaches to Genome, Institut Français de Bioinformatique, Évry, France
[7]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[8]EMBL-European Bioinformatics Institute, Hinxton, and Bioinformatics and Genomics, Royal Botanic Gardens, Surrey, United Kingdom
[9]IT Center for Science, Espoo, Finland
[10]Centre for Molecular and Biomolecular Informatics, Radboudumc, Nijmegen, Netherlands
[11]Swedish University of Agricultural Sciences, Uppsala, Sweden
Competing interests: SH none; BR none; TKA none; AG none; JVH none; GR none; PJK none; EK none; GV none; EBR none

Abstract

The completion of the human genome sequence triggered worldwide efforts to unravel the secrets hidden in its deceptively simple code. Numerous bioinformatics projects were undertaken to hunt for genes, predict their protein products, function and post-translational modifications, analyse protein-protein interactions, etc. Many novel analytic and predictive computer programmes fully optimised for manipulating human genome sequence data have been developed, whereas considerably less effort has been invested in exploring the many thousands of other available genomes, from unicellular organisms to plants and non-human animals. Nevertheless, a detailed understanding of these organisms can have a significant impact on human health and well-being.

New advances in genome sequencing technologies, bioinformatics, automation, artificial intelligence, etc., enable us to extend the reach of genomic research to all organisms. To this aim gather, develop and implement new bioinformatics solutions (usually in the form of software) is pivotal. A helpful model, often used by the bioinformatics community, is the so-called hackathon. These are events when all stakeholders beyond their disciplines work together creatively to solve a problem. During its runtime, the consortium of the EU-funded project AllBio - Broadening the Bioinformatics Infrastructure to cellular, animal and plant science - conducted many successful hackathons with researchers from different Life Science areas. Based on this experience, in the following, the authors present a step-by-step and standardised workflow explaining how to organise a bioinformatics hackathon to develop software solutions to biological problems.

Introduction

The vast advances in technologies of the past decade enabled researchers to reach genomic research to quit all organisms. Among other large-scale sequencing initiatives for plants, microbes or animals like fish, the Earth Biogenome Project[1] (EBP) was launched to sequence the DNA of all life on Earth within the next decade. This long-term project will lead both to a greater understanding of Earth's biodiversity and responsible stewardship of its resources, tackling the new millennium's most crucial scientific and social challenges. While the focus of the EBP is on collecting genomic data, other initiatives have centred on data

[1]www.earthbiogenome.org

Article history
Received: 02 December 2020
Accepted: 12 May 2021
Published: 19 October 2021

analysis. For example, the EU-funded project AllBio[2] - Broadening the Bioinformatics Infrastructure to unicellular, animal and plant science (FP7 GA 289452) - concentrated on non-human genomes, applying human-genome-derived computational solutions to non-human organisms. This project involved collecting a range of biological problems, the so-called test cases, and the relevant computational solutions. Some of these test cases were worked out in detail during hackathons. A hackathon is a short (1-day to 1-week) event where stakeholders with diverse skills and backgrounds gather to develop and implement solutions (usually in the form of software) to relevant problems. The term hackathon is a composite of the words 'hack' (meaning exploratory programming) and 'marathon' (a common metaphor for long and intensive events). Hackathons are typical in informatics communities but still relatively new to the life sciences. In part, this may be because there are still considerable communication gaps between life- and computational-science researchers. Bioinformatics hackathons or bio-hackathons aim to address such gaps by bringing IT professionals (and interested amateurs) and life science scientists together to communicate and exchange ideas around practical research questions. These type of events can indeed be highly productive for interdisciplinary teams to solve well-defined problems or to accelerate solution provision in a particular area (hackseq Organising Committee 2016 (2017); Friedberg *et al.*, 2015; Poncette *et al.*, 2020; Braune *et al.*, 2021), generate innovations (Lyndon *et al.*, 2018) or serve as educational tools (Silver *et al.*, 2016; Wang *et al.*, 2018). Here, the co-development principle involving the problem providers and developers in the entire process will ensure that a suitable solution is created. Online events might work for the technical solution creation but will most likely lack the lively interdisciplinary interaction.

The short time usually available for bio-hackathons generally allows for the design and implementation of prototype solutions. For the outputs to be helpful, the developed code must have the potential to undergo subsequent development by interested parties. Therefore, all results should be made available via openly accessible platforms to allow researchers to improve the product after the event is terminated.

The philosophy of the AllBio project was to solicit life science scientists to identify topic challenges directly. Around 60 of such test cases were collected via questionnaires and interviews, out of which 15 (encompassing unicellular organisms, plants, and farm animals) were deemed solvable with adaptations to software or workflows initially designed for human-genome data (Bongcam-Rudloff *et al.*, 2019). Eight were subsequently addressed in bio-hackathons (Amar *et al.*, 2014; Gomez-Cabrero *et al.*, 2014; Leung *et al.*, 2015). A problem was considered solvable when:

- a generic question relating to the analysis of a unicellular, animal or plant genome had been well defined;

[2]www.allbioinformatics.eu

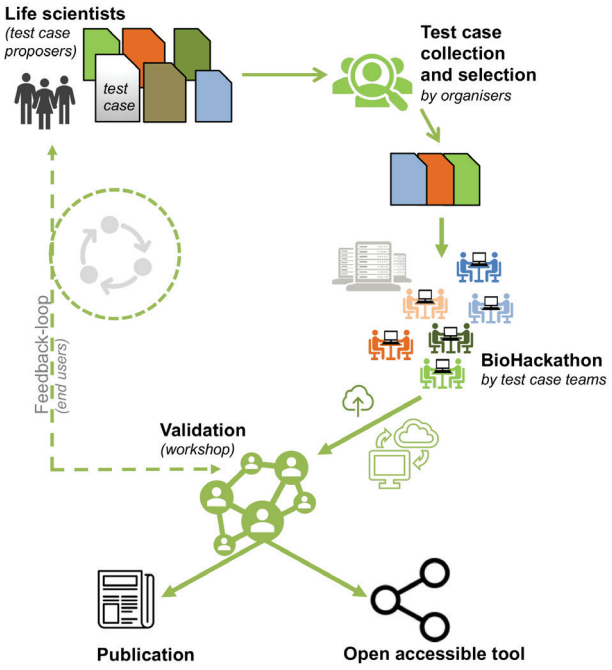


Figure 1. Identification of test cases. AllBio workflow illustrating the fate of test-cases proposed by life science scientists. After initial interviews, the test cases were collected and assessed for their tractability. The bio-hackathon teams comprised the proposer (life science scientist), a leader (bioinformatician), hackers (programmers) and, usually, a local organiser. Where a tool or meta-tool arose from work, it was proposed for testing during a validation workshop. Ultimately, the Team prepared an open-source tool and published or otherwise disseminated the results.

- a community of domain-expert bioscientists and bioinformaticians had been formed; and
- scientific meetings (in vivo or in-silico) had already taken place, and collaborations had begun.

The workflow for AllBio bio-hackathons involved collecting and selecting the test cases, preparing and organising the events, and finally - in the chance of success - publish the results (Fig. 1).

During the AllBio project, a rigorous regime of evaluating past events allowed each bio-hackathon to build on lessons learned from previous ones. This iterative process demonstrated that the events must be well-prepared and long in advance for bio-hackathons to be successful. The biological problems they set out to tackle must be tractable. They must have access to requisite computational infrastructure and sufficient time to complete the necessary tasks. Other essential pre-requisites are efficient leadership, an appropriate mix of skills/expertise, and effective communication strategies. A preparatory phase should precede bio-hackathons to check feasibility and practicability, *e.g.*, can the data be moved around and read? There should also be commitments afterwards to finalise any tools (or outputs), to test and validate them with end-users, and to disseminate the results. Based on the experience gained in AllBio, we present ten rules that we believe are

Hollmann et al. (2021) EMBnet.journal 26, e983
http://dx.doi.org/10.14806/ej.26.0.983

crucial when organising bioinformatics hackathons, or bio-hackathons: these fall into four main categories - the Problem, the Team, the bio-hackathon and the Answer, which are described in detail below. There will, of course, be other important considerations (funding, *etc.*), but we focus here on the practicalities of organising successful bio-hackathons. The presentation of the described process is following the scheme of the Ten Simple Rules series of PLOS[3].

The Ten Rules

The Problem

Rule 1: Understand the biological Problem (s) and select the theme

It might seem self-evident to state that a good starting point is to understand the Problem before trying to address it. But, solving biological problems via hackathons requires a spectrum of understanding that encompasses the biology of the Problem (including in vivo aspects), the nature of the data available, computational requirements and expected output(s), and how all of these can be brought together to implement a viable solution. One of the keys to success is that those responsible for implementing the technical solution(s) must appreciate, at least at some level, the underlying biology. Ultimately, this requires some investment of time to allow them to begin to understand the language of those whose biological problems they are trying to solve. One way to help achieve this, even for small events, might be to run a small cycle of webinars before the event to give participants more information about the theme. This is likely to facilitate team building and may also provide opportunities to come up with new ideas for possible approaches and solutions.

Rule 2: Ensure that the Problem is tractable

Bio-hackathons are driven by practical research questions, but not all biological problems are amenable to solution by hackathons. An early step in setting up any such event should therefore be to estimate whether the size of the Problem is compatible with the hackathon format. For example, while de novo software design is generally not the goal of hackathons (design of new algorithms tends to require more than just a few days), proof-of-concept implementations can fit the format quite well. Ideally, therefore, the necessary software components must already exist so that bio-hackathon sessions can readily combine them into bespoke workflows. Ideally, workflows should not contain any single point of failure. Notably, both the biological datasets and the software components must be available without restrictions.

The Team

Rule 3: Put together the right Team with carefully assigned roles

Start building the Team as soon as possible. Ideally, establish the core group two months before the event. Think about life- and computational-science colleagues and students who have the requisite skills and knowledge in the Problem area. Generate a checklist with the minimal requirements needed to ensure that the complete project can be implemented during the event. This will form the basis for participant selection. If necessary, promote the bio-hackathon widely (*e.g.*, using social media), providing as much information about the event as possible (including when, where, what, how, fees - if needed - and registration forms). Some incentives might be helpful to engage bio-hackathon participants, such as cooperation with university groups that might be willing to give credit points for participation or formulating problems whose solutions are suitable for academic publication and crediting those participants as authors.

Biohackathon teams are generally most effective when they comprise no more than eight to ten participants. In general, they should include a proposer or biological Problem owner, typically a life science scientist, whose needs will drive the event. A leader, usually a bioinformatician. The hackers, bioinformaticians and computer scientists and, ideally, an overall organiser/coordinator. Those with computational skills should include at least one IT professional or bioinformatician and programmers with experience in scripting, workflow design, use of ontologies, evaluation of data quality, and so on. These professionals must be able to communicate effectively with the leader and remain focused on the primary objective.

The bio-hackathon leader is responsible for monitoring and guiding the workflow during the event. The organiser must take responsibility for the overall coordination of the event, maintaining good communication within the Team (rule 4), orchestrating the validation (rule 9) and dissemination (rule 10) activities. The organiser must be local to the venue of the bio-hackathon and will be responsible for many mundane practical tasks: reserving the venue, testing bandwidth in the meeting room before the actual hackathon, providing travel instructions, communicating with the compute provider, selecting the participants and dealing with subsistence/refreshment issues, *etc.* One person may assume several roles, but it is vital that each partner knows his/her role and that all roles are maintained before, during and after the hackathon itself. To facilitate discussion and assignment of tasks as the project progresses, we suggest adopting a convenient communication platform, *e.g.*, Trello[4], Slack[5] or

[3]https://collections.plos.org/collection/ten-simple-rules/

[4]https://trello.com
[5]https://slack.com

comparable platform such as ownCloud[6], GoogleDrive[7] or Dropbox[8].

Rule 4: Communicate effectively and establish the ground rules

Communication – before, during and after hackathons – is key. The value of good communication, and the impact of not getting it right, is hard to over-emphasise. Bio-hackathons include partners from different disciplines who tend to speak very different languages. If a bio-hackathon is to be maximally productive, it is critical to take time, early on, to identify and resolve potential language barriers. Frequent conversations before the bio-hackathon (in person if possible, or electronically if not) are essential to understand, define and refine the biological question, identify and shape the overall analytical approach, and thence to build ownership of the tasks. As the technical partners assimilate the nature of the biological Problem and the biological partners begin to appreciate the heart of the technical challenges, the Team's purpose, focus and cohesiveness will mature.

If multiple projects are being tackled in one bio-hackathon, ensure that all requirements have been established beforehand, including the process of team-building, the time-frame available for each Problem (equal conditions for every Team, so that each has the same relative chance of success), and the rules for allowing participants to move between teams.

Rule 5: Prepare the ground-work well in advance

Bio-hackathons are generally time-limited; good preparatory work is therefore essential. A crucial part of the preparation is to test the necessary software and hardware before the event to prevent problems that could reduce the time available for hands-on work. Any heavy computational tasks should be pre-computed to allow participants to hit the ground running with real data. Bio-hackathon leaders must, therefore, comprehensively understand all the components in advance, arrange to have them tested in good time, and ensure that both software tools and hardware facilities are adequate for the tasks at hand. For example, CPU-intensive tasks might require massive pre-calculations or specialised equipment (such as all-against-all BLAST[9]) computations on datasets with millions of sequences, or the assembly of large genomes). Just as important is verification of the quality of any datasets to be used during the event, as poor-quality datasets are likely to jeopardise the success of bio-hackathon sessions. To not waste valuable time, any task that can be tackled by a participant in isolation (without requiring the insight of the entire Team) should be completed in advance. It is vital to test all software and hardware before the event. Work with the hackers to establish the hardware requirements. Ensure that hardware equipment/components can be provided or temporarily replaced if need be.

Prepare a budget forecast for the event. The budget will be dedicated to the rental of premises, IT requirements and subsistence. Gather options of suitable venues and their prices. Look at the premises and find out what the rental includes. Fix the premises for the scheduled date.

Decide the total amount you can spend on subsistence. We recommend creating a spreadsheet of all costs. If you have no funds available, you will need to set a fee (which will ultimately be determined by the number of participants, including lecturers, organisers, and so on). If you do have to set fees, you should also be aware of the potential fiscal risks. Involve your administration in the process to ensure that you do not run into trouble: they will know best how to treat fee income. If feasible, search for potential sponsors – *e.g.*, companies with an interest in your bio-hackathon theme.

The accuracy in silico simulation datasets is of great importance for benchmarking bioinformatics tools as well as for experimental design. For that reason, it is a good recommendation to create simulated sequencing data (mock data). For this purpose, there are now several freely available software packages to simulate mock data. Two examples are ART (Huang *et al.*, 2012) and InSilicoSeq (Gourlé *et al.*, 2019). When selecting the bio-hackathon venue, the proper mock data can be chosen accordingly. If the hackathon is organised in an academic environment with high computational capability, the mock data could be of substantial size. The data simulated can be the minimal required to perform the proper testings if, for travel logistics, a hotel close to an airport is chosen.

We recommend creating a checklist for all tasks to be done before, during and after the event. Spread responsibility between the organisers, but ensure that they do their job seriously. Discuss and agree on the rules and procedures, and take care that rules are followed strictly. Figure 2 collates the organisational workflow for a complete bio-hackathon cycle, including the preparatory, implementation and follow-up phase.

The bio-hackathon

Rule 6: Choose a convenient location

Bio-hackathons should take place at convenient locations for the registered number of participants, and locations have to fulfil all scientific/computing and non-scientific (housing, food, *etc.*) needs. University/national computing centres are likely to offer excellent computational facilities but may have restrictive opening hours. Hotels, on the other hand, while often very convenient in many aspects, may overestimate the bandwidth they can provide, so this needs to be tested extensively upfront.

Specific requirements to consider include:

- location convenient for participants to reach (minimise travel time and cost);

[6]https://owncloud.org
[7]https://drive.google.com
[8]https://www.dropbox.com
[9]https://blast.ncbi.nlm.nih.gov/Blast.cgi

Hollmann et al. (2021) EMBnet.journal 26, e983
http://dx.doi.org/10.14806/ej.26.0.983

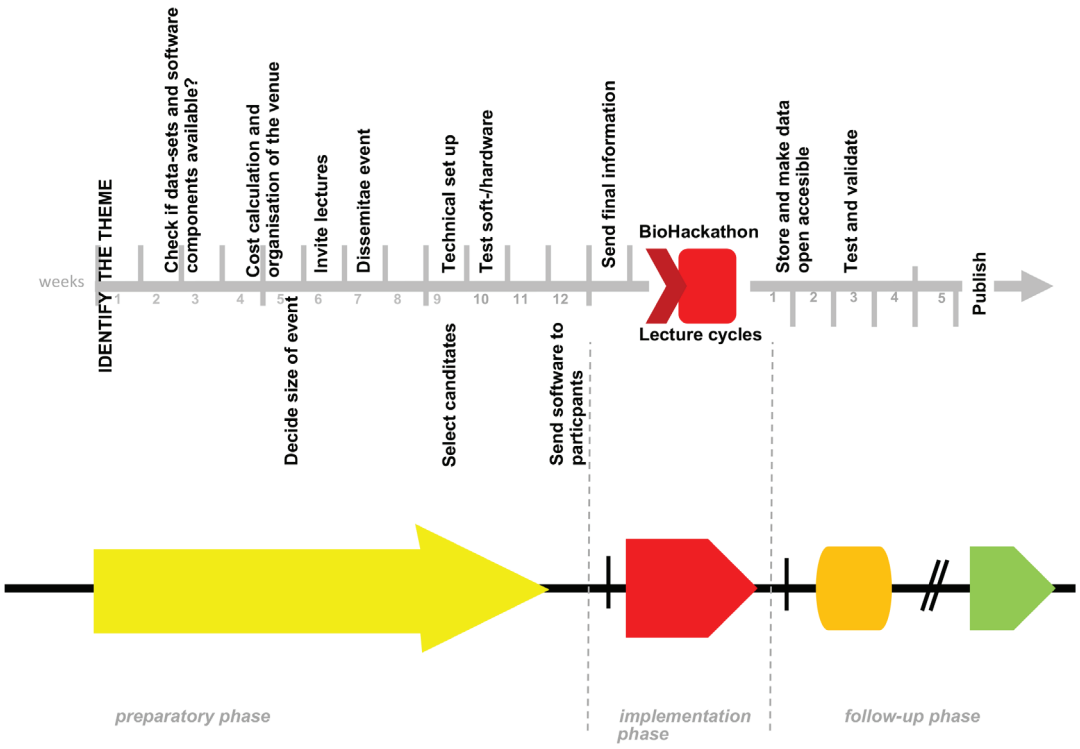


Figure 2. Workflow. The scheme demonstrates an optimal workflow for bio-hackathons, including the preparatory, implementation and follow-up phase for a complete cycle. Each phase is subdivided into different consecutive steps: in particular, the preparatory phase comprises a broad spectrum of tasks, including the selection of challenges, recruiting of participants, organisation of the venue and technical set-up, as well as the creation of webinars to prepare participants for the event.

- short distance between accommodation and meeting venue (if the venue is not the hotel);
- venue technically well equipped (beamer, screen, *etc.*), with liberal opening hours (often, much work is done outside regular working hours, and it is essential to facilitate this);
- venue has sufficient and stable bandwidth;
- food and drink are either available at the venue or allowed to be brought in. Often, many productive discussions occur informally over dinner, so arrangements that encourage the participants to keep together while eating are strongly preferred.

Of course, these events can also be conducted remotely if, for reasons such as the current pandemic situation, a face-to-face event is not possible. If so, Rule 6 is omitted and needs to be replaced by the organisation and establishment of a suitable online meeting tool (*e.g.* Zoom, Big Blue Buttom[10], Microsoft Teams[11], Slack. However, since bio-hackathons are based on an intensive and iterative exchange between biologists and computer experts, it is always preferable to hold them offline.

Rule 7: Ensure appropriate computer access

All bio-hackathons are not equal: some will have greater computational requirements than others. Some analyses might run efficiently on participants' laptops; some might require access to large clusters, supercomputers, dedicated hardware, or the cloud, which universities or national computer centres may be willing to provide. Regardless, the pre-requisites are i) fast internet connection at the hackathon venue, and ii) possibility for remote login to the computes facilities before and after the event. This last point is essential to prepare the ground-work beforehand, whereas any remaining work can be completed later. The local organiser should ensure (and check) that logins are available for all participants and ideally perform a test run before the bio-hackathon.

Similarly, if participants use their laptops, the requisite software should be installed before the event. It is recommended to create a Virtual Machine to provide a common computing environment for participants. To gain an overview of the software and hardware that will be needed during the bio-hackathon, we recommend gathering information about technical requirements via the registration form. Share this information with the hackers at the latest ten days before the event.

Rule 8: Ensure the duration is sufficient to obtain valuable outputs

Bio-hackathons are short, intensive working sessions, typically spanning a few days. Several considerations determine the duration of these events: the complexity of the workflow, how much computer work is envisaged (and how much can be done in advance), the funds

[10]https://www.bigbluebutton-hosting.de/
[11]https://www.microsoft.com/de-de/microsoft-teams/group-chat-software

available, how much time participants can commit, and whether writing documentation or article outlines are also intended to be part of the exercise. The expected outputs must therefore be clearly defined early on, and the duration of the event adjusted accordingly. It generally works well to organise hackathons over a weekend, as this affords participants greater flexibility with their schedules.

To kick off the event, plan to run a series of short lectures to better inform participants about the theme of the bio-hackathon and introduce its biological and computational components. Ensure the availability of suitably qualified lecturers. Disseminate information about these lectures to the participants and a broader audience at the latest two weeks before the hackathon. This may stimulate greater interest in the event and gain visibility within the community.

The lecture hall and workspaces might be at different locations. Ensure that you provide sufficient and detailed information about where and when to go to each place. If there is insufficient space to accommodate additional lecture series participants comfortably, focus on briefing the Team. This can also be done in the form of webinars before the event.

The Answer

Rule 9: Validate the results

Bio-hackathons aim to address particular biological problems. The events may focus on prototyping ideas, or they may lead to the production of tools or meta-tools that will ultimately be made available to the community. Before public release, validation events should be organised, in which participants are given opportunities to test the tool(s) with a variety of different datasets. Even though validation is normally done after hackathons, it should nevertheless be part of the initial planning to ensure that validation data exist, and that the software set-up is sufficiently generic to allow its use in validation. In an ideal case, most (if not all) of the original bio-hackathon Team should be present or (remotely) available during validation sessions.

Rule 10: Disseminate the results

Peer-reviewed publications are still the primary vehicles for disseminating scientific results, and reusable outputs from bio-hackathons are a good stimulus for article publication. However, public accessibility of all workflows must also be part of the dissemination strategy. Therefore, only open access publication platforms such as F1000research[12] should be used for publications. To maximise the outcomes impact, all workflows should also be properly documented and licensed, and inputs and outputs should be appropriately described following the FAIR principles (Wilkinson *et al.*, 2016) and using standardisation measures (Hollmann *et al.*, 2021). Ideally, alongside any publicly accessible documentation or article, small datasets that the workflows can use should also be included, inclusive of its corresponding Standard Operating Protocols (Hollmann *et al.*, 2020).

Optionally, Virtual Machine images to run workflows might also be provided. Results should be made available through openly accessible platforms such as SEEK[13], OpenAIRE[14], zenodo[15] or GitHub[16] that can guarantee longevity, as good workflows that answer biological questions often remain valuable for several years.

Potential pitfalls

The experience of the AllBio bio-hackathons provided an inside view of potential pitfalls that might limit the success of such events. A primary challenge is the careful selection of appropriate Problems; not all are suitable for inclusion in a bio-hackathon. It requires expert knowledge from both the biology and bioinformatics fields to evaluate the challenges and avoid frustration for the participants.

A specific function that bio-hackathons can perform is to enable interdisciplinary collaboration between the participants from the different expert fields. Sufficient time needs to be dedicated to training participants and finding a common language to discuss the challenges and develop efficient solutions.

Other more practical aspects may limit the success of events: *e.g.*, some early AllBio bio-hackathons struggled to deliver concrete outputs because:

- their teams were too small (≤5 people);
- the Team had no real leadership;
- the datasets on which they were obliged to work were too large to be processed fruitfully within the given time frame;
- the opening hours of computing centres limited the time available for productive work;
- the distance between hackathon venues and participants' hotels posed time and cost constraints.

A barrier to success may also occur if the meeting organiser/leader is no longer available after the event. The validation and follow-up phase is essential for summarising the results and ensuring the quality of solutions that have been developed. Moreover, publication of the results, whether via a journal article or upload to a repository, needs to be completed after the bio-hackathon. Costs associated with the dissemination of results need to be considered in the overall budget plan.

Conclusions

Bio-hackathons were powerful tools in the AllBio project for articulating and solving problems in the scientific community. They highlighted the need to consider the different disciplinary backgrounds of all participants, hence the vital role of the preparatory phase for ensuring

[12]https://f1000research.com/

[13]https://seek4science.org/
[14]https://explore.openaire.eu/participate/deposit-publications
[15]https://zenodo.org
[16]https://github.com/

the success of events. They also provided excellent opportunities, especially for young researchers, to learn new skills at the interface between disciplines, participate in advancing their field of research, and gain unique hands-on training with real challenges.

Some of the rules listed here may seem obvious, trivial, or even superfluous; nevertheless, all proved crucial in real-life scenarios. The ten rules provide practical guidelines for future bio-hackathon organisers, including preparations before, during and after the event itself.

Key Points

- New advances in sequencing technologies, bioinformatics, automation, artificial intelligence, *etc.*, to tackle this there is a need for continuous development of new bioinformatics solutions.
- Bio-hackathons are a helpful model to create bioinformatics solutions, often used by the bioinformatics community,
- Based on the work from the ALLBIO project this article present a step-by-step and standardised workflow explaining how to organise a bioinformatics hackathon.

Acknowledgments

The AllBio project was funded by the European Framework Programm GA 289452. The hackathon team members gratefully acknowledge SARA for providing computer facilities and hosting one hackathon. We thank NBIC's BioAssist programme for helping us find excellent programmers and allow them to spend adequate time on the projects.

References

1. Amar, D., Frades, I., Danek, A. *et al.* Evaluation and integration of functional annotation pipelines for newly sequenced organisms: the potato genome as a test case. BMC Plant Biol **14**, 329 (2014). http://dx.doi.org/10.1186/s12870-014-0329-9
2. Braune K, Rojas PD, Hofferbert J, Valera Sosa A, Lebedev A *et al.* (2021) Interdisciplinary Online Hackathons as an Approach to Combat the COVID-19 Pandemic: Case Study. J Med Internet Res. **23**(2):e25283. https://www.doi.org/10.2196/25283
3. Bongcam-Rudloff E *et al.* (2019) Broadening the bioinformatics infrastructure to unicellular, animal, and plant science D5.3 - Final summary report on solved test cases. Zenodo, http://dx.doi.org/10.5281/zenodo.3525052
4. Friedberg I, Wass MN, Mooney SD, Radivojac P (2015) Ten simple rules for a community computational challenge. PLoS Comput Biol. **11**(4):e1004150. http://dx.doi.org/10.1371/journal.pcbi.1004150
5. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M *et al.* (2014) Data integration in the era of omics: current and future challenges. BMC Syst Biol. **8**(Suppl 2):I1. http://dx.doi.org/10.1186/1752-0509-8-S2-I1
6. Gourlé H, Karlsson-Lindsjö O, Hayer J, Bongcam-Rudloff E (2019) Simulating Illumina metagenomic data with InSilicoSeq. Bioinformatics **35**(3),521-522. http://dx.doi.org/10.1093/bioinformatics/bty630
7. Hackseq Organizing Committee 2016 (2017) hackseq: Catalysing collaboration between biological and computational scientists via hackathon. F1000Research **28**;6:197. https://f1000research.com/articles/6-197/v2
8. Hollmann S, Frohme M, Endrullat C, Kremer A, D'Elia D *et al.* (2020) Ten simple rules on how to write a standard operating procedure. PLoS Comput Biol **16**(9):e1008095. http://dx.doi.org/10.1371/journal.pcbi.1008095
9. Hollmann S, Kremer A, Baebler S, *et al.* (2021) The need for standardisation in life science research - an approach to excellence and trust. F1000Research **9**:1398. http://dx.doi.org/10.12688/f1000research.27500.2
10. Huang W, Li L, Myers JR, Marth GT (2012) ART: a next-generation sequencing read simulator. Bioinformatics **28**(4),593-4. http://dx.doi.org/10.1093/bioinformatics/btr708
11. Leung W Y, Marschall T, Paudel Y, Falquet L, Mei H *et al.* (2015) SV-AUTOPILOT: optimised, automated construction of structural variation discovery and benchmarking pipelines. BMC genomics **16**(1), 238. http://dx.doi.org/10.1186/s12864-015-1376-9
12. Lyndon MP, Cassidy MP, Celi LA, Hendrik L, Kim YJ *et al.* (2018) Hacking Hackathons: Preparing the next generation for the multidisciplinary world of healthcare technology. Int J Med Inform. **112**, 1-5. http://dx.doi.org/10.1016/j.ijmedinf.2017.12.020
13. Poncette AS, Rojas PD, Hofferbert J, Valera Sosa A, Balzer F *et al.* (2020) Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations. J Med Internet Res. **22**(3):e17004. https://www.doi.org/10.2196/17004
14. Silver JK, Binder DS, Zubcevik N, Zafonte RD(2016) Healthcare Hackathons Provide Educational and Innovation Opportunities: A Case Study and Best Practice Recommendations. J Med Syst. **40**(7):177. https://www.doi.org/10.1007/s10916-016-0532-3
15. Wang JK, Pamnani RD, Capasso R, Chang RT (2018) An Extended Hackathon Model for Collaborative Education in Medical Innovation. J Med Syst. **42**(12):239. http://dx.doi.org/10.1007/s10916-018-1098-z
16. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data **3**:160018. http://dx.doi.org/10.1038/sdata.2016.18

EMBnet.journal
Bioinformatics in Action

Mitsis et al. (2020) EMBnet.journal 26, e934
http://dx.doi.org/10.14806/ej.26.0.934

Reviews

Exosomics

Thanasis Mitsis[1], Katerina Pierouli[1], Kalliopi Io Diakou[1], Eleni Papakonstantinou[1], Flora Bacopoulou[2], George P. Chrousos[2,3], Dimitrios Vlachakis[1,2,3]✉

[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; KP none; KID none; EP none; FB none; GPC none; DV none

Abstract

Extracellular vesicles have been the focus of a large number of studies in the past five years. Exosomes, a subgroup of extracellular vesicles, are of particularly high interest because they partake in a wide number of biological pathways. Produced by a variety of cells, exosomes have an important role in both physiological and pathological conditions. Exosome cargo heavily defines the vesicles' unique characteristics, and the cargo with the most intriguing prospects in its' biomedical applications is the non-coding RNAs. Non-coding RNAs, and specifically microRNAs are implicated in the regulation of many biological processes and have been associated with numerous diseases. Exosomes containing such important cargo can be used as biomarkers, therapeutic biomaterials, or even drug carriers. The potential media use of exosomes seems promising. However, some obstacles should be overcome before their clinical application. Synthetic exosome-like biomolecules may be a solution, but their production is still in their beginning stages. This review provides concise information regarding the current trends in exosome studies.

Introduction

Extracellular Vesicles (EVs) are membrane-bound vesicles secreted by cells into the extracellular space and have the ability to transport various molecules, such as DNA, RNA, and proteins, between cells (Zaborowski *et al.*, 2015; Doyle and Wang, 2019). EVs are, thus, essential mediators of cell-cell communication (Goran Ronquist, 2019; Raposo and Stahl, 2019). They can be classified into three main classes, which are microvesicles, exosomes, and apoptotic bodies (Yáñez-Mó *et al.*, 2015).

Currently, one of the most investigated classes of EVs is exosomes. Exosomes are single-membrane nano-sized vesicles with a diameter of ~30 to ~200nm with a topology similar to that of a cell (Pegtel and Gould, 2019). The reason for the intensive research that has taken place on exosomes is their specific role in cell communication. Intercellular communication through exosomes is important in both physiological and pathological biological function in humans (Camussi *et al.*, 2010). Exosomes appear to be of high importance in development, immunity, homeostasis, cancer, viral replication, tissue regeneration, and neurodegenerative diseases (de la Torre Gomez *et al.*, 2018; Pegtel and Gould, 2019). These abilities that exosomes possess showcase potential clinical applications, both as biomarkers and as therapeutic substance carriers (Zhang *et al.*, 2019).

As mentioned above, EVs can carry different types of molecules between cells. One exosome cargo that has attracted much attention for its potential clinical applications is the non-coding RNA (ncRNA), predominantly microRNAs (miRNAs) (Gallo *et al.*, 2012; Silva and Melo, 2015; Beuzelin and Kaeffer, 2018; Wang *et al.*, 2019). MiRNAs are a class of endogenous ncRNA molecules, approximately 20-22nt in length (Huang *et al.*, 2011) that have significant applicability as possible drug targets, modulators of drug resistance, and biomarkers for a wide variety of pathological conditions (Hanna *et al.*, 2019). The above information implies that modifications in the miRNA cargo of exosomes can provide various benefits in human health and an alternative to traditional drugs (Li *et al.*, 2018). Furthermore, synthetic exosome-mimics can be produced on a large scale, thus providing

Article history
Received: 23 May 2020
Accepted: 02 June 2020
Published: 04 August 2020

Mitsis et al. (2020) EMBnet.journal 26, e934
http://dx.doi.org/10.14806/ej.26.0.934

a feasible platform for a novel method of drug delivery (Li *et al.*, 2018).

Exosome Function

Exosomes are produced and released by various types of cells in the body, such as nervous system cells, such as Schwann cells (Ge *et al.*, 2012), astrocytes and neurons (Faure *et al.*, 2006; Mignot *et al.*, 2006), by epithelial cells, by fibroblasts and adipocytes, as well as by cells of the immune and hematopoietic systems, where their secretion was first detected,in particular by reticulocytes (Johnstone *et al.*, 1989; Fader *et al.*, 2005; Mignot *et al.*, 2006), B lymphocytes, T cells (Raposo *et al.*, 1996; Laulagnier *et al.*, 2004; Clayton *et al.*, 2005; Chaput *et al.*, 2006; Mignot *et al.*, 2006), platelets (Zitvogel *et al.*, 1998; Heijnen *et al.*, 1999; Laulagnier *et al.*, 2004; Clayton *et al.*, 2005; Chaput *et al.*, 2006; Mignot *et al.*, 2006), mast cells (Raposo *et al.*, 1996; Zitvogel *et al.*, 1998; Skokos *et al.*, 2003; Laulagnier *et al.*, 2004; Clayton *et al.*, 2005; Chaput *et al.*, 2006; Mignot *et al.*, 2006), dendritic cells (Raposo *et al.*, 1996; Zitvogel *et al.*, 1998; Chaput *et al.*, 2006; Mignot *et al.*, 2006), and macrophages (Nguyen *et al.*, 2003; Skokos *et al.*, 2003). Exosomes have also been detected in many types of biological fluids, such as breast milk, amniotic fluid, urine, blood, semen, bronchoalveolar lavage, synovial fluid and in the cerebrospinal fluid (Qin and Xu, 2014; Ellwanger *et al.*, 2017; Isola and Chen, 2017). Nowadays, it has been proven that the primary function of exosomes is the communication between cells, especially when they are distant from each other. Specifically, exosomes move from one cell that secretes them to another cell that internalises them, thereby transferring proteins and genetic material. Exosomes are also capable of transferring and spreading pathogens between cells, such as viruses and prions (Qin and Xu, 2014).

Due to the presence of exosomes in most cell types, they are involved in various procedures both in physiological and pathological conditions. One of the most important processes in which exosomes participate is the immune response. Immune cells secrete exosomes that are responsible for their inter-communication (Raposo *et al.*, 1996). From the beginning of an organism's infection, cells that recognise antigens, such as dendritic cells (DCs), are responsible for presenting the antigen to other immune cells. Antigen presentation occurs by secretion of exosomes containing membrane Major Histocompatibility Complex (MHC) molecules, which are recognised by T cell receptors and provoke their activation. Also, exosomes released by DCs that have recognised an antigen, carry the antigen to other DCs. Respectively, T helper cells activate B cells leading to increased secretion of exosomes containing MHC complexes to their membrane. In particular, it has been shown that exosomes secreted by B cells activate CD4+ T cells, which undermines the crucial role of exosomes in modulating the immune response. Upon completion of the immune response, exosomes are released by the DCs whose role is to promote its suppression, granted that they stimulate T cell apoptosis and lead to the conversion of T helper cells into regulatory T lymphocytes, thereby balancing pro-inflammatory and anti-inflammatory cells (Corrado *et al.*, 2013).

Another crucial role of exosomes is in the brain and nervous system functions. Here, exosomes participate in and assist the communication of neural cells with other types of cells, mainly between cells responsible for nerve axis integrity and myelination. Additionally, communication between neurons and oligodendrocytes, which are involved in the myelinating process, also depends on the secretion of exosomes. According to this mechanism, secretion of exosomes is signalled by glutamate, which as a neurotransmitter, leads to activation of glial ionotropic glutamate receptors. Through this process, exosomes are internalised by the neurons and release their cargo, which is now available for use. It has been reported that oligodendrocytic exosomes contain enzymes that resist oxidative stress, such as catalases and superoxide dismutase-1, thereby increasing neuronal tolerance to oxidative stress (Fruhbeis *et al.*, 2013; Frohlich *et al.*, 2014). Some studies have also recorded higher activation and expression of signalling pathways, such as the AKT and ERK pathways, in neurons that internalise exosomes (Frohlich *et al.*, 2014).

In the cardiovascular system, exosomes have been observed to contain TNF-α in hypoxic conditions (Yu *et al.*, 2012). Cardiomyocytes secrete exosomes with their secretion increasing rapidly under hypoxic conditions, while their contents change (Gupta and Knowlton, 2007). Under normal circumstances, no production of TNF-α is present in the heart tissue, as opposed to hypoxia. In this case, this factor is produced and secreted by the cardiomyocytes and transferred via exosomes to other healthy cells in which it induces apoptosis (Yu *et al.*, 2012). Thus through this mechanism, exosome secretion by the cells under stress conditions leads to the propagation of an inflammatory reaction. Exosomes also have the potential to induce modifications in the gene expression of recipient cells due to the genetic material, DNA and RNA, that can be transferred through them (Waldenstrom *et al.*, 2012). Therefore, exosomes constitute a non-specific cell type way of communicating in the heart (Danielson and Das, 2014).

The role of exosomes is also important in pathological situations, where they are involved in the development and spread of diseases. Their role has been clarified mainly in neurodegenerative (Vella *et al.*, 2008) and cardiovascular diseases (Halkein *et al.*, 2013), liver disease (Masyuk *et al.*, 2013) and cancer (Hannafon and Ding, 2013).

In neurodegenerative diseases, such as Parkinson and Alzheimer, exosomes may be responsible for disease spread. In Alzheimer disease (AD), accumulation of amyloid β (Aβ) molecules takes place, resulting in plaque formation in the brain (Bellingham *et al.*, 2012). In this particular case, exosomes are involved in the transfer of

amyloid β molecules to other neural cells of the brain resulting in the local spread of the disease. Also, increased secretion has been observed through the detection of an exosomal marker, Alix, in the brain of people with AD, as opposed to healthy ones in whom this marker is not detected (Aguzzi and Rajendran, 2009). Similarly, in Parkinson disease, α-synuclein enters exosomes that provide a catalytic environment through their lipids content. The result is a faster transfer of α-synuclein to other neural cells and consequent accumulation in the brain (Grey *et al.*, 2015).

A similar function of exosomes also occurs in cases of heart failure, and a more specific example is peripartum cardiomyopathy (PPCM), which occurs in pregnant or postpartum women. In this particular condition, exosomes function as carriers of a specific miRNA, miRNA-146a, which is produced by a prolactin fragment. These exosomes are taken up by cardiomyocytes and release their content into these cells, causing a decrease in cell metabolic activity and alterations in gene expression, ultimately resulting in heart failure (Halkein *et al.*, 2013).

At the heart of the research on exosomes is the study of their role in cancer. The exosomes secreted by cancer cells are transported to other cells of the same or other tissues, transferring both genetic material and proteins, that cause tumour proliferation and metastasis (Iero *et al.*, 2008; Hood *et al.*, 2009; Hood *et al.*, 2011; Kalluri, 2016; Steinbichler *et al.*, 2017; Whiteside, 2017). Examples are several types of cancers such as prostate and breast cancer in which proteins that induce fibroblast differentiation into myofibroblasts are transported through exosomes (Webber *et al.*, 2010; Vong and Kalluri, 2011) into different cells to activate Wnt signalling and cause activation and increased motility and activation of cancer cells resulting in metastases mainly to the lungs (Luga *et al.*, 2012; Kahlert and Kalluri, 2013). Similar in vitro studies have been conducted to study the promotion of metastasis by components of the exosomes (Jung *et al.*, 2009; Grange *et al.*, 2011). One example is the case in which exosomes from melanoma cells promote bone marrow cell tumorigenesis and metastases (Peinado *et al.*, 2012). Finally, there is a role of exosomes secreted by cancer cells in causing immunosuppression, which leads to suppression of the T cell response (Chalmin *et al.*, 2010). Exosome RNA content

Aside from DNA, proteins and lipids, exosomes possess a substantial RNA content. RNA species, such as messenger RNA (mRNA), miRNA, and long ncRNA (lncRNA) were shown to be present in exosomes in multiple studies. More modern techniques have revealed the presence of additional RNA species within exosomes, such as small nuclear RNA (snRNA), piwi-interacting RNA (piRNA), vault RNA, transfer RNA (tRNA), small nucleolar RNA (snoRNA), Y-RNA, SRP-RNA, small conditional RNA (scRNA), 7SK-RNA, as well as fragmented RNAs (Turchinovich *et al.*, 2019). Moreover, certain modifications of exosomal RNA, such as the 3'-end nucleotide additions and the 5'-terminal oligopyrimidine, have been reported and are possibly tied to RNA quality control processes (Koppers-Lalic *et al.*, 2014; Baglio *et al.*, 2016).

RNAs transcribed in a cell and released into an exosome can be received by another cell, resulting in the transfer of the RNA to the recipient cell in its functional form. This process has been described by various studies over several years, while more recent evidence has shed light on the mechanisms of RNA loading into the exosome. More precisely, RNA-binding proteins appear to bind specific subsets of RNAs. An example of this mechanism is the function of Gag and Gag-like proteins. These proteins impact the RNA content of the exosome by binding genomic RNA and other RNAs and transferring them into exosomes (Ashley *et al.*, 2018; Pastuzyn *et al.*, 2018). This exosomal transfer of RNA can play a crucial role in severe pathological conditions, such as cancer progression and metastasis.

Distribution of lncRNAs in exosomes has been strongly related to the parent cell type (Chen *et al.*, 2016) while also being subjected to regulation by changes in the cellular environment and possibly involved in disease pathogenesis (Hewson *et al.*, 2016). Studies have also shown that lncRNAs contained within exosomes can impact the function of cellular proteins involved in cell signalling, nucleosomal architecture and cell metabolism. Notably, several lncRNAs observed within exosomes have been found to function in cancer cell signalling (Hewson *et al.*, 2016). Kogure *et al.* reported that the exosomal lncRNA TUC339 from liver cancer cells could affect the microenvironment of the tumour, resulting in changes in adhesion and growth of tumour cells through the horizontal information transfer via exosomes (Kogure *et al.*, 2013). Another study shed light on the effect of lncRNA on gastric cancer cells, pointing out that the transfer of lncRNA ZFA1 through exosomes promoted the progression of this type of cancer (Pan *et al.*, 2017).

MiRNAs, the best-known class of RNA exosomal content, have been described by many studies as biomarkers and important components in intercellular communication. It has been shown that the proportion of miRNA is higher within exosomes than within the parent cell (Goldie *et al.*, 2014). As not all miRNAs are present in exosomes and changes in the cellular environment regulate their export, it has been speculated that specific miRNAs exit the cell in a tightly controlled process (Perez-Boza *et al.*, 2018).

MiRNAs within the exosome, function in two broad ways. One is the conventional negative regulation leading to changes in the expression of target genes. The second, more recently described function comes into view when miRNAs function is observed in their exosomal rather than intracellular state. Such studies were conducted on miR-29 and miR-21 contained in cancer cell-secreted exosomes, which were found to possess the ability to act as ligands, activating immune cells via Toll-like Receptor (TLR) binding (Fabbri *et al.*, 2012).

Mitsis et al. (2020) EMBnet.journal 26, e934
http://dx.doi.org/10.14806/ej.26.0.934

Several subspecies of miRNAs with roles in exocytosis, hematopoiesis tumorigenesis, and angiogenesis have been documented in intercellular communication via exosomes (Waldenstrom and Ronquist, 2014). Oshima *et al.* reported different levels of specific miRNA populations in exosomes derived from different cancer cell lines (Ohshima *et al.*, 2010). Moreover, different levels of specific miRNAs were reported in exosomes from the serum of healthy individual and glioblastoma patients (Skog *et al.*, 2008). Similar differences in specific exosomal miRNA levels have been reported between ovarian cancer and benign tumour cells (Taylor and Gercel-Taylor, 2008) as well as between exosomes from the plasma of healthy individuals and of non-small-cell lung carcinoma patients (Silva *et al.*, 2011).

Lastly, another study reported a close relationship between the expression of miR-134, a microRNA found in exosomes, and breast cancer, suggesting that this miRNA species can be used as a biomarker for diagnosis as well as a possible target for drug therapy (O'Brien *et al.*, 2015).

Exosome Applications

Interest in exosome research has escalated in the last decade because of their potential therapeutic applications (Li *et al.*, 2019). Exosomes may be indeed used as biomarker resources and as therapeutic biomolecule carriers (Zhang *et al.*, 2019). The key exosome feature that can be exploited is the fact that different cell types display differences in their exosome cargo (Sancho-Albero *et al.*, 2019). For instance, it has been demonstrated that exosomal miRNAs that partake in essential biological functions are lineage-specific and can override specific physiological mechanisms, and thus have the potential for a variety of clinical uses (Narayanan *et al.*, 2018).

Biomarkers are accurate and measurable indicators of health or pathological state (Comabella and Montalban, 2014). Biomarkers may include DNA, RNA, proteins, and metabolites. A particular non-invasive procedure of identifying biomarkers is the use of bodily fluids, such as serum, plasma, saliva and urine. Monitoring proteins in bodily fluids such as plasma, though, is a difficult procedure in complex disorders like cancer owing to the dynamic range of proteins contained, which may obstruct the detection of low abundance proteins. A promising way to overcome such difficulties is the use of EVs found in biological fluids, in particular exosomes (Boukouris and Mathivanan, 2015). As mentioned above, exosome cargo can provide extensive information on the state of the parental cell. Since pathological conditions lead to cells manufacturing disease-associated products, exosomes could contain a specific number of these products.

Furthermore, pathogens like viruses can take advantage of exosomes to infect host cells (Isola and Chen, 2017). Thus, assays for disease-associated molecules contained in exosomes may provide a high specificity biomarker test. The use of exosomes has many advantages compared to traditional biomarker tests because they are less complex samples than the whole bodily fluids, and their cargo is highly stable in storage conditions (Boukouris and Mathivanan, 2015). Current studies have showcased the potential of using exosomes as biomarkers in cancer prognosis and diagnosis, but more research is needed to evaluate the feasibility of such tasks (Huang and Deng, 2019; Jalalian *et al.*, 2019; Wong and Chen, 2019).

Naturally occurring exosomes could also be used as therapeutic biomaterials (Conlan *et al.*, 2017) because they may have therapeutic abilities characteristic of their counterpart live cells. Significant examples are mesenchymal stem cell (MSCs)derived exosomes (Zhao *et al.*, 2019). These cells are used as cellular therapy due to their regenerative and immunomodulatory effects. Granted that the vital mechanism behind mesenchymal stem cells derives from their paracrine ability, it is thought that various factors contained in their respective EVs orchestrate the main actions of MSCs (Hong *et al.*, 2019). The use of such exosomes may reduce side effects, including infusional toxicity (Mendt *et al.*, 2019).

Moreover, naturally occurring exosomes could be used as biomolecule carriers (Akuma *et al.*, 2019). There are various methods to load exosomes with the desired biomolecules and target specific cells. In the case of miRNAs, they can be loaded into exosomes through several methods including transfection of isolated exosomes with commercialised reagents, electroporation, active packaging through the use of proteins or conserved sequences of exosome enriched RNAs (eRNAs), transfection of the parental cells and the production of hybrid exosomes with liposomes (Liu and Su, 2019). In the case of small molecules, like chemotherapy drugs, loading methods may include direct mixing, ultrasonic treatment, and incubation with parental cells (Liu and Su, 2019). Regarding the targeting specificity of these exosomes, it can be determined through the selection of distinct parental cells, construction of targeting molecules or chemical modifications on the exosome surface (Liu and Su, 2019). All these methods have as a final goal the transport of therapeutic molecules to pathological cell targets and can be potentially applied as a therapeutic possibility to a large and diverse number of diseases (Samanta *et al.*, 2018). In the case of cardiovascular diseases, MSC-derived exosomes could be potentially applied. A study in mice showcased that purified MSC-derived exosomes can mitigate complications caused by reperfusion injury in myocardial ischemia after surgical blood flow restoration (Goran Ronquist, 2019). Specifically, the administration of MSC-derived exosomes just before reperfusion restores ATP and NADH levels while simultaneously reduce oxidative stress. Exosomes have also been proposed as therapeutic biomolecules for autoimmune diseases by exploiting their ability as immunomodulatory agents. In type 1 diabetes mellitus, SMCs might protect pancreatic islets of patients from

autoimmune targeting and therefore slowing disease progression (Xu *et al.*, 2019). In neurological, immune disorders, exosomes could deliver anti-inflammatory drugs to target brain cells. In a particular study, exosomes used to encapsulate curcumin or an inhibitor of the signal transducer and activator of transcription 3 (stat3) were noninvasively delivered to microglia cells and induced the apoptosis of the targeted microglial cells. This strategy could delay experimental autoimmune encephalomyelitis, an animal model of multiple sclerosis progression in mice (Zhuang *et al.*, 2011). Another autoimmune disease that could provide a potential use for therapeutic exosomes is rheumatoid arthritis. A research study has shown that IL-10-treated dendritic cells-derived exosomes may be able to suppress the onset of murine collagen-induced arthritis, an animal model of rheumatoid arthritis, as well as to reduce the severity of established arthritis (Kim *et al.*, 2005). Exosomes could also help diagnose or even be a potential treatment for developmental brain disorders. Specifically, in Rett syndrome, a developmental brain disorder with autism-like symptoms, Rett-affected exosomes lack essential neurodevelopmental proteins, while the administration of 'healthy' exosomes to a culture-dish model of Rett syndrome displayed therapeutic effects (Sharma *et al.*, 2019). Lastly, exosomes have been thoroughly studied for their therapeutic application in cancer. It has been shown that dendritic cell-derived exosomes can prime naïve T-cells and activate natural killer cells to shrink tumours (Gao and Jiang, 2018). Moreover, exosomes can deliver synthetic anticancer drugs to targeted cancer cells (Lu *et al.*, 2018).

While the clinical use of naturally occurring exosomes seems to be an up-and-coming field of study, it is important to be prescient in their utilisation as they take part in a large number of physiological pathways. Their multifaceted abilities might have adverse effects on a patient's immune response, cancer progression, drug resistance and metabolism (Conlan *et al.*, 2017). Furthermore, several challenges may also arise due to the difficulty in production, isolation, and storage on a commercial scale (Yamashita *et al.*, 2018). These difficulties can be addressed through the production of synthetic exosome-like biomaterials (Garcia-Manrique *et al.*, 2018). However, although natural exosomes clinical trials have just begun, synthetic exosomes are still in their first steps, demanding the development of standardised production protocols, studying their modes of actions and performing safety checks (Garcia-Manrique *et al.*, 2018).

Concluding Remarks

Exosomes are an intriguing field of study. Their cargo and unique abilities imply vast potential in their use as biomarkers, natural therapeutic vehicles and drug carriers. Before advancing in their clinical application, though, the mechanisms dictating their role in physiological and pathological conditions should be better elucidated.

Key Points

- Exosomes are single-membrane nano-sized vesicles secreted by a great variety of cell types that have an essential role in intercellular signaling both in physiological and pathological conditions.
- Exosomes are carriers of various biomolecules such as DNA, RNA, lipids, and proteins, with mentioned cargo characterizing their function.
- miRNAs, a class of non-coding RNAs, are the best-known class of RNA exosomal content and may influence gene expression and immune response.
- Exosomes can be used as biomarkers, where their cargo can provide extensive information on the state of the parental cell.
- Exosomes can also be used as therapeutic biomaterials, where their natural cargo or artificially loaded biomolecules can provide them with therapeutic abilities.

Acknowledgements

DV would like to acknowledge funding from Microsoft Azure for Genomics Research Grant (CRM:0740983), the Amazon Web Services Cloud for Genomics Research Grant (309211522729) and the AdjustEBOVGP-Dx (RIA2018EF-2081): Biochemical Adjustments of native EBOV Glycoprotein in Patient Sample to Unmask target Epitopes for Rapid Diagnostic Testing. A European & Developing Countries Clinical Trials Partnership (EDCTP2) under the Horizon 2020 "Research and Innovation Actions" DESCA.

References

1. Aguzzi A and Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. Neuron **64**(6), 783-790. http://dx.doi.org/10.1016/j.neuron.2009.12.016.
2. Akuma P, Okagu OD and Udenigwe CC (2019) Naturally Occurring Exosome Vesicles as Potential Delivery Vehicle for Bioactive Compounds. Frontiers in Sustainable Food Systems **3**(23). http://dx.doi.org/10.3389/fsufs.2019.00023.
3. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V *et al.* (2018) Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons. Cell **172**(1-2), 262-274.e211. http://dx.doi.org/10.1016/j.cell.2017.12.022.
4. Baglio SR, van Eijndhoven MA, Koppers-Lalic D, Berenguer J, Lougheed SM *et al.* (2016) Sensing of latent EBV infection through exosomal transfer of 5'pppRNA. Proc Natl Acad Sci U S A **113**(5), E587-596. http://dx.doi.org/10.1073/pnas.1518130113.
5. Bellingham SA, Guo BB, Coleman BM and Hill AF (2012) Exosomes: vehicles for the transfer of toxic proteins associated with neurodegenerative diseases? Front Physiol **3**, 124. http://dx.doi.org/10.3389/fphys.2012.00124.
6. Beuzelin D and Kaeffer B (2018) Exosomes and miRNA-Loaded Biomimetic Nanovehicles, a Focus on Their Potentials Preventing Type-2 Diabetes Linked to Metabolic Syndrome. Front Immunol **9**, 2711. http://dx.doi.org/10.3389/fimmu.2018.02711.
7. Boukouris S and Mathivanan S (2015) Exosomes in bodily fluids are a highly stable resource of disease biomarkers. Proteomics Clin Appl **9**(3-4), 358-367. http://dx.doi.org/10.1002/prca.201400114.

8. Camussi G, Deregibus MC, Bruno S, Cantaluppi V and Biancone L (2010) Exosomes/microvesicles as a mechanism of cell-to-cell communication. Kidney Int **78**(9), 838-848. http://dx.doi.org/10.1038/ki.2010.278.
9. Chalmin F, Ladoire S, Mignot G, Vincent J, Bruchard M *et al.* (2010) Membrane-associated Hsp72 from tumor-derived exosomes mediates STAT3-dependent immunosuppressive function of mouse and human myeloid-derived suppressor cells. J Clin Invest **120**(2), 457-471. http://dx.doi.org/10.1172/JCI40483.
10. Chaput N, Flament C, Viaud S, Taieb J, Roux S *et al.* (2006) Dendritic cell derived-exosomes: biology and clinical implementations. Journal of Leukocyte Biology **80**(3), 471-478. http://dx.doi.org/10.1189/jlb.0206094.
11. Chen M, Xu R, Ji H, Greening DW, Rai A *et al.* (2016) Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line. Sci Rep **6**, 38397. http://dx.doi.org/10.1038/srep38397.
12. Clayton A, Turkes A, Navabi H, Mason MD and Tabi Z (2005) Induction of heat shock proteins in B-cell exosomes. J Cell Sci **118**(Pt 16), 3631-3638. http://dx.doi.org/10.1242/jcs.02494.
13. Comabella M and Montalban X (2014) Body fluid biomarkers in multiple sclerosis. Lancet Neurol **13**(1), 113-126. http://dx.doi.org/10.1016/s1474-4422(13)70233-3.
14. Conlan RS, Pisano S, Oliveira MI, Ferrari M and Mendes Pinto I (2017) Exosomes as Reconfigurable Therapeutic Systems. Trends Mol Med **23**(7), 636-650. http://dx.doi.org/10.1016/j.molmed.2017.05.003.
15. Corrado C, Raimondo S, Chiesi A, Ciccia F, De Leo G *et al.* (2013) Exosomes as intercellular signaling organelles involved in health and disease: basic science and clinical applications. Int J Mol Sci **14**(3), 5338-5366. http://dx.doi.org/10.3390/ijms14035338.
16. Danielson KM and Das S (2014) Extracellular Vesicles in Heart Disease: Excitement for the Future ? Exosomes and microvesicles **2**(1), 10.5772/58390. http://dx.doi.org/10.5772/58390.
17. de la Torre Gomez C, Goreham RV, Bech Serra JJ, Nann T and Kussmann M (2018) "Exosomics"-A Review of Biophysics, Biology and Biochemistry of Exosomes With a Focus on Human Breast Milk. Front Genet **9**, 92. http://dx.doi.org/10.3389/fgene.2018.00092.
18. Doyle LM and Wang MZ (2019) Overview of Extracellular Vesicles, Their Origin, Composition, Purpose, and Methods for Exosome Isolation and Analysis. Cells **8**(7), 727. http://dx.doi.org/10.3390/cells8070727.
19. Ellwanger JH, Veit TD and Chies JAB (2017) Exosomes in HIV infection: A review and critical look. Infect Genet Evol **53**, 146-154. http://dx.doi.org/10.1016/j.meegid.2017.05.021.
20. Fabbri M, Paone A, Calore F, Galli R, Gaudio E *et al.* (2012) MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proceedings of the National Academy of Sciences **109**(31), 12278-12279. http://dx.doi.org/10.1073/pnas.1209414109.
21. Fader CM, Savina A, Sanchez D and Colombo MI (2005) Exosome secretion and red cell maturation: Exploring molecular components involved in the docking and fusion of multivesicular bodies in K562 cells. Blood Cells Mol Dis **35**(2), 153-157. http://dx.doi.org/10.1016/j.bcmd.2005.07.002.
22. Faure J, Lachenal G, Court M, Hirrlinger J, Chatellard-Causse C *et al.* (2006) Exosomes are released by cultured cortical neurones. Mol Cell Neurosci **31**(4), 642-648. http://dx.doi.org/10.1016/j.mcn.2005.12.003.
23. Frohlich D, Kuo WP, Fruhbeis C, Sun JJ, Zehendner CM *et al.* (2014) Multifaceted effects of oligodendroglial exosomes on neurons: impact on neuronal firing rate, signal transduction and gene regulation. Philos Trans R Soc Lond B Biol Sci **369**(1652) http://dx.doi.org/10.1098/rstb.2013.0510.
24. Fruhbeis C, Frohlich D, Kuo WP, Amphornrat J, Thilemann S *et al.* (2013) Neurotransmitter-triggered transfer of exosomes mediates oligodendrocyte-neuron communication. PLoS Biol **11**(7), e1001604. http://dx.doi.org/10.1371/journal.pbio.1001604.
25. Gallo A, Tandon M, Alevizos I and Illei GG (2012) The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. PLoS One **7**(3), e30679. http://dx.doi.org/10.1371/journal.pone.0030679.
26. Gao D and Jiang L (2018) Exosomes in cancer therapy: a novel experimental strategy. American journal of cancer research **8**(11), 2165-2175.
27. Garcia-Manrique P, Gutierrez G and Blanco-Lopez MC (2018) Fully Artificial Exosomes: Towards New Theranostic Biomaterials. Trends Biotechnol **36**(1), 10-14. http://dx.doi.org/10.1016/j.tibtech.2017.10.005.
28. Ge R, Tan E, Sharghi-Namini S and Asada HH (2012) Exosomes in Cancer Microenvironment and Beyond: have we Overlooked these Extracellular Messengers? Cancer microenvironment : official journal of the International Cancer Microenvironment Society **5**(3), 323-332. http://dx.doi.org/10.1007/s12307-012-0110-2.
29. Goldie BJ, Dun MD, Lin M, Smith ND, Verrills NM *et al.* (2014) Activity-associated miRNA are packaged in Map1b-enriched exosomes released from depolarised neurons. Nucleic Acids Res **42**(14), 9195-9208. http://dx.doi.org/10.1093/nar/gku594.
30. Goran Ronquist K (2019) Extracellular vesicles and energy metabolism. Clin Chim Acta **488**, 116-121. http://dx.doi.org/10.1016/j.cca.2018.10.044.
31. Grange C, Tapparo M, Collino F, Vitillo L, Damasco C *et al.* (2011) Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. Cancer Res **71**(15), 5346-5356. http://dx.doi.org/10.1158/0008-5472.CAN-11-0241.
32. Grey M, Dunning CJ, Gaspar R, Grey C, Brundin P *et al.* (2015) Acceleration of alpha-synuclein aggregation by exosomes. J Biol Chem **290**(5), 2969-2982. http://dx.doi.org/10.1074/jbc.M114.585703.
33. Gupta S and Knowlton AA (2007) HSP60 trafficking in adult cardiac myocytes: role of the exosomal pathway. Am J Physiol Heart Circ Physiol **292**(6), H3052-3056. http://dx.doi.org/10.1152/ajpheart.01355.2006.
34. Halkein J, Tabruyn SP, Ricke-Hoch M, Haghikia A, Nguyen NQ *et al.* (2013) MicroRNA-146a is a therapeutic target and biomarker for peripartum cardiomyopathy. J Clin Invest **123**(5), 2143-2154. http://dx.doi.org/10.1172/JCI64365.
35. Hanna J, Hossain GS and Kocerha J (2019) The Potential for microRNA Therapeutics and Clinical Research. Front Genet **10**, 478. http://dx.doi.org/10.3389/fgene.2019.00478.
36. Hannafon BN and Ding WQ (2013) Intercellular communication by exosome-derived microRNAs in cancer. Int J Mol Sci **14**(7), 14240-14269. http://dx.doi.org/10.3390/ijms140714240.
37. Heijnen HF, Schiel AE, Fijnheer R, Geuze HJ and Sixma JJ (1999) Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. Blood **94**(11), 3791-3799. http://dx.doi.org/10.1182/blood.V94.11.3791.
38. Hewson C, Capraro D, Burdach J, Whitaker N and Morris KV (2016) Extracellular vesicle associated long non-coding RNAs functionally enhance cell viability. Noncoding RNA Res **1**(1), 3-11. http://dx.doi.org/10.1016/j.ncrna.2016.06.001.
39. Hong P, Yang H, Wu Y, Li K and Tang Z (2019) The functions and clinical application potential of exosomes derived from adipose mesenchymal stem cells: a comprehensive review. Stem Cell Res Ther **10**(1), 242. http://dx.doi.org/10.1186/s13287-019-1358-y.
40. Hood JL, Pan H, Lanza GM, Wickline SA, Consortium for Translational Research in Advanced I *et al.* (2009) Paracrine induction of endothelium by tumor exosomes. Lab Invest **89**(11), 1317-1328. http://dx.doi.org/10.1038/labinvest.2009.94.
41. Hood JL, San RS and Wickline SA (2011) Exosomes released by melanoma cells prepare sentinel lymph nodes for tumor metastasis. Cancer Res **71**(11), 3792-3801. http://dx.doi.org/10.1158/0008-5472.CAN-10-4455.
42. Huang T and Deng C-X (2019) Current Progresses of Exosomes as Cancer Diagnostic and Prognostic Biomarkers. International

journal of biological sciences **15**(1), 1-11. http://dx.doi.org/10.7150/ijbs.27796.

43. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM *et al.* (2011) Biological functions of microRNAs: a review. J Physiol Biochem **67**(1), 129-139. http://dx.doi.org/10.1007/s13105-010-0050-6.
44. Iero M, Valenti R, Huber V, Filipazzi P, Parmiani G *et al.* (2008) Tumour-released exosomes and their implications in cancer immunity. Cell Death Differ **15**(1), 80-88. http://dx.doi.org/10.1038/sj.cdd.4402237.
45. Isola AL and Chen S (2017) Exosomes: The Messengers of Health and Disease. Current neuropharmacology **15**(1), 157-165. http://dx.doi.org/10.2174/1570159x14666160825160421.
46. Jalalian SH, Ramezani M, Jalalian SA, Abnous K and Taghdisi SM (2019) Exosomes, new biomarkers in early cancer detection. Analytical Biochemistry **571**, 1-13. http://dx.doi.org/10.1016/j.ab.2019.02.013.
47. Johnstone RM, Bianchini A and Teng K (1989) Reticulocyte maturation and exosome release: transferrin receptor containing exosomes shows multiple plasma membrane functions. Blood **74**(5), 1844-1851. http://dx.doi.org/10.1182/blood.V74.5.1844.1844.
48. Jung T, Castellana D, Klingbeil P, Cuesta Hernandez I, Vitacolonna M *et al.* (2009) CD44v6 dependence of premetastatic niche preparation by exosomes. Neoplasia **11**(10), 1093-1105. http://dx.doi.org/10.1593/neo.09822.
49. Kahlert C and Kalluri R (2013) Exosomes in tumor microenvironment influence cancer progression and metastasis. J Mol Med (Berl) **91**(4), 431-437. http://dx.doi.org/10.1007/s00109-013-1020-6.
50. Kalluri R (2016) The biology and function of exosomes in cancer. J Clin Invest **126**(4), 1208-1215. http://dx.doi.org/10.1172/JCI81135.
51. Kim SH, Lechman ER, Bianco N, Menon R, Keravala A *et al.* (2005) Exosomes derived from IL-10-treated dendritic cells can suppress inflammation and collagen-induced arthritis. J Immunol **174**(10), 6440-6448. http://dx.doi.org/10.4049/jimmunol.174.10.6440.
52. Kogure T, Yan IK, Lin W-L and Patel T (2013) Extracellular Vesicle-Mediated Transfer of a Novel Long Noncoding RNA TUC339: A Mechanism of Intercellular Signaling in Human Hepatocellular Cancer. Genes & cancer **4**(7-8), 261-272. http://dx.doi.org/10.1177/1947601913499020.
53. Koppers-Lalic D, Hackenberg M, Bijnsdorp IV, van Eijndhoven MAJ, Sadek P *et al.* (2014) Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. Cell Rep **8**(6), 1649-1658. http://dx.doi.org/10.1016/j.celrep.2014.08.027.
54. Laulagnier K, Motta C, Hamdi S, Roy S, Fauvelle F *et al.* (2004) Mast cell- and dendritic cell-derived exosomes display a specific lipid composition and an unusual membrane organisation. Biochem J **380**(Pt 1), 161-171. http://dx.doi.org/10.1042/bj20031594.
55. Li SP, Lin ZX, Jiang XY and Yu XY (2018) Exosomal cargo-loading and synthetic exosome-mimics as potential therapeutic tools. Acta Pharmacol Sin **39**(4), 542-551. http://dx.doi.org/10.1038/aps.2017.178.
56. Li X, Corbett AL, Taatizadeh E, Tasnim N, Little JP *et al.* (2019) Challenges and opportunities in exosome research-Perspectives from biology, engineering, and cancer therapy. APL bioengineering **3**(1), 011503-011503. http://dx.doi.org/10.1063/1.5087122.
57. Liu C and Su C (2019) Design strategies and application progress of therapeutic exosomes. Theranostics **9**(4), 1015-1028. http://dx.doi.org/10.7150/thno.30853.
58. Lu M, Xing H, Xun Z, Yang T, Zhao X *et al.* (2018) Functionalised extracellular vesicles as advanced therapeutic nanodelivery systems. Eur J Pharm Sci **121**, 34-46. http://dx.doi.org/10.1016/j.ejps.2018.05.001.
59. Luga V, Zhang L, Viloria-Petit AM, Ogunjimi AA, Inanlou MR *et al.* (2012) Exosomes mediate stromal mobilisation of autocrine Wnt-PCP signaling in breast cancer cell migration. Cell **151**(7), 1542-1556. http://dx.doi.org/10.1016/j.cell.2012.11.024.
60. Masyuk AI, Masyuk TV and Larusso NF (2013) Exosomes in the pathogenesis, diagnostics and therapeutics of liver diseases. J Hepatol **59**(3), 621-625. http://dx.doi.org/10.1016/j.jhep.2013.03.028.
61. Mendt M, Rezvani K and Shpall E (2019) Mesenchymal stem cell-derived exosomes for clinical use. Bone Marrow Transplant **54**(Suppl 2), 789-792. http://dx.doi.org/10.1038/s41409-019-0616-z.
62. Mignot G, Roux S, Thery C, Segura E and Zitvogel L (2006) Prospects for exosomes in immunotherapy of cancer. J Cell Mol Med **10**(2), 376-388. http://dx.doi.org/10.1111/j.1582-4934.2006.tb00406.x.
63. Narayanan K, Kumar S, Padmanabhan P, Gulyas B, Wan ACA *et al.* (2018) Lineage-specific exosomes could override extracellular matrix mediated human mesenchymal stem cell differentiation. Biomaterials **182**, 312-322. http://dx.doi.org/10.1016/j.biomaterials.2018.08.027.
64. Nguyen DG, Booth A, Gould SJ and Hildreth JE (2003) Evidence that HIV budding in primary macrophages occurs through the exosome release pathway. J Biol Chem **278**(52), 52347-52354. http://dx.doi.org/10.1074/jbc.M309009200.
65. O'Brien K, Lowry MC, Corcoran C, Martinez VG, Daly M *et al.* (2015) miR-134 in extracellular vesicles reduces triple-negative breast cancer aggression and increases drug sensitivity. Oncotarget **6**(32), 32774-32789. http://dx.doi.org/10.18632/oncotarget.5192.
66. Ohshima K, Inoue K, Fujiwara A, Hatakeyama K, Kanto K *et al.* (2010) Let-7 microRNA family is selectively secreted into the extracellular environment via exosomes in a metastatic gastric cancer cell line. PLoS One **5**(10), e13247. http://dx.doi.org/10.1371/journal.pone.0013247.
67. Pan L, Liang W, Fu M, Huang ZH, Li X *et al.* (2017) Exosomes-mediated transfer of long noncoding RNA ZFAS1 promotes gastric cancer progression. J Cancer Res Clin Oncol **143**(6), 991-1004. http://dx.doi.org/10.1007/s00432-017-2361-2.
68. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV *et al.* (2018) The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell **172**(1-2), 275-288.e218. http://dx.doi.org/10.1016/j.cell.2017.12.024.
69. Pegtel DM and Gould SJ (2019) Exosomes. Annual Review of Biochemistry **88**(1), 487-514. http://dx.doi.org/10.1146/annurev-biochem-013118-111902.
70. Peinado H, Aleckovic M, Lavotshkin S, Matei I, Costa-Silva B *et al.* (2012) Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat Med **18**(6), 883-891. http://dx.doi.org/10.1038/nm.2753.
71. Perez-Boza J, Lion M and Struman I (2018) Exploring the RNA landscape of endothelial exosomes. **24**(3), 423-435. http://dx.doi.org/10.1261/rna.064352.117.
72. Qin J and Xu Q (2014) Functions and application of exosomes. Acta Pol Pharm **71**(4), 537-543.
73. Raposo G, Nijman HW, Stoorvogel W, Liejendekker R, Harding CV *et al.* (1996) B lymphocytes secrete antigen-presenting vesicles. The Journal of experimental medicine **183**(3), 1161-1172. http://dx.doi.org/10.1084/jem.183.3.1161.
74. Raposo G and Stahl PD (2019) Extracellular vesicles: a new communication paradigm? Nature Reviews Molecular Cell Biology **20**(9), 509-510. http://dx.doi.org/10.1038/s41580-019-0158-7.
75. Samanta S, Rajasingh S, Drosos N, Zhou Z, Dawn B *et al.* (2018) Exosomes: new molecular targets of diseases. Acta pharmacologica Sinica **39**(4), 501-513. http://dx.doi.org/10.1038/aps.2017.162.
76. Sancho-Albero M, Navascues N, Mendoza G, Sebastian V and Arruebo M (2019) Exosome origin determines cell targeting and the transfer of therapeutic nanoparticles towards target cells. **17**(1), 16. http://dx.doi.org/10.1186/s12951-018-0437-z.

77. Sharma P, Mesci P, Carromeu C, McClatchy DR, Schiapparelli L *et al.* (2019) Exosomes regulate neurogenesis and circuit assembly. Proc Natl Acad Sci U S A **116**(32), 16086-16094. http://dx.doi.org/10.1073/pnas.1902513116.
78. Silva J, Garcia V, Zaballos A, Provencio M, Lombardia L *et al.* (2011) Vesicle-related microRNAs in plasma of nonsmall cell lung cancer patients and correlation with survival. Eur Respir J **37**(3), 617-623. http://dx.doi.org/10.1183/09031936.00029610.
79. Silva M and Melo SA (2015) Non-coding RNAs in Exosomes: New Players in Cancer Biology. Current genomics **16**(5), 295-303. http://dx.doi.org/10.2174/1389202916666150707154719.
80. Skog J, Würdinger T, van Rijn S, Meijer DH, Gainche L *et al.* (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nature cell biology **10**(12), 1470-1476. http://dx.doi.org/10.1038/ncb1800.
81. Skokos D, Botros HG, Demeure C, Morin J, Peronet R *et al.* (2003) Mast cell-derived exosomes induce phenotypic and functional maturation of dendritic cells and elicit specific immune responses in vivo. J Immunol **170**(6), 3037-3045. http://dx.doi.org/10.4049/jimmunol.170.6.3037.
82. Steinbichler TB, Dudas J, Riechelmann H and Skvortsova, II (2017) The role of exosomes in cancer metastasis. Semin Cancer Biol **44**, 170-181. http://dx.doi.org/10.1016/j.semcancer.2017.02.006.
83. Taylor DD and Gercel-Taylor C (2008) MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. Gynecol Oncol **110**(1), 13-21. http://dx.doi.org/10.1016/j.ygyno.2008.04.033.
84. Turchinovich A, Drapkina O and Tonevitsky A (2019) Transcriptome of Extracellular Vesicles: State-of-the-Art. Frontiers in Immunology **10**(202) http://dx.doi.org/10.3389/fimmu.2019.00202.
85. Vella LJ, Sharples RA, Nisbet RM, Cappai R and Hill AF (2008) The role of exosomes in the processing of proteins associated with neurodegenerative diseases. Eur Biophys J **37**(3), 323-332. http://dx.doi.org/10.1007/s00249-007-0246-z.
86. Vong S and Kalluri R (2011) The role of stromal myofibroblast and extracellular matrix in tumor angiogenesis. Genes Cancer **2**(12), 1139-1145. http://dx.doi.org/10.1177/1947601911423940.
87. Waldenstrom A, Genneback N, Hellman U and Ronquist G (2012) Cardiomyocyte microvesicles contain DNA/RNA and convey biological messages to target cells. PLoS One **7**(4), e34653. http://dx.doi.org/10.1371/journal.pone.0034653.
88. Waldenstrom A and Ronquist G (2014) Role of exosomes in myocardial remodeling. Circ Res **114**(2), 315-324. http://dx.doi.org/10.1161/circresaha.114.300584.
89. Wang M, Yu F, Ding H, Wang Y, Li P *et al.* (2019) Emerging Function and Clinical Values of Exosomal MicroRNAs in Cancer. Mol Ther Nucleic Acids **16**, 791-804. http://dx.doi.org/10.1016/j.omtn.2019.04.027.
90. Webber J, Steadman R, Mason MD, Tabi Z and Clayton A (2010) Cancer exosomes trigger fibroblast to myofibroblast differentiation. Cancer Res **70**(23), 9621-9630. http://dx.doi.org/10.1158/0008-5472.CAN-10-1722.
91. Whiteside TL (2017) Exosomes carrying immunoinhibitory proteins and their role in cancer. Clin Exp Immunol **189**(3), 259-267. http://dx.doi.org/10.1111/cei.12974.
92. Wong C-H and Chen Y-C (2019) Clinical significance of exosomes as potential biomarkers in cancer. World journal of clinical cases **7**(2), 171-190. http://dx.doi.org/10.12998/wjcc.v7.i2.171.
93. Xu H, Jia S and Xu H (2019) Potential therapeutic applications of exosomes in different autoimmune diseases. Clin Immunol **205**, 116-124. http://dx.doi.org/10.1016/j.clim.2019.06.006.
94. Yamashita T, Takahashi Y and Takakura Y (2018) Possibility of Exosome-Based Therapeutics and Challenges in Production of Exosomes Eligible for Therapeutic Application. Biol Pharm Bull **41**(6), 835-842. http://dx.doi.org/10.1248/bpb.b18-00133.
95. Yáñez-Mó M, Siljander PRM, Andreu Z, Zavec AB, Borràs FE *et al.* (2015) Biological properties of extracellular vesicles and their physiological functions. Journal of extracellular vesicles **4**, 27066-27066. http://dx.doi.org/10.3402/jev.v4.27066.
96. Yu X, Deng L, Wang D, Li N, Chen X *et al.* (2012) Mechanism of TNF-alpha autocrine effects in hypoxic cardiomyocytes: initiated by hypoxia inducible factor 1alpha, presented by exosomes. J Mol Cell Cardiol **53**(6), 848-857. http://dx.doi.org/10.1016/j.yjmcc.2012.10.002.
97. Zaborowski MP, Balaj L, Breakefield XO and Lai CP (2015) Extracellular Vesicles: Composition, Biological Relevance, and Methods of Study. Bioscience **65**(8), 783-797. http://dx.doi.org/10.1093/biosci/biv084.
98. Zhang Y, Liu Y, Liu H and Tang WH (2019) Exosomes: biogenesis, biologic function and clinical potential. Cell Biosci **9**, 19. http://dx.doi.org/10.1186/s13578-019-0282-2.
99. Zhao T, Sun F, Liu J, Ding T, She J *et al.* (2019) Emerging Role of Mesenchymal Stem Cell-derived Exosomes in Regenerative Medicine. Curr Stem Cell Res Ther **14**(6), 482-494. http://dx.doi.org/10.2174/1574888x14666190228103230.
100. Zhuang X, Xiang X, Grizzle W, Sun D, Zhang S *et al.* (2011) Treatment of brain inflammatory diseases by delivering exosome encapsulated anti-inflammatory drugs from the nasal region to the brain. Molecular therapy : the journal of the American Society of Gene Therapy **19**(10), 1769-1779. http://dx.doi.org/10.1038/mt.2011.164.
101. Zitvogel L, Regnault A, Lozier A, Wolfers J, Flament C *et al.* (1998) Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes. Nat Med **4**(5), 594-600. http://dx.doi.org/10.1038/nm0598-594.

a way in

Vivienne Baillie Gerritsen

As children in Scotland, back home from school and when the weather was dry, we would fling our schoolbags into the hall and grab a few golf clubs, a ball and a tee. There was no need for a change of clothes or shoes, whatever we were wearing was good enough. The course was along the coast on the edge of the North Sea, and the balls we used were found in the dunes where they had been lost by more experienced players. We had four clubs – a driver, two irons and a putter. Putting was the best part of the game. You would aim carefully for the long slim rod with the little red flag, taking into account the odd clump of grass or small mounds on the green and then, holding your breath, watch the little white ball as it made its meandering way to the base of the rod to drop, almost as an afterthought, into the hole. Well… it so happens that viruses infect cells in a similar manner… Viruses need to get inside cells in order to multiply, and this is what brings on infection. Like the flagged pole that marked the way in for our golf ball, viruses recognise molecules on the surface of cells to which they bind, thus enabling them – or parts of them – to enter the host cells where they rapidly spread. The coronavirus which is wreaking havoc across the planet as I write these words, is able to recognise a protein on the surface of a variety of human cells known as angiotensin-converting enzyme 2, or ACE2.

Coronavirus, illustration by David Goodsell

Courtesy of the artist

Viruses are unable to multiply on their own, so they seek help. Cells are their only bet: plant cells, animal cells or even bacteria – every virus has its specificity. Why? Because cells have all the machinery viruses need to produce their little ones: virions. There are many different kinds of viruses and many ingenious ways of infecting cells but, to cut a long story short, the sole aim of a virus is to have its genome read and translated, so that the host cell produces everything the virus needs to make its progeny – by the thousands. Once the cell cannot contain them anymore, it literally bursts and the virions are released into the environment to repeat what their parents have just done. Infection spreads rapidly. What is more, spontaneous mutations are bound to occur and progeny – thanks to the brief time-lapse between generations and the laws of natural selection – are not only a little different from their parents but also frequently better at reproducing themselves. This is one of the reasons it can be so difficult to develop vaccines.

Why do viruses exist, you may wonder. It is a question no one can answer. Viruses may seem to cause more harm to Nature than they do good. However, this is not quite true. Their aim is not to cause disease but to replicate - and in so doing, they unfortunately also frequently cause damage, sometimes leading to death. It so happens, too, that our genomes - like those of plants and other animals - carry myriads of small stretches of genetic material which have been acquired over time through bouts of viral infection, and passed down to future generations. Some of this material turns out to be genetically beneficial. For instance, in mammals, the foetal placenta is separated from the mother's tissues thanks to a gene of viral origin. Some organisms also live in symbiosis with viruses - in this way, greenfly carry with them plant viruses that muffle the reactivity of plants they attack. What we know for sure is that viruses have been around for a very long time. In fact, they were probably part of the primordial genetic pool, and perhaps on the earth's surface before even cells appeared. But that is open to debate.

So, although viruses seem to be as old as life itself, no one knew what a virus looked like until the advent of electronic microscopy in the 1930s. Towards the close of the 19th century, scientists had begun to suspect that fluids like sap and pus held infectious substances that

the Dutch microbiologist Martinus Beijerinck (1851-1931) named "virus" from the Latin meaning slime, or poison. The first actual virus to be observed under the electron microscope was the plant tobacco mosaic virus, which looks like a small space shuttle. Its full structure was elucidated in 1955 on the basis of information gathered by the English chemist Rosalind Franklin (1920-1958). Following this, the architecture of many other viruses was elucidated as well as the vectors they use to reach their host cells – such as insects, body fluids, water or air. Today, over 5,500 viruses have been identified while new viruses and strains continue to be unveiled.

Coronavirus is spherical in shape. Its outside envelope protects the virus's genome, which the host cell will dutifully read and translate to produce hordes of coronavirions. The coronavirus which is creating global chaos at this very moment has been baptized SARS-CoV-2, for Severe Acute Respiratory Syndrome coronavirus "number 2", a cousin of SARS-CoV that spread across 26 countries in 2003 but was far less severe than SARS-CoV-2. Like all coronaviruses, SARS-CoV-2's surface is riddled with a protein known as the spike (S) protein that assembles as a crown around the virion, hence "corona". S proteins recognise angiotensin-converting enzyme 2, ACE2, on the surface of human cells that are abundant in our lungs and our cardiovascular and renal systems. Upon recognition, SARS-CoV-2 binds to the cell via ACE2. This may well initiate a structural change in both proteins finally resulting in the entry of ACE2 and the virus into the cell by way of a small intracellular bubble-like compartment known as an endosome. It is not hard to understand, then, why this particular coronavirus is able to cause respiratory distress, and is dangerous for those already suffering from a cardiovascular disease.

Like all viruses, SARS-CoV-2 is simply twisting a protein's fate – and a cell's – for its own benefit; ACE2 did not evolve for the sake of a virus. What does it do? ACE2 is a metallopeptidase that modifies angiotensin – a peptide hormone that causes vasoconstriction and an increase in blood pressure. ACE2 is claw-like in appearance; the "claw" protrudes from the cell membrane while a short tail-end anchors it in the cell's membrane. Two ACE2 monomers join to form an active dimer. ACE2 is an integral component of what is known as the renin-angiotensin system that controls blood pressure, as well as fluid and salt balance. In particular, ACE2 converts angiotensin into three smaller hormones Ang (1-7), Ang (1-9) and alamandine – all of which are involved in increasing vasodilation and reducing fibrosis, the exact opposite of what occurs with angiotensin. During viral infection by SARS-CoV-2, the S protein is thought to bind to ACE2 on the very tip of the claw, from where the virus will find its way in.

One of ACE2's original roles is to produce hormones that have positive cardiovascular effects – which makes it an ideal candidate for developing therapies against high blood pressure for instance. ACE2 is also an open door, though, for SARS-CoV-2. Understanding how exactly the S protein binds to ACE2 could provide a means to develop ligands or antibodies that would help suppress viral infection – without neutralizing ACE2 function. It is a complicated matter, given that ACE2 is probably not the only path SARS-CoV-2 uses to enter cells and that, like all viruses, the virus has this debilitating habit of mutating at a swift rate. Much has been understood over the past years and yet, as always, so little too. Research is moving fast these days to find something that will help to fight off SARS-CoV-2 infection and its disease, COVID-19. It is mind-boggling how a such a tiny parcel of chemistry – and not of our making – has managed to bring human society to near asphyxiation. The rest of nature, however, seems to be getting a well-deserved breath of fresh air.

Cross-references to UniProt

Angiotensin-converting enzyme 2, *Homo sapiens* (Human): Q9BYF1

References

1. Mendoza-Torres E., Oyarzún A., Mondaca-Ruff D., et al.
ACE2 and vasoactive peptides: novel players in cardiovascular/renal remodeling and hypertension
Therapeutic Advances in Cardiovascular Disease 9:217-237(2015)
PMID: 26275770

2. Yan R., Zhang Y., Li Y., Xia L., Guo Y., Zhou Q.
Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2
Science 10.1126/science.abb2762 (2020)
PMID: 32132184

3. Guo Y.-R., Cao Q.-D., Zhong-Si H., et al.
The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status
Military Medical Research https://doi.org/10.1186/s40779-020-00240-0 (2020)
PMID: 32169119

proteinspotlight

SIB Swiss Institute of Bioinformatics

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

beneath us

Vivienne Baillie Gerritsen

While our life is spent above ground level, myriads of other creatures spend theirs below it. There are those, too, who share their time between both worlds – like moles, ants and trees for example. Soil offers protection but it is also a great source for food. Plants and fungi are perhaps among the organisms which make the best out of this state of affairs as they forage underground for essential nutrients while, above, leaves suck in sunlight for energy and mushrooms ripen to spread spores. Seeking sustenance underground requires a system that can rummage through earth, like roots in plants or rhizomorphs in fungi. In the recent years, there has been much talk about tree roots which are able to form intricate networks underground. The same goes for the mycelia of fungi. One particular fungus, *Armillaria gallica,* created a buzz in the 1990s when scientists announced that they had found a colony whose rhizomorphs seemingly stretched over tens of acres. However, as rhizomorphs grow, they also spend a lot of time fending off microbes that also want to prosper. Consequently, fungi yield numerous antimicrobial products, among them: melleolides, whose synthesis depends on an enzyme known as protoilludane synthase, or PRO1.

Fairy Ring, by Amy Ross

Courtesy of the artist, www.amyross.com

Armillaria gallica and related *Armillaria* species are also called honey mushrooms because of its colour. Save for the patterns and hues on its body, its shape resembles the classic mushroom illustrated in fairy tales : a short stocky stem with a bulbous head. However, unlike the poisonous toadstool, honey mushrooms are edible. Found in temperate regions in Asia, North America and Europe, it is a plant parasite and attacks the roots of a broad host range of trees that have been damaged, thus gradually causing wood rot but also playing an important role in the ecosystem as it injects carbon back into the soil. It is not the mushroom *per se*, in other words the fruit body, which attacks trees but filaments known as rhizomorphs – very similar to plant roots and from which the mushrooms bud – that zigzag their way through the soil underground to feed off roots.

In the early 1990s, scientists stumbled over a surprising colony of *Armillaria* while carrying out research on the effects that extremely low frequency radio stations, such as those used for submarine communication, could have on the environment. Further research revealed that the underground rhizomorph network, collectively known as the mycelium, from one individual *Armillaria gallica* colony seemed to stretch over an area of 37 acres – which is very roughly 15 football fields –, weigh about 95 tonnes and be about 1,500 years old. Recent molecular studies, however, have not only demonstrated that the colony is far more widespread and heavier but that it is also far more elderly; the meandering mycelium spreads underneath an astounding 75 acres, weighs 400 tonnes and has been dated back 2,500 years!

We tend to forget that beneath the ground, under mushrooms, trees or indeed any garden, there is a more or less vast and complex network continuously rooting around, literally, for nutrients. For a colony's mycelium to reach such huge proportions, it must have a system – or two – not only to make it grow but also to ward off other organisms equally eager to find nutrients in the same environment, notably microbes. Intriguingly, *A.gallica* rhizomorphs happen to be bioluminescent, and luminescence heightens when

rhizomorph growth is disturbed which may indicate that it could be a means to avert heterotrophs – although scientists suggest that it could also be used to attract insects for spore dispersal, or could simply be the by-product of another biochemical reaction.

To be able to develop to such an extent, rhizomorphs must secrete squads of cell-wall degrading enzymes to enable them to riddle their way across various layers of tree roots. They must also fire battalions of virulence factors that would begin by guiding them towards roots to sup up its nutrients while shunning the tree's own defense system. But this is not all, *Armillaria* will have needed to develop its own defense system to fight off microbes equally eager to extract nutrients from the same environment, i.e. antimicrobial products, and these are known as melleolides.

Melleolides are terpenes. Terpenes form that characteristic acrid scent released by tree resin. The name terpene originates from the Greek terbinthine meaning resin, itself extracted from terebinth, a tree species found in Greece. Interestingly, terpenes were given their name by the German organic chemist August Kekulé who was the first to lay down the revolutionary theory of chemical structure in the 1850s. Chemical formulae had been around for some time already, but no one had understood how atoms were actually located with respect to one another – until Kekulé introduced the inspired notion of chemical valence, or affinity, and was able to draft the structure of benzene. The chemical structure of terpenes themselves – of which we currently know about 50,000 – was subsequently identified by the German chemist Otto Wallach who found a bottle of essential oils lying around in Kekulé's laboratory where he was working.

How does *Armillaria* synthesize its own melleolides? It begins by the cyclization of the universal precursor farnesyl diphosphate ($C_{15}H_{28}O_7P_2$) – and this happens to be a chemical reaction which is one of the most complex that occurs in Nature. In *Armillaria*, this particular step is catalyzed by a terpene synthase, notably dela(6)-protoilludene synthase, or PRO1, which produces an intermediate metabolite known as 6-protoilludane. 6-protoilludane does not have antimicrobial activity itself, but it constitutes the first essential step leading to the synthesis of melleolides. Subsequent steps generate yet further structural diversity ultimately giving rise to as many as 50 different melleolides – perhaps even more – and thus illustrate the major role melleolides have in *Armillaria's* defense system.

Melleolides are very diverse and show antimicrobial activity against a wide range of microorganisms, including bacteria and viruses. As yet, their mode of action has been poorly investigated; only melleolides with antifungal activity are known to inhibit translation. Antimicrobial products are of great interest to the medical world, and fungi have been a precious reservoir for a little over a century now. Remember the revolutionary finding by the Scottish microbiologist Alexander Fleming who managed to isolate penicillin antibiotics from *Penicillium* fungi in the 1920s – the first antibiotics ever to be used against bacterial infections. The thing is, over time, bacteria are becoming more and more resistant to the antibiotics we know and we need to find other sources, or indeed engineer them ourselves – but to be able to do that you need to understand the intimacy of their molecular chemistry.

A surprising discovery: despite its good age, the genome of the *A.gallica* colony discovered in the 1990s has proved to be remarkably resistant to genetic change with an overall mutation rate that is unexpectedly low. This has made a few researchers wonder about cancer and how tumours behave. Cancer tumours also develop within a restricted environment yet their genome undergoes multiple mutations over very short periods of time. So where does the difference lie? Could it be due to a DNA repair system which is highly effective in *Armillaria*? Maybe. But in tumours, progression thrives less on a dithering DNA repair system than an overenthusiastic DNA replication system. But it is food for thought.

Cross-references to UniProt

Delta(6)-protoilludene synthase, *Armillaria gallica* (Bulbous honey fungus): P0DL13

References

1. Dörfer M., Gressler M., Hoffmeister D.
Diversity and bioactivity of Armillaria sesquiterpene aryl ester natural products
Mycological Progress 18:1027-1037(2019)
DOI: https://doi.org/10.1007/s11557-019-01508-z

2. Engels B., Heinig U., Grothe T. et al.
Cloning and characterization of an Armillaria gallica cDNA encoding protoilludane synthase […]
The Journal of Biological Chemistry 286: 6871-6878(2011)
PMID: 21148562

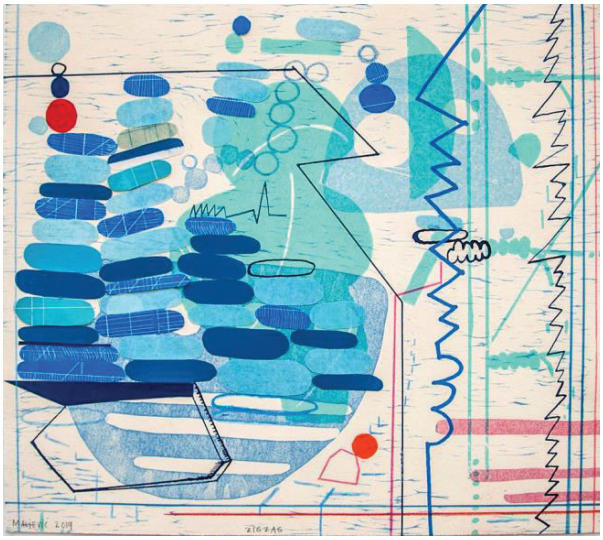
proteinspotlight

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

spotting patterns

Vivienne Baillie Gerritsen

When something is not right, signals of some sort are sent out. While many are obvious – an engine gone dead, water flooding a basement, flames protruding from a rooftop – others are more subtle: the slim crack in a dam before its collapse, the ray of sun about to set off a forest fire, the faint smell of gas before an explosion. Humans have developed ways of sensing and measuring these more discreet signals in the hope of predicting and preventing catastrophes. On a far smaller scale, cells too have ways of sensing something that has gone wrong. This can be in the form of entities such as molecules – either foreign or synthesized by the cell itself in situations of stress – which trigger off defence mechanisms. It can also be something far less tangible, such as patterns. Patterns? Imagine small weaving inaccuracies in cloth for example. This is not unlike unusual conformations which may arise in stretches of DNA or RNA (known as Z-DNA or Z-RNA) when cells are dealing with stress such as viral infection for instance. While, in the 17th Century, Antonie van Leeuwenhoek perfected the magnifying lens to spot stray stitches in cloth, over time cells have perfected their own lens to spy Z-DNA and Z-RNA: a protein known as ZBP1.



Zig Zag (2019), by Maja Maljevic

(linocut, woodcut, silkscreen, etching, colour pencil, ink)

Courtesy of the artist

In the 1970s, barely twenty years after the double-helical structure of DNA had been discovered, while observing X-ray images, the American biochemist Robert Wells noticed a peculiar twist in the backbone structure of a synthetic double-stranded DNA polymer. Instead of the usual smooth helical right-handed structure of DNA, it looked more like a zig zag – hence the name that it was later given: Z-DNA – despite the fact that the base-pairs and the sugar backbone were identical to those in the more classical DNA (also known as B-DNA). The discovery begged the question: does Z-DNA actually exist in live cells? And if so, does it have any biological relevance? The answer to the first question turned out to be: yes. The second has been subject to much debate but, almost 50 years later, it is now believed that stretches of DNA can flip to the left-handed Z conformation to relieve torsional strain during transcription – in the way a wet towel spontaneously unwinds when it has been wrung out tightly. Z-DNA would therefore be yet another mechanism used by cells in the process of decoding genetic information.

So short stretches of Z-DNA can be observed close to where transcription occurs in cells. But what is happening on the molecular scale? What is it that gives Z-DNA its zig zag configuration? Each Z-DNA stretch begins and ends with what has been coined a B-Z junction – where one base pair has snapped to relieve strain. The resulting end-bases stick out on either side of the DNA structure instead of facing in. In between the two junctions, the double helix is still held together by the traditional Watson Crick base pairs, but instead of the sugar backbone twisting smoothly as it traces major and minor grooves, the grooves traced by the

backbone of Z-DNA are of similar width while the orientation of the base pairs relative to the sugar backbone are uncharacteristic. As a consequence, the backbone of flipped DNA adopts a zig zag conformation.

As the years rolled by and research did too, it soon became apparent that double-stranded RNA could also adopt the Z conformation. This has been observed in the course of viral infection – such as influenza A virus (IAV) for example – during which Z-RNAs are produced. These stretches of Z-RNA are not part of the viral genome but seem to be by-products of the virus as it multiplies. Their formation provide telltale signals – albeit unintentional on behalf of the virus – that something is not right and infected cells are swift to pick them up. This is where the so-called Z-DNA Binding Protein, or ZBP1, comes into action. Human ZBP1 is a pathogen 'pattern' sensor in that it is able to recognize and bind to Z-RNAs produced during IAV infection. Once bound, specific pathways are activated, ultimately leading to host cell death and inflammation – both of which are distinct immune responses although interconnected and mutually regulated. Host cell death may sound counterintuitive when a host is fighting off infection, but it is one way – as long as it is controlled – of ensuring that viral replication is stunted.

How does ZBP1 know when to recognize stretches of Z-RNA in a host cell? ZBP1 seems to be expressed when cells are undergoing stress. The protein is found mainly in a cell's cytoplasm within what are known as stress granules and processing bodies. These entities form when, under stressful conditions, a cell decides to defer translation or perhaps even resume it. RNA molecules stored within stalled translation complexes are held in stress granules, and if a cell decides to stop translation altogether, the contents of the stress granules are destroyed thanks to the processing bodies. Following IAV infection, however, ZBP1 rapidly accumulates in the host cell's nucleus where the replicating RNA virus is generating Z-RNAs.

How exactly ZBP1 senses Z-RNA has still to be understood, but two domains in its sequence are important: Zα and Zβ, which are similar to one another and situated at the protein's amino-terminus. These domains are now known to bind specifically to Z-RNA or to Z-DNA; this is explained by the fact that ZBP1 binds to the sugar backbone rather than to the bases. As a consequence, ZBP1 is not specific to a nucleotide-base sequence but rather to the left-handed conformation of Z-RNA, or indeed of Z-DNA. Downstream of Zα and Zβ are two other domains known as RHIM domains, which are necessary to complete the immune response. Once bound to Z-RNA and by way of the RHIM domains, ZBP1 activates the beginning of a cascade ultimately leading to cell death by the disruption of the nuclear envelope and leakage of the cell's DNA into the cytosol.

The system is a very subtle way of sensing danger and launching an immune response. Though human ZBP1 is known to bind to Z-RNA during IAV infection – which is unique in that IAV is an RNA virus which replicates in the host cell's nucleus – ZBP1 can also bind to Z-DNA, and Nature has graced life with both DNA and RNA viruses. Besides sensing viral infection, ZBP1 seems to be associated with the onset and progression of certain cancers, as well as several autoimmune diseases. This does not really come as a surprise since, if ZBP1 were to be defective, cell death could fire ahead unbridled subsequently causing severe damage to the host, despite the fact that a virus has been wiped out. Currently, therapeutic strategies for targeting Z-DNA and the proteins that bind to it to manipulate the expression of local genes is being considered. Still, though ZBP1 has participated in demonstrating the biological relevance of Z-DNA, it continues to puzzle scientists. How is it activated? But more intriguingly, perhaps, how is a protein capable of sensing a change of pattern in DNA structure?

Cross-references to UniProt

Z-DNA binding protein 1, *Homo sapiens (Human)* : Q9H171

References

1. Zhang T., Yin C., Boyd D.F., et al.
Influenza virus Z-RNAs induce ZBP1-mediated necroptosis
Cell 181:1115-1129(2020)
PMID: 32200799

2. Kuriakose T., Kanneganti T.-D.
ZBP1: Innate sensor regulating cell death and inflammation
Trends in Immunology 39:123-134(2018)
PMID: 29236673

proteinspotlight

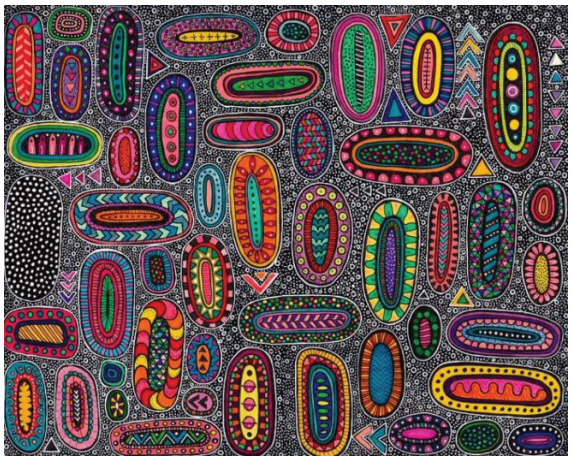
SIB Swiss Institute of Bioinformatics

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

stress

Vivienne Baillie Gerritsen

Stress is a warning signal. It is like setting off a fire alarm or dialling an emergency phone number. The point is to trigger a reaction which, in the best of cases, will set things right again. We have all sensed stress at various times in our lives and, if we are attentive, our bodies react by firing off different signals: muscle stiffness, fatigue, headaches, indigestion… Though we may instinctively think of stress as something which emerges from external sources – such as busy timetables, relationships, disease or starvation – stress can also arise from sources within us. In fact, living organisms recognised the benefits of stress as a warning signal long before the word existed. One example are mitochondria. Mitochondria are small organelles whose major role is to produce biological energy, otherwise known as ATP. Consequently, healthy mitochondria are paramount to life. If, for one reason or another, they become malfunctional, they may choose to trigger off what is known as an 'integrated stress response', or ISR. This indicates to the cell that they need help, and the cell will do its best to fix the situation. A key protein involved in activating mitochondrion-induced ISR is known as 'death ligand signal enhancer', or DELE1.



"microscopic" by Hello Angel

courtesy of the artist

Mitochondria are intriguing organelles, almost bestowed with a mind of their own. They are found in eukaryotic cells where they differ considerably in size and number – from one copy to thousands –, depending on the type of cell, its cycle and the physicochemical environment. They are supported by microtubules, and their rounded shape – from spherical to pellet-like – changes continuously depending on whether they are dividing (mitochondrial fission), fusing (mitochondrial fusion), or indeed dying (mitophagy). Characteristically, they have two membranes, an outer membrane and an inner membrane – both structurally and chemically similar to a cell's plasma membrane – which are separated by an intermembrane space. As the inner membrane has a greater surface than its outer counterpart, it looks as though it has been squeezed into a space too tight, which lends it this distinctive zig-zagged aspect, like a long neatly-packed sausage.

A surprising fact: mitochondria carry their own DNA. The vertebrate mitochondrial genome is invariably circular and stored within the inner membrane, known as the matrix. Unlike a cell's nucleus, mitochondria are usually equipped with several copies of their own DNA, on which are found a few of the genes for proteins required for ATP synthesis as well as the rRNA and tRNAs required for synthesizing proteins. Other proteins needed by mitochondria are imported from the cell's cytosol – such as DELE1, the protein involved in dealing with mitochondrial stress.

Though a mitochondrion's most distinctive job is to produce biological energy (ATP) by a pathway known as cellular respiration, it is also involved in many other metabolic pathways. Among them: programmed cell death, the regulation of cellular metabolism, steroid synthesis and hormonal signalling. Mitochondria were first observed in the

1840s and established as cell organelles – or bioblasts – *per se* in the 1890s. In 1898, they were termed mitochondria, from the Greek *mitos*, meaning 'thread', and *chondrion*, 'granule', which is what they look like under the microscope. The respiratory chain –the seat of ATP production, to cut a long story short – was discovered in the 1920s, but it took another 30 years or so before scientists realised that mitochondria, themselves, were the actual orchestrators of ATP production.

Mitochondrial fusion and fission are frequently an answer to various forms of stress. A damaged mitochondrion can fuse to a healthy one, thus mingling their contents, diluting the damage and replacing it by healthy components. Fission, on the other hand, can be used as a means of disposing of a mitochondrion's damaged elements which are transferred to one extremity, which then buds off. Each of these reactions arise from sending out a stress warning signal, or integrated response signal (IRS), which – in the best of cases – will lead to the restoration of cellular homeostasis. DELE1 is intimately involved in this process.

When a mitochondrion senses stress, DELE1 binds to its inner membrane where it is cleaved into a shorter active form: DELE1s. This shorter form is composed mainly of tetratricopeptide repeats (TPRs). These are structural motifs – typically bunches of alpha-helices – which form scaffolds that mediate protein-protein interactions. The active form of DELE1 is able to exit the mitochondrion and accumulates in the cell's cytosol where it binds – via its TPR domains – to an enzyme known as HRI. HRI is the first component of a pathway which ultimately leads to an integrated response signal. In this light, DELE1 can be seen as a messenger capable of relaying a stress signal from the mitochondrion to the cell.

Mitochondrion dynamics is a fascinating paragraph of cell biology, as is the origin of mitochondria. The most widely accepted hypothesis today is known as the endosymbiotic hypothesis. This hypothesis suggests that mitochondria were once prokaryotic cells, such as bacteria, with metabolic pathways – like cellular respiration – which eukaryotic cells then lacked. By becoming endosymbionts, these prokaryotic cells will have offered their hosts the ability to produce their own energy, which would have been an undeniable evolutionary advantage. Such a hypothesis is supported by several facts, one being that bacterial genomes are also circular. In the past years, ageing and age-associated diseases have been linked to mitochondria – or, perhaps more specifically, to the ability cells still have of reacting to mitochondrion damage. Understanding the intricacies of the ISR pathway could pave the way to therapeutic strategies – not to postpone ageing, but to subdue age-associated diseases.

Cross-references to UniProt

DAP3-binding cell death enhancer 1, *Homo sapiens* (Human): Q14154

References

1. Fessler E., Eckl E-M., Schmitt S., et al.
A pathway coordinated by DELE1 relays mitochondrial stress to the cytosol
Nature 579:433-437(2020)
PMID: 32132706

2. Guo X., Aviles G., Liu Y., et al.
Mitochondrial stress is relayed to the cytosol by an OMA1-DELE1-HRI pathway
Nature 579:427-432(2020)
PMID: 32132707

3. Youle R.J., van der Bliek A.M.
Mitochondrial fission, fusion and stress
Science 337:1062-1065(2012)
PMID: 22936770

proteinspotlight

Swiss Institute of Bioinformatics

Protein Spotlight (ISSN 1424-4721) is a montly review written by the **Swiss-Prot** team of the **SIB Swiss Institute of Bioinformatics**. Spotlight articles describe a specific protein or family of proteins on an informal tone. http://web.expasy.org/spotlight/

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