



Setup and use of advanced HPC Servers for biocomputing, bioinformatics and data management in **Biomedical Research**

Prof. Dr. Javier De Las Rivas

jrivas@usal.es

PI of the Bioinformatics & Functional Genomics Lab
Centro de Investigación del Cáncer
(CiC, CSIC/USAL) Salamanca (Spain)



06/05/2026



– **Setup and use of advanced HPC Servers for biocomputing, bioinformatics and data management in Biomedical Research**

Dr. Javier De Las Rivas

jrivas@usal.es

Cancer Research Center (CiC, CSIC/USAL)

Consejo Superior de Investigaciones Científicas (**CSIC**) & University of Salamanca (**USAL**)
Salamanca (Spain)

– **Current application of Artificial Intelligence tools and resources in Biomedical Research**

Dr. Alberto Berral Gonzalez

aberralgonzalez@usal.es

Cancer Research Center (CiC, CSIC/USAL)

Consejo Superior de Investigaciones Científicas (**CSIC**) & University of Salamanca (**USAL**)
Salamanca (Spain)



1 • WHO ARE WE ? : a multidisciplinary team of **biochemists, biotechnologists, computer scientists, informatic engineers, mathematician-statisticians** as well as **medical doctors**, who work with complex **biomedical and biomolecular data** using computers

2 • WHAT TYPE OF COMPUTERS WE USE ? :
for **Biological data & Biomedical research**

3 • BRIEF OVERVIEW of our **experience & scientific production** in **Computational Biology & Bioinformatics** applied to **Biomedical Research**

4 • BRIEFING about the current development of the **AI** in **Biological data & Biomedical research**

1 • WHO ARE WE ?:

We are a multidisciplinary team of **biochemists, biotechnologists, computer scientists, informatic engineers, mathematician-statisticians** as well as **medical doctors**, who work with **complex, large-scale (“omic”) biomedical and biomolecular data** and using computers.

We “speak”:

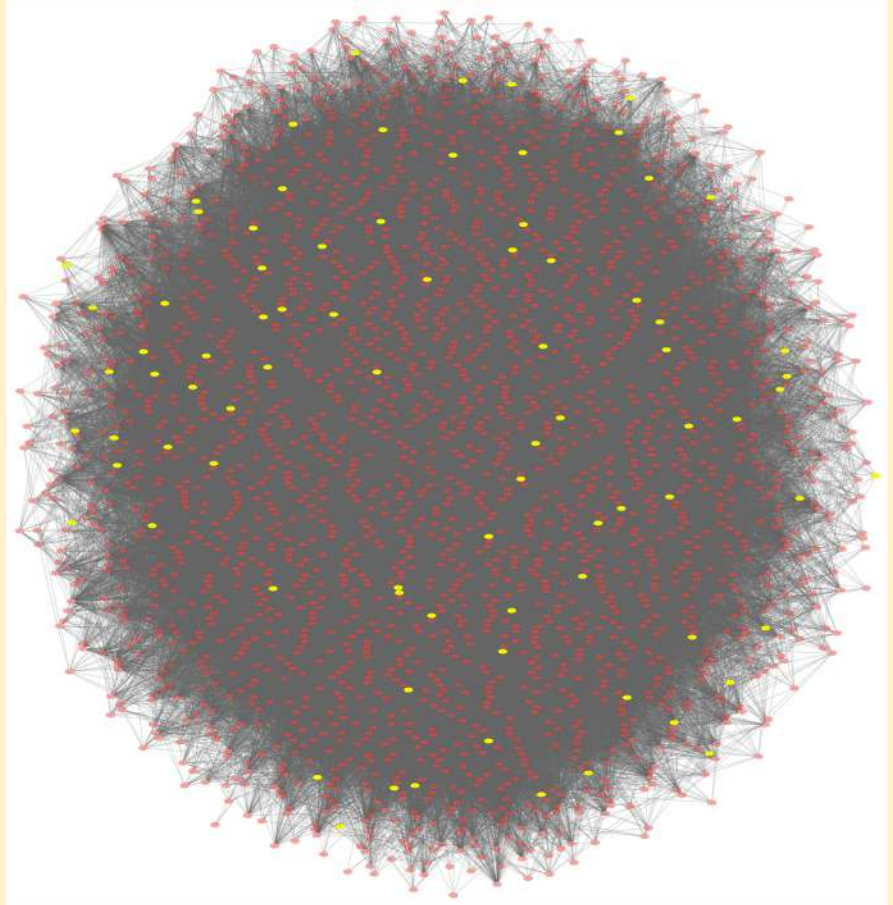
R, Python, JAVA, JavaScript,
Ruby, MySQL, shell-script,
HTML/PHP ...



1 • WHO ARE WE ?:

We analyse
complex and high-dimensional
biological & biomedical data

For example this is a regulatory
network of human genes
obtained for a study in ovarian
cancer
that includes:
2,414 nodes (genes) &
114,482 edges (associations)



We analyse complex and high-dimensional biological & biomedical data

Current **biomedical research** is a **petabyte factory**

30M+

PubMed
scientific
articles
~4.000 new/day

150 GB

per one human
whole genome
raw sequencing

10 yrs

& \approx 2.6 billion\$
drug
development

x100

omics data
growth over the
last decade
2015-2025

Biomedical research generates more data than human can process



HPC + AI bridges the gap between data and scientific insight

1 • WHO ARE WE ?:

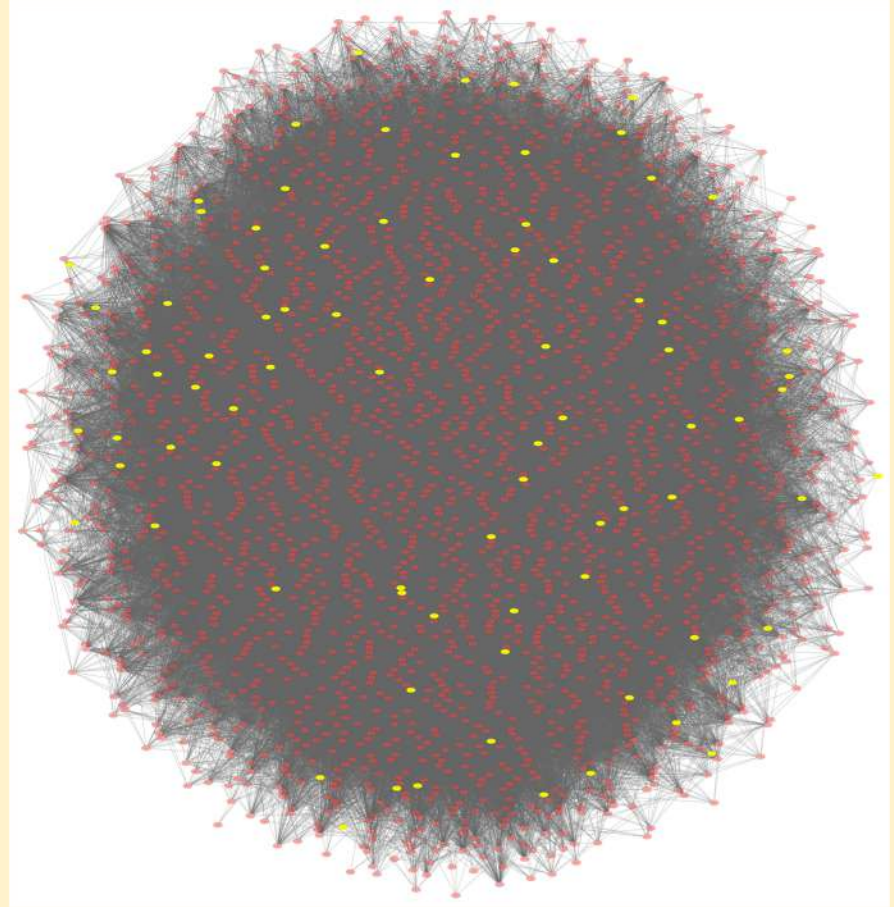
We analyse complex and high-dimensional biological & biomedical data

For example this is a regulatory network of human genes obtained for a study in ovarian cancer that includes:

2,414 nodes (genes) &
114,482 edges (associations)

This type of data can only be analysed using adequate computing servers:
In the last 10-12 years (from 2014) we have acquired excellent machines to SIE

www.sie.es



2 • WHAT TYPE OF COMPUTERS WE USE ?:

The complex bio-data can only be analysed using adequate computing servers:
In the last years (from 2014-2024) we have adquired machines to SIE www.sie.es



2 • WHAT TYPE OF COMPUTERS WE USE ?:

The complex bio-data can only be analysed using adequate computing servers:
In the last years (from 2014-2024) we have adquired machines to SIE www.sie.es

- in 2014 [cicabu] - SIE LADON ABU DABI 4 processors Opteron 6380 server with 64 cores & 256 GB RAM memory DDR3/1600 ECC + 2 HD SSD 512GB storage
- in 2016 [cicngs] - SIE LADON ABU DABI 4 processors Opteron 6380 server with 64 cores & 512 GB RAM memory DDR3/1600 ECC + 5 HD SAS 7200rpm with 24TB in RAID5 for storage
- in 2019 [cicdna] - SIE LADON NAPLES 2 processors AMD EPYC 7000 server with 128 cores & 512 GB RAM memory DDR4/2666 ECC + 1 SSD Micron M5100MAX with 480GB storage & 3 HD SAS 7200rpm with 12TB storage
- in 2024 [cicair] - SIE LADON MILAN 2 processors AMD EPYC 7763 server with 128 cores & 2 TB RAM memory DDR4/3200 ECC + 2 SSD CS900 SATA with 2TB storage & 2 Kioxia NVMe 12TB storage + 2 GPUs NVIDIA RTX 7,168 CUDA with 20GB VRAM GDDR6 & NVIDIA PRO5000 14,080 CUDA with 48GB VRAM GDDR7

2 • WHAT TYPE OF COMPUTERS WE USE ?:

The complex bio-data can only be analysed using adequate computing servers:
In the last years (from 2014-2024) we have acquired machines to SIE www.sie.es

- in 2014 [cicabu] - SIE LADON ABU DABI 4 processors Opteron 6380 server with 64 cores & 256 GB RAM memory DDR3/1600 ECC + 2 HD SSD 512GB storage
- in 2016 [cicngs] - SIE LADON ABU DABI 4 processors Opteron 6380 server with 64 cores & 512 GB RAM memory DDR3/1600 ECC + 5 HD SAS 7200rpm with 24TB in RAID5 for storage

2 • WHAT TYPE OF COMPUTERS WE USE ?:

- in 2019 [cicdna] - SIE LADON NAPLES 2 processors AMD EPYC 7000 server with 128 cores & 512 GB RAM memory DDR4/2666 ECC + 1 SSD Micron M5100MAX with 480GB storage & 3 HD SAS 7200rpm with 12TB storage



- in 2024 [cicair] - SIE LADON MILAN 2 processors AMD EPYC 7763 server with 128 cores & 2 TB RAM memory DDR4/3200 ECC + 2 SSD CS900 SATA with 2TB storage & 2 Kioxia NVMe 12TB storage + 2 GPUs NVIDIA RTX 7,168 CUDA with 20GB VRAM GDDR6 & NVIDIA PRO5000 14,080 CUDA with 48GB VRAM GDDR7



3 • BRIEF OVERVIEW of our experience & scientific production in Computational Biology & Bioinformatics applied to Biomedical Research

We are a multidisciplinary team of biochemists, biotechnologists, computer scientists, informatic engineers, mathematician-statisticians as well as medical doctors, who work with complex, large-scale (“omic”) biomedical and biomolecular data and using computers.

We “speak”:

R, Python, JAVA, JavaScript, Ruby, MySQL, shell-script, HTML/PHP ...



3 • BRIEF OVERVIEW of our experience & scientific production in Computational Biology & Bioinformatics applied to Biomedical Research





Javier De Las Rivas 

 FOLLOW

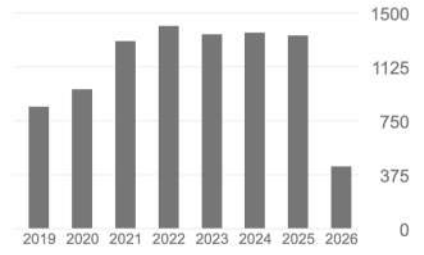
Cancer Research Center (IBMCC, CSIC/[USAL](#) & IBSAL), Salamanca, Spain
 Verified email at usal.es - [Homepage](#)

[Bioinformatics](#) [Functional Genomics](#) [Cancer Genomics](#) [Network Biology](#) [Interactomics](#)

<input type="checkbox"/>	TITLE  	CITED BY	YEAR
<input type="checkbox"/>	A proteome-scale map of the human interactome network T Rolland, M Taşan, B Charloteaux, SJ Pevzner, Q Zhong, N Sahni, S Yi, ... Cell 159 (5), 1212-1226	1645	2014
<input type="checkbox"/>	A reference map of the human binary protein interactome K Luck, DK Kim, L Lambourne, K Spirohn, BE Begg, W Bian, R Brignall, ... Nature 580 (7803), 402-408	1421	2020
<input type="checkbox"/>	Protein–protein interactions essentials: key concepts to building and analyzing interactome networks J De Las Rivas, C Fontanillo PLoS computational biology 6 (6), e1000807	969	2010
<input type="checkbox"/>	PSICQUIC and PSISCORE: accessing and scoring molecular interactions B Aranda, H Blankenburg, S Kerrien, FSL Brinkman, A Ceol, E Chautard, ... Nature methods 8 (7), 528-529	336	2011
<input type="checkbox"/>	Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis LA Corchete, EA Rojas, D Alonso-López, J De Las Rivas, NC Gutiérrez, ... Scientific reports 10 (1), 19737	314	2020

Cited by [VIEW ALL](#)

	All	Since 2021
Citations	16319	7224
h-index	64	35
i10-index	174	111



Public access [VIEW ALL](#)

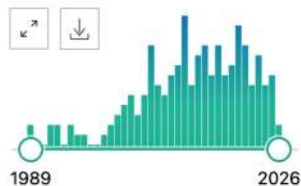
6 articles	119 articles
not available	available

Based on funding mandates

• BRIEF OVERVIEW of our experience & scientific publications



RESULTS BY YEAR



PUBLICATION DATE

- 1 year
- 5 years
- 10 years
- Custom Range

- Experimental assessment of AI-based interactome mapping.
1 Lambourne L, Yadav A, Wang Y, Desbuleux A, Kim DK, Laval F, Spirohn-Fitzgerald K, Cafarelli T, Pons C, Kovács IA, Jaikhani N, Schlabach S, De Ridder D, Luck K, Botchkarev VV Jr, Debnath O, Bian W, Shen Y, Yang Z, Mee MW, Helmy M, Jacob Y, Lemmens I, Rolland T, McClain GG, Coté AG, Gebbia M, Kishore N, Knapp JJ, Mellor JC, Memisoglu G, Reimand J, Tavernier J, Cusick ME, Zhong Q, Aloy P, Hao T, Charlotteaux B, Roth FP, **De Las Rivas J**, Falter-Braun P, Hill DE, Calderwood MA, Twizere JC, Vidal M.
Nat Commun. 2026 Apr 4. doi: 10.1038/s41467-026-70942-x. Online ahead of print.
PMID: 41935050 **Free article.**
- Deletion of 9p drives B-ALL through heterozygous inactivation of Pax5 and Cd72 in preleukemic cells.
2 Ruiz-Corzo B, Casado-García A, Oak N, Somoza-Cotillas P, López-Álvarez de Neyra A, Martínez-Cano J, Pérez-Pons A, Sánchez EG, Blanco O, Alonso-López D, **De Las Rivas J**, Riesco S, Prieto-Matos P, García Criado FJ, García Cenador MB, Orfao A, Ramírez-Orellana M, Coboleda C, Vicente-Dueñas C, Nichols KE, Sánchez-García I.
JCI Insight. 2026 Feb 17;11(7):e199464. doi: 10.1172/jci.insight.199464. eCollection 2026 Apr 8.
PMID: 41701537 **Free PMC article.**
- Pharmacogenomic Drug-Target Network Analysis Reveals Similarity Profiles Among FDA-Approved Cancer Drugs.
3 Berral-González A, Arroyo MM, Alonso-López D, Rivas-López MJ, Sánchez-Santos JM, **De Las Rivas J**.
Pharmaceutics. 2025 Nov 3;17(11):1421. doi: 10.3390/pharmaceutics17111421.
PMID: 41304759 **Free PMC article.**
- Challenges in the preclinical design and assessment of CAR-T cells.
4 Tomai R, **De Las Rivas J**, Fetica B, Bergantim R, Filipic B, Gagic Z, Nikolic K, Gulei D, Kegyes D, Nistor M, Muresan XM, Cenariu D, Feder R, Pavel-Tanasa M, Cianga A, Tigu AB, Munteanu R, Tanase A, Einsele H, Tomuleasa C.
Front Immunol. 2025 Aug 8;16:1564998. doi: 10.3389/fimmu.2025.1564998. eCollection 2025.
PMID: 40861448 **Free PMC article.** Review.



Nature Communications

<https://doi.org/10.1038/s41467-026-70942-x>

Article in Press

Experimental assessment of AI-based interactome mapping

Received: 19 August 2025

Accepted: 10 March 2026

Cite this article as: Lambourne, L., Yadav, A., Wang, Y. *et al.* Experimental assessment of AI-based interactome mapping. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-70942-x>

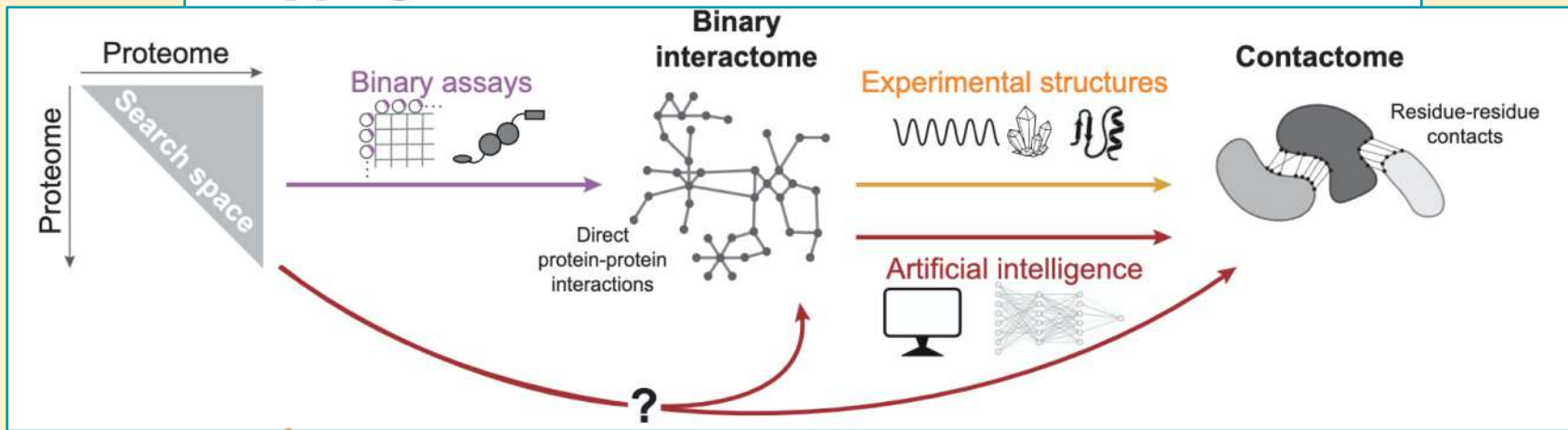
Luke Lambourne, Anupama Yadav, Yang Wang, Alice Desbuleux, Dae-Kyum Kim, Florent Laval, Kerstin Spirohn-Fitzgerald, Tiziana Cafarelli, Carles Pons, István A. Kovács, Noor Jailkhani, Sadie Schlabach, David De Ridder, Katja Luck, Vladimir V. Botchkarev Jr., Olivia Debnath, Wenting Bian, Yun Shen, Zhipeng Yang, Miles W. Mee, Mohamed Helmy, Yves Jacob, Irma Lemmens, Thomas Rolland, Gregory G. McClain, Atina G. Coté, Marinella Gebbia, Nishka Kishore, Jennifer J. Knapp, Joseph C. Mellor, Gonen Memisoglu, Jüri Reimand, Jan Tavernier, Michael E. Cusick, Quan Zhong, Patrick Aloy, Tong Hao, Benoit Charloteaux, Frederick P. Roth, Javier De Las Rivas, Pascal Falter-Braun, David E. Hill, Michael A. Calderwood, Jean-Claude Twizere & Marc Vidal

Nature Communications

<https://doi.org/10.1038/s41467-026-70942-x>

Article in Press

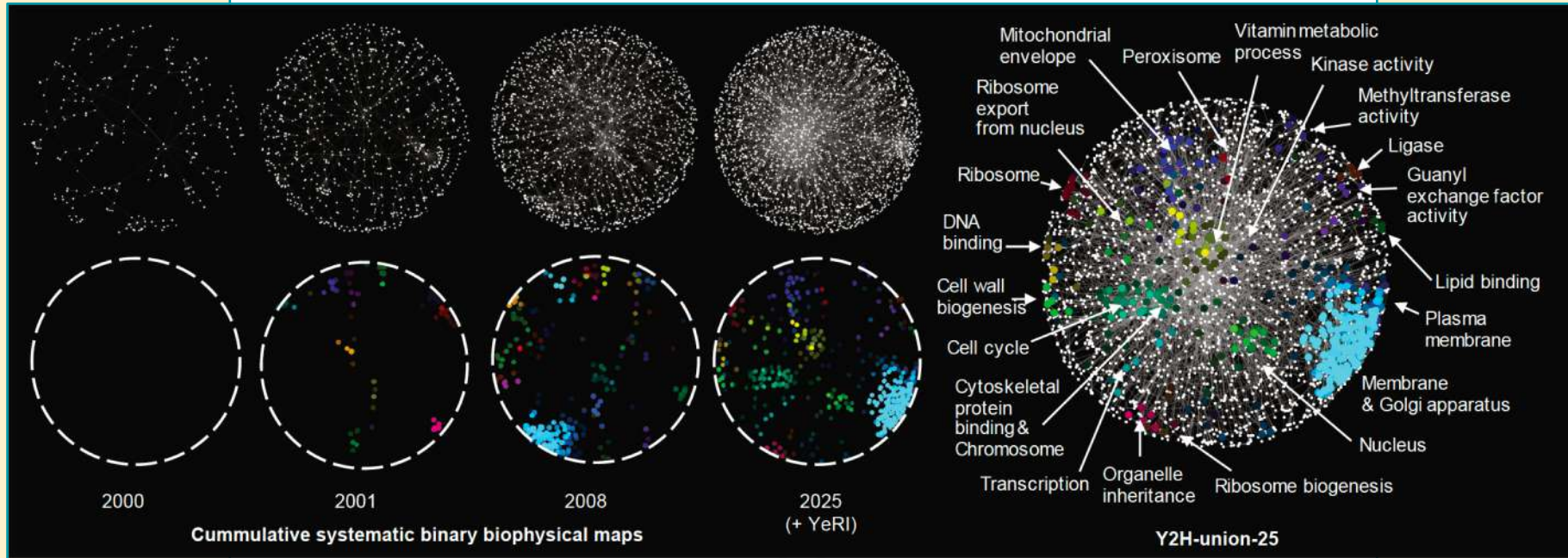
Experimental assessment of AI-based interactome mapping



Nature Communications

<https://doi.org/10.1038/s41467-026-70942-x>

Article in Press

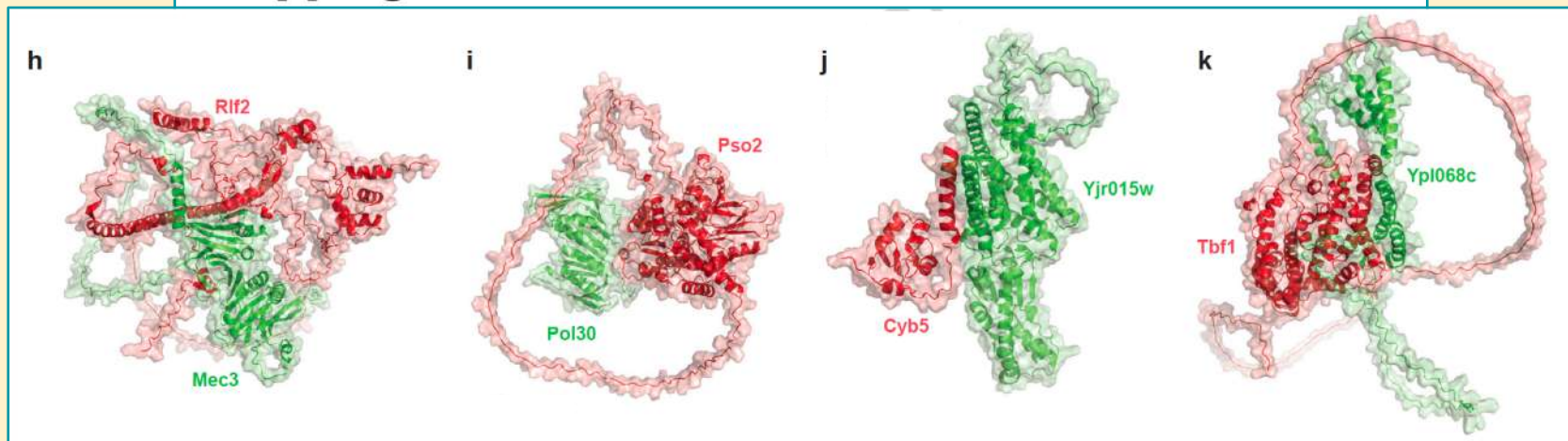


Nature Communications

<https://doi.org/10.1038/s41467-026-70942-x>

Article in Press

Experimental assessment of AI-based interactome mapping





1 • WHO ARE WE ? : a multidisciplinary team of **biochemists, biotechnologists, computer scientists, informatic engineers, mathematician-statisticians** as well as **medical doctors**, who work with complex **biomedical and biomolecular data** using computers

2 • WHAT TYPE OF COMPUTERS WE USE ? :
for **Biological data & Biomedical research**

3 • BRIEF OVERVIEW of our **experience & scientific production** in **Computational Biology & Bioinformatics** applied to **Biomedical Research**

4 • BRIEFING about the current development of the **AI** in **Biological data & Biomedical research**



4 • BRIEFING about the current development of the AI in Biological data & Biomedical research

These **LLModels** are trained on massive biological datasets and biological information to **learn the underlying “grammar” of life**, that is:

- **DNA** and **RNA** sequences and function
- **Genes** expression and function
- **Proteins** sequences, structures and functions, etc
- **Biomolecular** interactions/association (e.g. drug - protein target)
- **Cells, cell-types, cells lineages**, performance and evolution
- ...

The integration of **Artificial Intelligence (AI)** and **Large Language Models (LLMs)** into **bioinformatics and computational biology** has shifted in 2025-2026 from specialized, task-specific algorithms to **Foundation Models**.



4 • BRIEFING about the current development of the AI in Biological data & Biomedical research





Briefings in Bioinformatics, 2025, 26(4), bbaf357

<https://doi.org/10.1093/bib/bbaf357>

Review

Bridging artificial intelligence and biological sciences: a comprehensive review of large language models in bioinformatics

Anqi Lin^{1,†}, Junpu Ye^{1,†}, Chang Qi², Lingxuan Zhu³, Weiming Mou^{3,4}, Wenyi Gan⁵, Dongqiang Zeng^{6,7}, Bufu Tang⁸, Mingjia Xiao⁹,
Guangdi Chu¹⁰, Shengkun Peng¹¹, Hank Z.H. Wong¹², Lin Zhang^{13,14}, Hengguo Zhang ¹⁵, Xinpei Deng¹⁶, Kailai Li³, Jian Zhang³,
Aimin Jiang^{17,*}, Zhengrui Li^{18,*}, Peng Luo ^{1,19,*}



Contemporary applications and advances of LLMs in bioinformatics & biomedicine into 5 main domains:

These models are trained on massive biological datasets and biological information to learn the underlying “grammar” of life, that is: DNA and RNA sequences and function, genes expression and function, proteins sequences, structures and functions, etc.

DNA/RNA sequence analysis, functional prediction and structure prediction



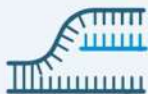
- HyenaDNA
- DNAGPT
- CombSAFE
- BERT-Enhancer
- DNABERT
- RoBERTa
- RNABERT
- GeoBoost2

Protein sequence analysis, functional prediction and structure prediction



- PTMGPT
- ProGen
- ESM
- ProtGPT
- EpiBERTope
- ELMo
- BepiPred-3.0
- TransPPMP
- ProteinBERT
- MSA Transformer
- BERT4Bitter

Multi-Omics Data Analysis



- BERT-TFBS
- scGPT
- GP-GPT
- pLM4Alg
- scELMo
- MuLan-Methyl
- iEnhancer-BERT
- scBERT
- TooT-PLM-ionCT
- UESoIDS
- iDNA-ABT
- POOE
- mtX-COBRA
- DeepGene Transformer

Computational Approaches in Drug Discovery and Design



- MolGPT
- ChemBERTa
- SMILES-BERT
- PISTON
- MRCF
- K-BERT
- PharmBERT
- DTI-BERT
- TransDTI

Biomedical Literature Mining



- Galactica
- BioBERT
- ClinicalBERT
- GENEVIC
- MarkerGenie
- LBERT
- GenCLiP 3
- BioBERT



4 • BRIEFING about the current development of the AI in Biological data & Biomedical research

These **LLModels** are trained on massive biological datasets and biological information to **learn the underlying “grammar” of life**, that is:

- **DNA** and **RNA** sequences and function
- **Genes** expression and function
- **Proteins** sequences, structures and functions, etc
- **Biomolecular** interactions/association (e.g. drug - protein target)
- **Cells, cell-types, cells lineages**, performance and evolution
- ...

The integration of **Artificial Intelligence (AI)** and **Large Language Models (LLMs)** into **bioinformatics and computational biology** has shifted in 2025-2026 from specialized, task-specific algorithms to **Foundation Models**.



4 • BRIEFING about the AI for Biological & Biomedical data

Received: 18 May 2024 | Accepted: 25 June 2024

DOI: 10.1002/qub.2.69

PERSPECTIVE



Foundation models for bioinformatics

Ziyu Chen^{1,2} | Lin Wei^{1,2} | Ge Gao^{1,2}

¹State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Biomedical Pioneering Innovative Center (BIOPIC) & Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), Peking University, Beijing, China

²Changping Laboratory, Beijing, China

Correspondence

Ge Gao.
Email: gaog@mail.cbi.pku.edu.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2022ZD0115004

Abstract

Transformer-based foundation models such as ChatGPTs have revolutionized our daily life and affected many fields including bioinformatics. In this perspective, we first discuss about the direct application of textual foundation models on bioinformatics tasks, focusing on how to make the most out of canonical large language models and mitigate their inherent flaws. Meanwhile, we go through the transformer-based, bioinformatics-tailored foundation models for both sequence and non-sequence data. In particular, we envision the further development directions as well as challenges for bioinformatics foundation models.

KEYWORDS

ChatGPT, foundation models, large language models, transformer



4 • BRIEFING about the AI for Biological & Biomedical data

INFORMATION SCIENCE

National Science Review

12: nwaf028, 2025

<https://doi.org/10.1093/nsr/nwaf028>

Advance access publication 25 January 2025

Foundation models in bioinformatics

Fei Guo^{1,2}, Renchu Guan^{ID 3}, Yaohang Li⁴, Qi Liu⁵, Xiaowo Wang^{ID 6}, Can Yang^{ID 7}
and Jianxin Wang^{ID 1,2,*}

ABSTRACT

With the adoption of foundation models (FMs), artificial intelligence (AI) has become increasingly significant in bioinformatics and has successfully addressed many historical challenges, such as pre-training frameworks, model evaluation and interpretability. FMs demonstrate notable proficiency in managing large-scale, unlabeled datasets, because experimental procedures are costly and labor intensive. In various downstream tasks, FMs have consistently achieved noteworthy results, demonstrating high levels of accuracy in representing biological entities. A new era in computational biology has been ushered in by the application of FMs, focusing on both general and specific biological issues. In this review, we introduce recent advancements in bioinformatics FMs employed in a variety of downstream tasks, including genomics, transcriptomics, proteomics, drug discovery and single-cell analysis. Our aim is to assist scientists in selecting appropriate FMs in bioinformatics, according to four model types: language FMs, vision FMs, graph FMs and multimodal FMs. In addition to understanding molecular landscapes, AI technology can establish the theoretical and practical foundation for continued innovation in molecular biology.

Keywords: foundation model, bioinformatics, genomics, transcriptomics, proteomics, drug discovery, single-cell analysis

Received: 18 May 2024 | Accepted: 25 June 2024

DOI: 10.1093/nsr/nwaf028

PERSPECTIVE



Foundation models for bioinformatics

Ziyu Chen^{1,2} | Lin Wei^{1,2} | Ge Gao^{1,2}

¹State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Biomedical Pioneer Innovative Center (BIOPIC) & Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), Peking University, Beijing, China

²Changping Laboratory, Beijing, China

Correspondence

Ge Gao,

Email: gao@mail.cbi.pku.edu.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2022ZD0115004

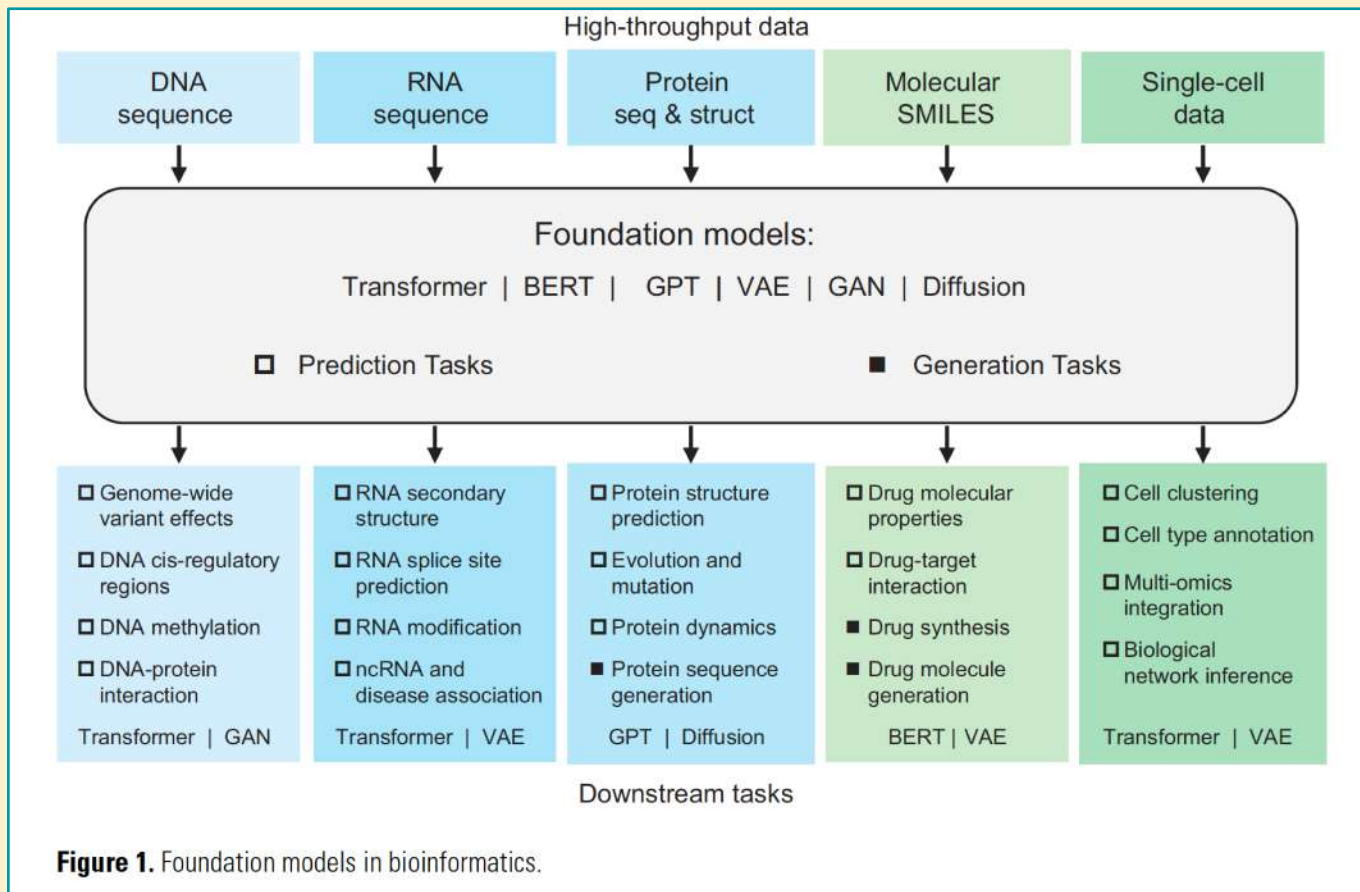
Abstract

Transformer-based foundation models such as ChatGPTs have revolutionized our daily life and affected many fields including bioinformatics. In this perspective, we first discuss about the direct application of textual foundation models on bioinformatics tasks, focusing on how to make the most out of canonical large language models and mitigate their inherent flaws. Meanwhile, we go through the transformer-based, bioinformatics-tailored foundation models for both sequence and non-sequence data. In particular, we envision the further development directions as well as challenges for bioinformatics foundation models.

KEYWORDS

ChatGPT, foundation models, large language models, transformer

4 • BRIEFING about the AI for Biological & Biomedical data





– **Setup and use of advanced HPC Servers for biocomputing, bioinformatics and data management in Biomedical Research**

Dr. Javier De Las Rivas

jrivas@usal.es

Cancer Research Center (CiC, CSIC/USAL)

Consejo Superior de Investigaciones Científicas (**CSIC**) & University of Salamanca (**USAL**)
Salamanca (Spain)

– **Current application of Artificial Intelligence tools and resources in Biomedical Research**

Dr. Alberto Berral Gonzalez

aberralgonzalez@usal.es

Cancer Research Center (CiC, CSIC/USAL)

Consejo Superior de Investigaciones Científicas (**CSIC**) & University of Salamanca (**USAL**)
Salamanca (Spain)