Giorgio Valentini – Curriculum vitae

SYNTHETIC CURRICULUM

Master Degree in Biology and Computer Science, PhD in Computer Science (University of Genoa). Full Professor at the Department of Computer Science, University of Milan (UNIMI). Member of the UNIMI Doctoral School of Informatics and scientific director for UNIMI of the European doctorate in Genomics and Bioinformatics in collaboration with the Joint Research Center of the European Union. Founder of AnacletoLab, Computational Biology and Bioinformatics Laboratory of the Department of Computer Science of the University of Milano. Principal Investigator of 15 national and international research projects funded by public and private institutions in the area of Bioinformatics, Artificial Intelligence and Big Data analytics. Member of ELLIS, the European Laboratory for Learning and Intelligent Systems; head of the UNIMI unit of CINI-InfoLife National Laboratory.
Member of the editorial board of Nature Scientific Reports and other bioinformatics and computer science journals.
In 2018 one of my papers has been awarded as one of the best papers of the year by the International Medical Informatics Association.
Research activity: development and application of Artificial Intelligence methods to bio-medical problems, with a special focus on Machine Learning methods for Personalized and Precision Medicine, Network Medicine and Systems Biology, in the context of collaborations with various European and American research centers and universities.
My most recent research concerns GPAI models and their trustworthy applications in biology and medicine. I am author of about 200 scientific publications with peer-review in journals, book chapters and international conferences in the field of Artificial Intelligence, Bioinformatics and Computational Biology.

ANALYTICAL CURRICULUM

Qualifications and courses of study

- PhD in Computer Science, University of Genoa (2003). Title: "Ensemble methods based on bias-variance analysis".
- Master Degree in Information Sciences, University of Genoa, grade: 110/110 cum laude.
- Master Degree in Biological Sciences, University of Genoa, grade: 110/110 cum laude.
- Classical Maturity, Classical Lyceum of Savona, grade: 56/60.

Research activity

My research lies at the intersection of artificial intelligence and bioinformatics and is driven by complex problems in molecular biology and medicine. By modeling these problems in a Machine Learning context, I design and develop new machine learning algorithms or adopt suitable existing algorithms to address relevant problems in the field of Bioinformatics, with a particular interest, especially in recent years, in the application of General Purpose Artificial Intelligence methods to Molecular Modeling, Medicine Genomics, and Precision and Personalized Medicine.
Although in my research activity the development of new Machine Learning methods is closely linked to real problems in the bio-medical field, I have also developed "pure" Machine Learning research lines, especially for the design and analysis of ensemble methods, graph-based semi-supervised algorithms, and Large Language Models.

Indices of scientific productivity

The productivity indexes from the Google Scholar database are : Numbero of citations: 6739; H – index: 42; i10 index: 101

MAIN SCIENTIFIC RESPONSIBILITY OF NATIONAL AND INTERNATIONAL RESEARCH PROJECTS

- PI of “AI digital technologies to support gene therapy and RNA drug analysis and discovery”, Spoke 7 “Biocomputing” of the PNRR “National Center for Gene Therapy and Drugs based on RNA Technology”, funded by the NextGenerationEU program of the EU (2023-2026)
- PI of “Adaptive AI methods for Digital Health (AIDH)”, project in the context of the PNRR Future Artificial Intelligence – FAIR project, funded by the NextGenerationEU program of the EU (2024-2025)
- Scientific coordinator for UNIMI of the EU Collaborative Doctoral Partnership in Genomics and Bioinformatics, funded by the European Commission in collaboration with the EU Joint Research Centers (2020-2025). The five-year collaboration contract is renewable at the end of the five-year period.
- Scientific responsible of the DAAD project “Graph Representation Learning methods for the analysis of biomedical heterogeneous graphs” funded by The German Ministry of University (2024)
- PI of the European PRACE project “ParBigMen: ParSMURF application to Big genomic and epigenomic data for the detection of pathogenic variants in Mendelian diseases”, funded by the European Union (2020-2022)
- Head of the UNIMI unit of the Finding-MS project within the European project ERA-PerMed Joint Transnational Call (JTC) 2018 (2020-2023)

I was PI of other 10 national and international projects in the field of AI applications in Biology and Medicine.
I participated also as member to EU funded Network of Excellence "Pattern Analysis, Statistical Modeling and Computational Learning (PASCAL and PASCAL2)" to the project “Statistical Learning Technologies for Digital Information Management Search”, funded by the National Scientific Foundation (NSF), USA and to other international and national research projects.

International collaborations:

I collaborate with my AnacletoLab research group with several research teams in Europe and in the United States. The main collaborations are the following:

- Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
- Department of Computer Science, Bioinformatics Centre for Systems and Synthetic Biology, Royal Holloway, University of London
- Berlin Institute of Health, Charitè University of Medicine, Berlin
- Department of Signal Theory and Communications, University Rey Juan Carlos, Madrid
- Division of Molecular Oncology, Department of Pathological Anatomy F. de Vitoria University (UFV), Madrid
- Division of Computer Science, National Technical University of Athens
- School of Electrical and Computer Engineering, National Technical University of Athens
- Institute of Medical Genetics and Applied Genomics at the University Hospital Tübingen, Germany
- Wellcome Trust Sanger Institute and the European Bioinformatics Institute (EBI) of Hinxton, UK
- European Molecular Biology Laboratory, Heidelberg, Germany
- Fundacao Getulio Vargas, Rio de Janeiro, Brazil
- Artificial Intelligence department of the University of Granada, Spain
- Computer Science Dept of Aalto University, Helsinki, Finland
- Queen Mary University of London
- Division of Env. Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
- Laboratory of Systems Pharmacology, Harvard Medical School, Boston, USA
- Department of Computer Science, Virginia Polytechnic Institute and State University, USA
- Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, USA
- IBM Research – Almaden, CA, USA
- Other national research groups both in the medical field (National Cancer Institute (INT), S. Raffaele, Humanitas and Policlinico Hospital of Milan, National Institute of Molecular Genetics) and Computer Science (University of Pisa, Palermo and Cagliari).

Participation in editorial committees of international journals

- Member of the Editorial board of Scientific Reports, Nature, (from 1 June 2018).
- Member of the editorial board of the Journal of Translational Medicine, Elsevier (from 1 Febuary 2022)

- Guest Editor of Artificial Intelligence in Medicine, Elsevier for the Special issue "Computational Intelligence and Machine Learning in Bioinformatics" (2008-2009).
- I was also a member of the Editorial board of other journals in the Bioinformatics and Artifcial Intelligence area from 2008 to 2024.

Conference and Workshop Organization and fellowships

I was chair of several bioinformatics conferences or organizer of workshops in interantional conferences, including ECML (European Conference on Machine Learning), ECAI (European Conference on Artifical intelligence), NIPS, CIBB (Computational Intelligence methods for Bioinformatics and Biostatistics), AIIA and others.

I was visiting researcher or visiting professor in several universities and research centers including Computer Science Dept. Oregon State University – USA, Computer Science Department of Royal Holloway, University of London, Principe Felipe Research Center, Valencia, Spain, Computer Science Department of the Aristotle University of Thessaloniki, Computer Science Department of Aalto Univeristy (Helsinki), European Center for Living Technologies (ECLT), Venezia, Charité – Universitätsmedizin Berlin, Departamento de Ciencias de la Computación y Inteligencia Artificial de la Universidad de Granada, National Technical University of Athtnes – Division of Computer Science, Berlin Institute of Health.

Participation in national and international scientific societies and associations

I am member of :
ELLIS - European Lab for Learning and Intelligent Systems
CINI AIIS - Artificial Intelligence and Intelligent Systems laboratory
Head of the UNIMI unit of the CINI InfoLife National laboratory
ISCB - International Society of Computational Biology
BITS - Italian Society of Bioinformatics
Data Mining and Big Data Analytics Technique Committee (DMTC) - IEEE Computational Intelligence Society
Scientific committee of the working group on Machine Learning and Data Mining, within AI * IA, Italian Association for Artificial Intelligence.
Data Science Research Center (DSRC) of the University of Milan

Selected recent Publications

(a full list is available at https://valentini.di.unimi.it/pub.html)

- *E Niyonkuru, J Caufield, L Carmody, M Gargano, S Toro, P Whetzel, H Blau, M Soto Gomez, E Casiraghi, L Chimirri, J Reese, G Valentini, M Haendel, C Mungall, P Robinson Leveraging generative AI to assist biocuration of medical actions for rare disease, Bioinformatics Advances, vol.5, issue 1, vbaf141, 2025*
- *F Torgano, M Soto-Gomez, M Zignani, J Gliozzo, E Cavalleri, M Mesiti, E Casiraghi, G Valentini RNA Knowledge-Graph analysis through homogeneous embedding methods, Bioinformatics Advances, vol.5, issue 1, 2025*
- *J Gliozzo, M Soto-Gomez, A Bonometti, A Patak, E Casiraghi, G Valentini miss-SNF: a multimodal patient similarity network integration approach to handle completely missing data sources, Bioinformatics, Volume vol. 41, Issue 4, April 2025*
- *M Nicolini, E Saitto, RE Jimenez Franco, E Cavalleri, AJ Galeano, D Malchiodi, A Paccanaro, PN Robinson, E Casiraghi and G Valentini Fine-tuning of conditional Transformers improves in silico enzyme prediction and generation, Computational and Structural Biotechnology Journal, vol. 27, pp. 1318-1334, 2025*
- *E Niyonkuru, M Soto-Gomez, E Casiraghi, S Antogiovanni, H Blau, J Reese, G Valentini and PN Robinson Replacing non-biomedical concepts improves embedding of biomedical concepts, PLoS One 20(5): e0322498, 2025*
- *J Gliozzo, M Soto-Gomez, V Guarino, A Bonometti, A Cabri, E Cavalleri, J Reese, PN Robinson, M Mesiti, G Valentini, E Casiraghi Intrinsic-Dimension Analysis for Guiding Dimensionality Reduction and Data Fusion in Multi-Omics Data Processing, Artificial Intelligence in Medicine, vol. 160, 2025*
- *M Soto-Gomez, C. Cano, J Reese, PN Robinson, G Valentini and E Casiraghi Biasing second-order random walk sampling for heterogeneous graph embedding , International Joint Conference on Neural Networks - IJCNN 2025*
- *F. Stacchietti, M. Nicolini, L. Chimirri, P. Robinson, E. Casiraghi and G. Valentini, Modular Deep Neural Networks with residual connections for predicting the pathogenicity of genetic variants in non coding genomic regions , Proc. of IWANN - 18th International Work-Conference on Artificial Neural Networks, LNCS, 2025*
- *M. Nicolini, F. Stacchietti, C. Cano, E. Casiraghi and G. Valentini A transformer-based model to predict micro RNA interactions , Proc. of IWANN - 18th Int. Work-Conference on Artificial Neural Networks, LNCS, 2025*

- *M. Nicolini, F. Stacchietti, E. Casiraghi and G. Valentini Computational Understanding of Pairwise Interactions in ncRNA Data , Proc. of CIBB - 20th conference on Computational Intelligence methods for Bioinformatics and Biostatistics, 2025*
- *E. Fassi, M. Nicolini, E. Saitto, G. Valentini, E. Casiraghi and G. Grazioso Protein Language Model-Generated Enzyme Sequences Exhibit high stability in Molecular Dynamics Simulation , Proc. of CIBB, 2025*
- *R. Jiménez, A. Galeano, M. Báez, S. Ferreyra, G. Melo, G. Valentini, E. Casiraghi, L. Cernuzzi and A. Paccanaro An Empirical Study of Remote Homology Detection using Protein Language Models , Proc. of CLEI, IEEE 2025*
- *E Cavalleri, A Cabri, M Soto-Gomez, S Bonfitto, P Perlasca, J Gliozzo, TJ Callahan, J Reese, PN Robinson, E Casiraghi, G Valentini, M Mesiti An ontology-based knowledge graph for representing interactions involving RNA molecules, Nature Scientific Data, 11, 906, 2024*
- *B Coleman, E Casiraghi, TJ Callahan, H Blau, L Chan, B Laraway, K Clark, Y Re'em, K Gersing, KJ Wilkins, N Harris, G Valentini, M Haendel, J Reese, PN Robinson Association of post-COVID phenotypic manifestations with new-onset psychiatric disease, Nature Transl Psychiatry, 14, 246, 2024*
- *TJ Callahan, … G Valentini, M Mesiti, LA Gillenwater, B Santangelo, NA Vasilevsky, R Hoehndorf, TD Bennett, PB Ryan, G Hripcsak, MG Kahn, M Bada, WA Baumgartner Jr An open source knowledge graph ecosystem for the life sciences, Nature Scientific Data, 11, 363, 2024*
- *J Reese, L Chimirri, ..., E Casiraghi, G Valentini, J Jacobsen, M Haendel, D Smedley, C Mungall, PN Robinson, Systematic benchmarking demonstrates large language models have not reached the diagnostic accuracy of traditional rare-disease decision support tools, MedRxiv, 2024/11/7, 2024*
- *J Reese, D Danis, J HCaufield, T Groza, E Casiraghi, G Valentini, CMungall, PN Robinson, On the limitations of large language models in clinical diagnosis, MedRxiv, 2023.07. 13.23292613, 2024*
- *L Chan, E Casiraghi, J Reese, Q Harmon, K Schaper, H Hegde, G Valentini, C Schmitt, A Motsinger-Reif, J Hall, C Mungall, P Robinson, M Haendel Predicting nutrition and environmental factors associated with female reproductive disorders using a knowledge graph and random forests, International Journal of Medical Informatics, 187, 105461, 2024*
- G. Valentini Exploring the similarity between genetic diseases improves their differential diagnosis and the understanding of their etiology, *Eur J Hum Genet, 2024*
- *L Cappelletti, …, TI Oprea, J Reese, G Valentini, PN Robinson Node-degree aware edge sampling mitigates inflated classification performance in biomedical random walk-based graph representation learning, Bioinformatics Advances, Volume 4, Issue 1, Oxford University Press, 2024*
- *G. Valentini, D. Malchiodi, J. Gliozzo, M. Mesiti, M. Soto-Gomez, A. Cabri, J. Reese, E. Casiraghi, P. Robinson The promises of large language models for protein design and modeling, Frontiers in Bioinformatics, vol. 3, 2023*
- *B. Antony, H. Blau, E. Casiraghi, ..., G. Valentini, A. Williams, P. Robinson, J. Reese, T. Murali Predictive models of long COVID, eBioMedicine - The Lancet Discovery Science, vol. 96, 104777, 2023*
- *L. Cappelletti, T. Fontana, E. Casiraghi, V. Ravanmehr, T. J.Callahan, C. Cano, M. P. Joachimiak, C. J. Mungall, P. N. Robinson, J. Reese, G. Valentini GRAPE for Fast and Scalable Graph Processing and Random Walk-based Embedding, Nature Computational Science 3, 552–568, 2023*
- *E. Casiraghi and G. Valentini A software resource for large graph processing and analysis, Nature Computational Science 3, 2023*
- *G. Karlebach, L. Carmody, J. C. Sundaramurthi, E. Casiraghi, P. Hansen, J. Reese, C. J. Mungall, G. Valentini, P. N. Robinson An expectation-maximization framework for comprehensive prediction of isoform-specific functions, Bioinformatics, Oxford University Press, 39(4), btad1322023, April 2023*
- *E. Casiraghi, R. Wong, M. Hall, B. Coleman, M. Notaro, M. D. Evans, J. S. Tronieri, H. Blau, B. Laraway, T. J. Callahan, L. E. Chan, C. T. Bramante, J. B. Buse, R. A. Moffitt, T. Stürmer, S. G. Johnson, Y. R. Shao, J. Reese, P. N. Robinson, A. Paccanaro, G. Valentini, J. D. Huling, K. J. Wilkins A method for comparing multiple imputation techniques: a case study on the U.S. National COVID Cohort Collaborative, Journal of Biomedical Informatics, 139:104295, 2023*
- *J. T. Reese, ..., G. Valentini, D, Sahner, C, G. Chute, C, Madlock-Brown, M. A Haendel, P. N. Robinson on behalf of the N3C Consortium Generalizable Long COVID Subtypes: Findings from the NIH N3C and RECOVER Programs, eBioMedicine - The Lancet Discovery Science, Vol. 87, 2023*
- *L. Cappelletti, A. Petrini, J. Gliozzo, E. Casiraghi, M. Schubach, M. Kircher and G. Valentini Boosting tissue-specific prediction of active cis-regulatory regions through deep learning and Bayesian optimization techniques, BMC Bioinformatics, 23:154, 2022*
- *J. Gliozzo, M. Mesiti, M. Notaro, A. Petrini, A. Patak, A. Puertas-Gallardo, A. Paccanaro, G. Valentini and E. Casiraghi Heterogeneous data integration methods for patient similarity networks, Briefings in Bioinformatics, Oxford University Press, Jul 18;23(4):bbac207, 2022*

- *V. Ravanmehr, H. Blau, …, E. Casiraghi, G. Valentini, C. Mungall, T, Oprea and P. Robinson Supervised learning with word embeddings derived from PubMed captures latent knowledge about protein kinases and cancer, NAR Genomics and Bioinformatics, Oxford Academic, 3(4), 2021*
- *M. Notaro, M. Frasca, A. Petrini, J. Gliozzo, E. Casiraghi, P.N. Robinson and G. Valentini HEMDAG: a family of modular and scalable hierarchical ensemble methods to improve Gene Ontology term prediction, Bioinformatics, 37(23), 2021*
- *D. Danis, J.O.B. Jacobsen, L. Carmody, M.A. Gargano, J.A. McMurry, A. Hegde, M.A. Haendel, G. Valentini, D. Smedley, P.N. Robinson Interpretable prioritization of splice variants in diagnostic next-generation sequencing, The American Journal of Human Genetics, Cell Press, 108(9), pages 1564-1577, 2021*
- *A. Petrini, M, Mesiti, M. Schubach, M. Frasca, D. Danis, M. Re, G. Grossi, L. Cappelletti, T. Castrignano', P. Robinson, G. Valentini parSMURF, a High Performance Computing tool for the genome-wide detection of pathogenic variants, GigaScience, Oxford Academic 9(5), 2020*
- *S. Vascon, M. Frasca, R. Tripodi, G. Valentini, M. Pelillo Protein Function Prediction as a Graph-Transduction Game, Pattern Recognition Letters vo. 134 pp. 96-105, 2020*
- *J. Gliozzo, P. Perlasca, M. Mesiti, E. Casiraghi, V. Vallacchi, E. Vergani, M. Frasca, G. Grossi, A. Petrini, M. Re, A. Paccanaro and G. Valentini Network modeling of patients' biomolecular profiles for clinical phenotype/outcome prediction, Scientific Reports, Nature Publishing 10:3612, 2020*
- *N. Zhou, Y. Jiang, T. Bergquist, ..., G. Valentini, ... P. Radivojac, I. Friedberg The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens, Genome Biology 20, article number: 244, 2019*
- *M. Notaro, M. Schubach, P.N. Robinson, G. Valentini Prediction of Human Phenotype Ontology terms by means of hierarchical ensemble methods, BMC Bioinformatics, vol. 18 (1), 2017 doi.org/10.1186/s12859-017-1854-y*
- *M. Schubach, M. Re, P.N. Robinson and G. Valentini Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants, Scientific Reports, Nature Publishing, 7:2959, 2017.*
- *D. Smedley, M, Schubach, J. Jacobsen, S. Kohler, T. Zemojtel, M. Spielmann, M. Jager, H. Hochheiser, N. Washington, J. McMurry, M. Haendel, C. Mungall, S. Lewis, T. Groza, G. Valentini and P.N. Robinson A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease, The American Journal of Human Genetics, 99:3, pp.595--606, September 2016. doi.org/10.1016/j.ajhg.2016.07.005*
- *Y. Jiang, P. Oron, ... G. Valentini, ... I. Friedberg and P. Radivojac An expanded evaluation of protein function prediction methods shows an improvement in accuracy, Genome Biology, 17:184 September 2016.*
- *G. Valentini, G. Armano, M. Frasca, J. Lin, M. Mesiti and M. Re RANKS: a flexible tool for node label ranking and classification in biological networks, Bioinformatics, 32(18), September 2016*
- *M. Mesiti, M. Re, G. Valentini Think globally and solve locally: secondary memory-based network learning for automated multi-species function prediction, GigaScience, 3:5, 2014*
- *G. Valentini, A. Paccanaro, H. Caniza, A. Romero, M. Re, An extensive analysis of disease-gene associations using network integration and fast kernel-based gene prioritization methods, Artificial Intelligence in Medicine, 61(2), 2014*
- *I. Cattinelli, G. Valentini, E. Paulesu, A. Borghese A Novel Approach to the Problem of Non-uniqueness of the Solution in Hierarchical Clustering , IEEE Transactions on Neural Networks and Learning Systems 24(7) pp.1166-1173, July 2013*
- *M. Frasca, A. Bertoni, M. Re, and G. Valentini, A neural network algorithm for semi-supervised node label learning from unbalanced data, Neural Networks 43, pp.84-98, July 2013*

Milano 21/8/2025

Giorgio Valentini