


PERSONAL INFORMATION

Fabio Cumbo



 Via Innocenzo XI, 8 – 00165 RM – Rome – Italy

 +39 335 669 1448

 fabio.cumbo@gmail.com

 <https://cumbo.me/>

 Skype [fabiocumbo](#)

Sex Male | Date of birth December 30, 1989 | Nationality Italian

WORKING EXPERIENCES

February 2019 – February 2020

Research Associate

Institute for Systems Analysis and Computer Science "Antonio Ruberti" - National Research Council of Italy - Rome - Italy

November 2018 – **Ongoing**

Postdoctoral Researcher

Segata Lab - Computational Metagenomics Laboratory - Department of Cellular, Computational, and Integrative Biology (CIBIO) - University of Trento - Trento, Italy

Development of new software tools for the characterisation of bacterial species
EU-ERC (MetaPG-716575)

March 2018 – December 2018

Professional Collaborator

ACTOR (Analytics, Control Technologies and Operations Research) S.R.L. – Rome – Italy

- Development of a technological platform to establish an early and non-invasive diagnosis of neurodegenerative diseases;
- Extraction and standardization of data from the IDA (Image and Data Archive) database powered by LONI (Laboratory of Neuro Imaging) funded by NIH and NIBIB;
- Creation of an ontology in order to better understand how these data are organized and to create an easy access service to the data themselves.

Collaborations (i) EBRI (European Brain Research Institute) – Rita Levi Montalcini Foundation – Rome – Italy, and (ii) Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy (CNR) – Rome – Italy

Keywords Bioinformatics – Ontologies – Machine Learning – Alzheimer and Parkinson’s disease – Diagnostics

April 2018 – September 2018

PhD Fellow

Institut für Informatik of the Albert-Ludwigs-Universität Freiburg –Freiburg im Breisgau – Germany

- Development of bioinformatics tools for the Galaxy platform

September 2017 – March 2018

Research Lab Assistant

Wartik Laboratory – Department of Biochemistry and Molecular Biology – The Pennsylvania State University – University Park Campus – 16802 PA – Pennsylvania – USA

- Development of a portal which allows to fast query massive sequence datasets using the Sequence Bloom Trees

Collaborations (i) The Galaxy Team and (ii) the Paul Medvedev’s Lab

Keywords Bioinformatics – Galaxy – Information Retrieval – Sequence Bloom Tree

March 2017 – September 2017

Intern

Wartik Laboratory – Department of Biochemistry and Molecular Biology – The Pennsylvania State University – University Park Campus – 16802 PA – Pennsylvania – USA

- Collaboration with the team of the Galaxy project to extend and implement new features for the Galaxy platform.
- Development of new statistical analysis and algorithms, and contribute to the development of Galaxy, Conda, and Bioconda projects.

Collaborations (i) The Galaxy Team and (ii) the Paul Medvedev’s Lab

Keywords Bioinformatics – Galaxy – Functional Data Analysis – Information Retrieval

February 2017 – November 2018

Professional Collaborator

Department of Engineering – International Telematic University UNINETTUNO – Rome – Italy

- Development of a software to automatically extract, extend, and standardize clinical and genomic data from the Genomic Data Commons Portal;
- This project is part of the Data-Driven Genomic Computing (GeCo), focusing on tertiary analysis for genomic data integration, and funded with an ERC Advanced Grant (September 2016 – August 2021)

Collaborations (i) Department of Electronics, Information and Bioengineering of the Polytechnic University of Milan and (ii) Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy

Keywords Bioinformatics – ERC – GeCo – TCGA2BED

September 2016 – November 2018

Teaching Assistant

Department of Engineering – International Telematic University UNINETTUNO – Rome – Italy

- Proposal of a training plan for the new Master’s Degree courses in Software Engineering (Big Data branch): (i) “Introduction to Big Data” and (ii) “Big Data Analytics and Visualization”;
- Teaching assistant and Lecturer for both the “Introduction to Big Data” and “Big Data Analytics and Visualization” courses;
- Proposed theses:

- “*Analysis and implementation of a web platform for the management and querying of genomic Big Data*” (Bachelor’s Degree): Candidate “Lorenzo Di Nardo”, Supervisor “Prof. Emanuel Weitschek”, Co-supervisor “Fabio Cumbo”;
- “*The structure of the Bloom Filters for the management and querying of Big Data*” (Master’s Degree): Candidate “Antonio Tranchida”, Supervisor “Prof. Emanuel Weitschek”, Co-supervisor “Fabio Cumbo”;
- “*Probabilistic data structures for the reference-free alignment of sequences*” (Master’s Degree): Candidate “Federico Ferranti”, Supervisor “Prof. Emanuel Weitschek”, Co-supervisor “Fabio Cumbo”;
- “*Hyperdimensional Computing for the Supervised Machine Learning*” (Master’s Degree): Candidate “Simone Truglia”, Supervisor “Prof. Emanuel Weitschek”, Co-supervisor “Fabio Cumbo”

Keywords Hadoop – Spark – MapReduce – Machine Learning – D3.js – Data Visualization

September 2016 – March 2017

Professional Collaborator

Marine Technology Research Institute of the National Research Council of Italy – INSEAN-CNR – Rome – Italy

- Development of a database containing data about military and merchant ships in which were used amiantus as a thermal insulator and data about officers and machinists affected by mesothelioma

Keywords Amiantus – Database – Mesothelioma – Military and Merchant Ships

October 2015 – November 2018

PhD Student

Department of Engineering – University of Roma Tre – Rome – Italy

- Development of an innovative platform for the acquisition, storage, management, integration, and analysis of heterogeneous biomedical data;
- Proposed theses:

- “*Analysis and development of a web service for the computation, visualization, and comparison of gene co-expression networks*” (Bachelor’s Degree): Candidate “Dalila Rosati”, Supervisor “Prof. Maurizio Patrignani”, Co-supervisor “Fabio Cumbo”;
- “*TCGAinBED Web: Managing and querying genomic Big Data*” (Bachelor’s Degree): Candidate “Luca Wissel”, Supervisor “Prof. Maurizio Patrignani”, Co-supervisor “Fabio Cumbo”

Collaborations (i) Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy and (ii) SYSBIO.IT – Center for Systems Biology – Milan – Italy

March 2015 – September 2015 Professional Collaborator

Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy

- Analysis and development of toolkits for bioinformatics analysis;
 - Extraction, storage, and management of genomic data from TCGA
- Keywords** Bioinformatics – The Cancer Genome Atlas – Data Extraction

December 2014 – November 2018 Professional Collaborator

SYSBIO.IT – Center for Systems Biology – Milan – Italy

- Development of COSYS, a platform for the interoperability of different software tools for a Systems Biology oriented analysis;
- The platform guarantees the data sharing between researchers of different European research centers;
- Part of ISBE (Infrastructure for Systems Biology – Europe), a large-scale European research infrastructure project on the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap

Collaborations (i) Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy and (ii) Department of Informatics, Systems and Communication (DISCo) of the University of Milano-Bicocca

Keywords Systems Biology – COSYS – ISBE

September 2014 – March 2015 Professional Collaborator

Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy

- Design and implementation of a software for the storage, management, and querying of genomic and clinical data;
- Application of the software to The Cancer Genome Atlas

Keywords Bioinformatics – The Cancer Genome Atlas – Data Extraction

February 2014 – February 2016 Professional Collaborator

Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy

- Development of a data extraction and analysis tool of genomic data from The Cancer Genome Atlas;
- Part of the Data-Centric Genomic Computing (GenData 2020) project funded by the Ministry of Education, University, and Research of Italy under the PRIN program

Collaborations Department of Electronics, Information and Bioengineering of the Polytechnic University of Milan

Keywords Bioinformatics – The Cancer Genome Atlas – Data Extraction

February 2013 – August 2014 Professional Collaborator

Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy

- Software development for the analysis of Protein-Protein Interaction Networks (PPI);
- Analysis of significant changes in the structure of protein complexes starting from temporal gene expression microarray data for the transgenic Mouse organism affected by Alzheimer’s disease

Collaborations EBRI (European Brain Research Institute) – Rita Levi Montalcini Foundation – Rome – Italy

Keywords Bioinformatics – AD11 Mouse Model – Alzheimer’s Disease – Microarray – Time Dynamics – Protein Complexes – CORUM – PPI

September 2011 – November 2012 Intern

Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy – Rome – Italy

- Design and development of algorithms for the computation of characteristic parameters in biological networks

Keywords Bioinformatics – PPI – Cytoscape – Network Theory

EDUCATION AND TRAINING

- October 2015 – April 2019 **PhD in Computer Science and Automation Engineering**
 Department of Engineering – University of Roma Tre – Rome – Italy
 Score: Excellent
- October 2012 – August 2014 **Master of Science Degree in Software Engineering**
 Department of Engineering – University of Roma Tre – Rome – Italy
 Score: 102/110
- October 2008 – October 2012 **Bachelor’s Degree in Software Engineering**
 Department of Engineering – University of Roma Tre – Rome – Italy
 Score: 89/110

PERSONAL SKILLS

Mother tongue(s) Italian

Other language(s)	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	B2	B2	B2	B2	B2

Levels: A1/A2: Basic user - B1/B2: Independent user - C1/C2 Proficient user
[Common European Framework of Reference for Languages](#)

- Job-related skills**
- Excellent knowledge of software engineering methods and respective tools (e.g.: UML) practicing in project design and proposals;
 - Familiar with agile methods of software development;
 - Excellent knowledge of the following programming and query languages: Java, J2EE, Python, PHP, R, HTML, CSS, SQL, XML, XPath, XQuery, and UML;
 - Good knowledge of virtual servers and networks setup with Hyper-V, VMWare Workstation, and Netkit;
 - Proficiency knowledge of PostgreSQL, MySQL, SQLite and the embedded H2 relational DBMSs;
 - Proficiency knowledge of MongoDB and Neo4j noSQL DBMSs;
 - Good knowledge of Apache and Tomcat web servers;
 - Excellent knowledge of the network analysis and visualization tools Cytoscape and Gephi;
 - Excellent knowledge of Weka software for Data Mining;
 - Good knowledge of the Apache Mahout, a scalable machine-learning library;
 - Good knowledge of the Apache Solr framework;
 - Excellent knowledge of the Apache Hadoop framework and the MapReduce programming paradigm;
 - Excellent knowledge of parallel and concurrent programming;
 - Excellent knowledge of versioning control softwares (SVN and GIT);
 - Excellent knowledge of the Galaxy framework;
 - Excellent knowledge of the TeX typesetting language

ADDITIONAL INFORMATION

- | | |
|--------------|--|
| Certificates | <ul style="list-style-type: none"> ▪ Certificate for the online course “Data Mining with Weka” provided by the Department of Computer Science of the University of Waikato, Hamilton, New Zealand ▪ Microsoft 070-640 Certificate: TS: Windows Server 2008, Active Directory, Configuring ▪ Microsoft 070-646 Certificate: Pro: Windows Server 2008, Server Administrator ▪ Microsoft 070-642 Certificate: TS: Windows Server 2008, Network Infrastructure, Configuring ▪ ECDL Certificate (European Computer Driving License) |
| Seminars | <ul style="list-style-type: none"> ▪ Speaker at the YES@IASI (Young Experts Seminar at IASI-CNR) event with a talk about “Fast querying of massive sequence datasets” – March 9, 2018 at the Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy |
| Conferences | <ul style="list-style-type: none"> ▪ Participant to the “Bioinformatics Italian Society 2015” (BITS 2015) with a poster about “The Cancer Genome Atlas Data Querying Tool” at the University of Milano-Bicocca – June 2015 ▪ Speaker at the “13th International Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics” at the University of Stirling, UK, with a talk titled “IRIS-TCGA: An Information Retrieval and Integration System for Cancer Genomic Data” – September 2016 ▪ Poster about “GDCWebApp: filtering, extracting, and converting genomic and clinical data from the Genomic Data Commons portal” at the Cold Spring Harbor Meeting on Genome Informatics – Cold Spring Harbor, Ny, USA – September 2017 ▪ Participation to the “9th International Workshop on Biological Knowledge Discovery from Big Data” (BIOKDD’18) in the context of the “29th International Conference on Database and Expert Systems Applications” (DEXA’18) in Regensburg, Germany – September 2018 ▪ Participation to the “ISMB/ECCB 2019 - International Society for Computational Biology” conference held in Basel, Switzerland with a Long Talk about “MetaRefSGB: a scalable framework to organize genomes from metagenomes and their annotations into species-level genome bins” – July 2019 ▪ Poster about “An Ontology to Organize Data on Alzheimer’s Disease from International Databases to Support Integrated Analysis” at the ISMB/ECCB 2019 - International Society for Computational Biology – Basel, Switzerland – July 2019 |
| PhD Schools | <ul style="list-style-type: none"> ▪ Participant to the “2nd SyBSyM Lake Como School – Systems Biology and Systems Medicine: Toward a Precision Medicine” at Villa del Grumello, Como, Italy – September 2016 ▪ Scientific Committee member and speaker at the “1st SYSBIO.IT School on Computational Systems Biology” at the University of Milano-Bicocca with a talk titled “Introduction to COSYS platform” about a web platform able to manage, simulate, analyse, and visualize biochemical models – June 2016 ▪ Participant to the “2nd School on Scientific Data Analytics and Visualization” at the CINECA center of Bologna, Italy – May 2016 ▪ Participant to the “3rd SYSBIO.IT School in Computational Biology” at the Institute for Systems Analysis and Computer Science “Antonio Ruberti” – National Research Council of Italy, Rome, Italy – May 2018 |
| Workshops | <ul style="list-style-type: none"> ▪ Participant to the course “Parallel I/O and Management of Large Scientific Data” organized by CINECA at the CINECA center of Rome, Italy – May 2015 ▪ Participant to the workshop “Hands On Big Data: Getting Started With NoSQL and Hadoop” organized by the Codemotion at the Polo Didattico in Rome, Italy – April 2015 |

Publications

- E. Cappelli, F. Cumbo, A. Bernasconi, A. Canakoglu, S. Ceri, M. Masseroli, E. Weitschek, "OpenGDC: unifying, modeling, integrating cancer genomic data and clinical metadata", MDPI Applied Sciences, 2020
- F. Cumbo, E. Weitschek, "An in-memory cognitive-based hyperdimensional approach to accurately classify DNA-Methylation data of cancer", The 11th International Workshop on Biological Knowledge Discovery from Big Data (BIOKDD 2020), Communications in Computer and Information Science, vol 1285, Springer, Cham, 2020
- E. Pasolli, F. De Filippis, I. Mauriello, F. Cumbo, A. Walsh, J. Leech, P. Cotter, N. Segata, D. Ercolini, "Large-scale genome-wide analysis links lactic acid bacteria from food with gut microbiome", Nature Communications, 2020
- F. Asnicar, A.M. Thomaz, F. Beghini, C. Mengoni, S. Manara, P. Manghi, Q. Chu, M. Bolzan, F. Cumbo, U. May, J.G. Sanders, M. Zolfo, E. Kopylova, E. Pasolli, R. Knight, S. Mirarab, C. Huttenhower, N. Segata, "Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0", Nature Communications, 2020
- E. Cappelli, E. Weitschek, F. Cumbo, "Extending knowledge on genomic data and metadata of cancer by exploiting taxonomy-based relaxed queries on domain-specific ontologies", in press on the post-proceedings of The 16th International Conference on Computational Intelligence methods for Bioinformatics and Biostatistics (CIBB 2019)
- S. Manara, F. Asnicar, F. Beghini, D. Bazzani, F. Cumbo, M. Zolfo, E. Nigro, N. Karcher, P. Manghi, M. I. Metzger, E. Pasolli, N. Segata, "Microbial genomes from gut metagenomes of non-human primates expand the primate-associated bacterial tree-of-life with over 1,000 novel species", BMC Genome Biology volume 20, Article number: 299 (2019)
- E. Cappelli, E. Weitschek, F. Cumbo, "Smart persistence and accessibility of genomic and clinical data", The 10th International Workshop on Biological Knowledge Discovery from Big Data (BIOKDD'19), Communications in Computer and Information Science, vol 1062. Springer, Cham, 2019
- L. Fiumi, F. Cumbo, C. Crenca, D. Gallo, C. Meoni, "The AMINAVI database: Know the presence of asbestos on board of ships, in the past and in the present", Epidemiologia & Prevenzione, June 2019
- C. Cumbo, F. Cumbo, "GMS – Gammadiae Management System: cataloguing and interpretation project of the so-called gammadiae starting from the iconographic evidences in the Roman catacombs", Conservar Patrimonio, January 2019
- E. Weitschek, F. Cumbo, E. Cappelli, G. Felici, P. Bertolazzi, "Classifying Big DNA Methylation Data: A Gene-Oriented Approach", Communications in Computer and Information Science vol. 903, DEXA-BIOKDD 2018, 9th International Workshop on Biological Knowledge Discovery from Data, Springer, October 2018
- R. Dale, B. Gruning, A. Sjobin, J. Rowe, B. A. Chapman, C. H. Tomkins-Tinch, R. Valieris, The Bioconda Team, J. Koster, "Bioconda: A Sustainable and Comprehensive Software Distribution for the Life Sciences", Nature Methods 2018
- F. Celli, F. Cumbo, E. Weitschek, "Classification of Large DNA Methylation Data Sets for Identifying Cancer Drivers: BIGBIOCL", Big Data Research 2018
- M. A. Cremona, A. Pini, F. Cumbo, K. D. Makova, F. Chiaromonte, S. Vantini, "IWTomics: Testing High-Resolution "Omics" Data at Multiple Locations and Scales", Bioinformatics 2018
- F. Cumbo, D. Vergni, D. Santoni, "Investigating Transcription Factor Synergism in Humans", DNA Research 2017
- F. Cumbo, M. S. Nobile, C. Damiani, R. Colombo, G. Mauri, P. Cazzaniga, "COSYS: A Computational Infrastructure for Systems Biology", Lecture Notes in Bioinformatics vol. 10477, Computational Intelligence Methods for Bioinformatics and Biostatistics, Springer, October, 2017
- F. Cumbo, E. Weitschek, P. Bertolazzi, G. Felici, "IRIS-TCGA: An Information Retrieval and Integration System for Genomic Data of Cancer", Lecture Notes in Bioinformatics vol. 10477, Computational Intelligence Methods for Bioinformatics and Biostatistics, Springer, October, 2017
- F. Cumbo, G. Fiscon, M. Masseroli, S. Ceri, E. Weitschek, "TCGA2BED: Extracting, Extending, Integrating, and Querying The Cancer Genome Atlas", BMC Bioinformatics 2016
- E. Weitschek, F. Cumbo, E. Cappelli, G. Felici, "Genomic Data Integration: A Case Study on Next Generation Sequencing of Cancer", 27th International Workshop on Database and Expert Systems Applications (DEXA) 2016
- E. Weitschek, F. Cumbo, G. Fiscon, V. Cestarelli, S. Ceri, M. Masseroli, "TCGA2BED and CAMUR for cancer NGS data processing", F1000Research, 1899-1899 (2016)
- F. Cumbo, G. Fiscon, S. Ceri, M. Masseroli, E. Weitschek, "TCGA2BED: converting and querying The Cancer Genome Atlas", BITS 2016: 13th Annual Meeting of the Bioinformatics Italian Society, 28-29 (2016)
- I. Arisi, M. D'Onofrio, R. Brandi, A. Cattaneo, P. Bertolazzi, F. Cumbo, G. Felici, C. Guerra, "Time Dynamics of Protein Complexes in the AD11 Transgenic Mouse Model for Alzheimer's Disease Like Pathology", BMC Neuroscience 16, n. 1 (2015): 28
- F. Cumbo, G. Fiscon, S. Ceri, M. Masseroli, E. Weitschek, "The Cancer Genome Atlas data querying tool", BITS 2015: 12th Annual Meeting of the Bioinformatics Italian Society, 120-121 (2015)
- F. Cumbo, G. Felici, P. Bertolazzi, "Selecting Relevant Nodes and Structures in Biological Networks. BiNAT: A New Plugin for Cytoscape", F1000Research 2014, 3:287
- F. Cumbo, P. Paci, D. Santoni, L. Di Paola, A. Giuliani, "GIANT: A Cytoscape Plugin for Modular Networks", PLOS ONE 9, no. 10 (2014): e105001