

UNIVERSITÀ DEGLI STUDI DI MILANO

Procedura di selezione per la chiamata a professore di I fascia da ricoprire ai sensi dell'art. 18, comma 1, della Legge n. 240/2010 per il settore concorsuale 05/E2 - Biologia Molecolare ,  
(settore scientifico-disciplinare BIO/11 - Biologia Molecolare)  
presso il Dipartimento di Bioscienze, Codice concorso 4650

## Giulio Pavesi CURRICULUM VITAE

### INFORMAZIONI PERSONALI (NON INSERIRE INDIRIZZO PRIVATO E TELEFONO FISSO O CELLULARE)

COGNOME	PAVESI
NOME	GIULIO
DATA DI NASCITA	08/10/1970

### TITOLI

#### TITOLO DI STUDIO

(indicare la Laurea conseguita inserendo titolo, Ateneo, data di conseguimento, ecc.)

Aprile 1998: Laurea in Scienze dell'Informazione con pieni voti e lode, Università degli Studi di Milano.  
Tesi: "Grammatiche context-free per la predizione della struttura delle proteine". Relatore Prof. G.Mauri

#### TITOLO DI DOTTORE DI RICERCA O EQUIVALENTI, OVVERO, PER I SETTORI INTERESSATI, DEL DIPLOMA DI SPECIALIZZAZIONE MEDICA O EQUIVALENTE, CONSEGUITO IN ITALIA O ALL'ESTERO

(inserire titolo, ente, data di conseguimento, ecc.)

Gennaio 2003: Dottorato di ricerca (quadriennale) in Informatica. Università degli Studi di Milano.  
Tesi: "String Algorithms in Computational Biology". Relatore Prof. G.Mauri.

#### ALTRI TITOLI CONSEGUITI

(inserire titolo, ente, data di conseguimento, ecc.)

Conseguita Abilitazione Scientifica Nazionale a Professore di Prima Fascia - SSD BIO/18 (05/I1) - Genetica - Giugno 2020.

Conseguita Abilitazione Scientifica Nazionale a Professore di Prima Fascia - SSD BIO/11 (05/E2) - Biologia Molecolare - Aprile 2017.

**NON** Conseguita Abilitazione Scientifica Nazionale a Professore di Prima Fascia - SSD INF/01 (01/B1) - Informatica - Aprile 2017. Giudizio unanime della commissione: pubblicazioni e attività scientifica del candidato non congruenti con le specifiche del settore scientifico disciplinare.

Aprile 2014-oggi: Professore Associato Confermato, SSD INF/01. Dipartimento di Bioscienze, Università degli Studi di Milano.

Aprile 2011-Aprile 2014: Professore Associato non Confermato, SSD INF/01. Dipartimento di Bioscienze, Università degli Studi di Milano.

Gennaio 2006-Aprile 2011: Ricercatore (confermato nel febbraio 2007). Dipartimento di Scienze Biomolecolari e Biotecnologie (successivamente Bioscienze), Università degli Studi di Milano, per modifica di afferenza dal Dipartimento di Informatica e Comunicazione.

Febbraio 2004-Dicembre 2005: Ricercatore non confermato, SSD INF/01. Dipartimento di Informatica e Comunicazione (D.I.Co.), Università degli Studi di Milano.

Novembre 2002-Gennaio 2004: Assegnista di ricerca (assegno di Ateneo), tematica "Algoritmi per l'analisi di sequenze biologiche". Dipartimento di Informatica, Sistemistica e Comunicazione (D.I.S.Co), Università di Milano-Bicocca.

Gennaio 1999-Novembre 2002: Dottorato di ricerca (quadriennale) in Informatica. Università degli Studi di Milano. Tesi dottorato: "String Algorithms in Computational Biology". Relatore Prof. G.Mauri.

Giugno 1998-Dicembre 1998: Borsa di studio nell'ambito del progetto europeo "CAPP" (Cellular Automata in Percolation Processes). Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano. Sviluppo di metodi di simulazione di sistemi complessi basati su automi cellulari.

Aprile 1998: Laurea in Scienze dell'Informazione con pieni voti e lode, Università degli Studi di Milano. Tesi: "Grammatiche context-free per la predizione della struttura delle proteine". Relatore Prof. G.Mauri.

Diplomi in Teoria e Solfeggio e Pianoforte Complementare (Conservatorio di Alessandria - 1986), nell'ambito del corso di diploma in violoncello non portato a termine (abbandonato al termine del quinto anno di corso sui dieci previsti).

## ATTIVITÀ DIDATTICA

### INSEGNAMENTI E MODULI

A partire dalla mia presa di servizio come Professore Associato all'Università degli Studi di Milano nell'anno accademico 2011-2012 ho svolto almeno 120 ore di didattica frontale annuale come previsto dai regolamenti di Ateneo e ministeriali. Precedentemente avevo ricoperto per affidamento didattico corsi dello stesso ambito. La didattica è stata quasi completamente nell'ambito di corsi di laurea triennale e magistrale di area biologica e biotecnologica, solitamente collocati nel SSD BIO/11. Unica eccezione, 3 CFU (24 ore) nell'ambito del corso di Bioinformatica nella laurea magistrale in Informatica (tenuti fino al 2013) e più recentemente 3 CFU (30 ore) di "Algoritmi per la Bioinformatica" tenuti a contratto nella laurea triennale di Ingegneria Informatica al Politecnico di Milano. Il rapporto tra CFU e ore frontali dei diversi corsi è variabile, a seconda del fatto che i singoli CFU siano considerati di lezioni (8 ore) o esercitazione pratica al PC (16 ore).

#### **Anno accademico 2020-2021**

Bioinformatica e Biostatistica (intero corso) - 9 CFU - ore frontali 88 - corso di Laurea Triennale in Biotecnologia - Università di Milano

Genomics and Transcriptomics - modulo "Transcriptomics" - 6 CFU - ore frontali 48 - corso di Laurea Magistrale in Bioinformatics for Computational Genomics - Università di Milano

Advanced Genomics and Epigenomics - modulo "Epigenomics" - 4 CFU - ore frontali 32 - corso di Laurea Magistrale in Bioinformatics for Computational Genomics - Università di Milano

#### **Anno accademico 2019-2020**

Bioinformatica e Biostatistica (intero corso) 9 CFU - ore frontali 88 - corso di Laurea Triennale in Biotecnologia - Università di Milano

Genomics and Transcriptomics - 6 CFU - ore frontali 48 - corso di Laurea Magistrale in Bioinformatics for Computational Genomics - Università di Milano

#### **Anni accademici 2017-2018/2018-2019**

Bioinformatica e Biostatistica (intero corso) 9 CFU - ore frontali 88 - corso di Laurea Triennale in Biotecnologia - Università di Milano

Advanced Bioinformatics for Biotechnology - ore frontali 48 - corso di Laurea Magistrale in Molecular Biotechnology and Bioinformatics - Università di Milano

#### **Anno accademico 2016-2017**

Bioinformatica e Biostatistica - Modulo Bioinformatica - 6 CFU - ore frontali 64 - corso di Laurea Triennale in Biotecnologia - Università di Milano

Biologia Computazionale - 3 CFU (modulo) - ore frontali 44 - corso di Laurea Triennale in Biotecnologia - Università di Milano

Advanced Bioinformatics for Biotechnology - 6 CFU - ore frontali 56 - corso di Laurea Magistrale in Molecular Biotechnology and Bioinformatics - Università di Milano

#### **Anno accademico 2015-2016**

Metodi Bioinformatici - 6 CFU - 48 ore - corso di Laurea Magistrale in Biotecnologie Molecolari e Bioinformatica - Università di Milano

Laboratorio Interdisciplinare - Bioinformatica - 2 CFU - ore frontali 32 - Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Biologia Computazionale - 6 CFU - ore frontali 72 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

#### **Anni accademici 2013-2014/2014-2015**

Bioinformatica e Biostatistica - Modulo Bioinformatica - 3 CFU - ore frontali 32 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Biologia Computazionale - 6 CFU - ore frontali 72 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Biologia Molecolare e Bioinformatica - Modulo "Bioinformatica" - 3 CFU - ore frontali 24 - Corso di Laurea Triennale in Scienze Biologiche (turno in lingua inglese)

Laboratorio Interdisciplinare - Bioinformatica - 2 CFU - ore frontali 32 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali

#### **Anni accademici 2011-2012/2012-2013**

Bioinformatica e Biostatistica - Modulo Bioinformatica - 3 CFU - ore frontali 32 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Biologia Molecolare e Bioinformatica - Modulo Bioinformatica - 3 CFU - ore frontali 24 - Corso di Laurea Triennale in Scienze Biologiche

Laboratorio Interdisciplinare - Bioinformatica - 2 CFU - ore frontali 32 - Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Bioinformatica - 3 CFU - ore frontali 24 -Laurea magistrale in Informatica, Università di Milano

#### **Affidamenti didattici come Ricercatore Universitario anteriori al 2011**

##### **Anni accademici 2009-2010 e 2010-2011**

Bioinformatica - 3 CFU - ore frontali 24 - Laurea magistrale in Informatica, Università di Milano

Bioinformatica e Biostatistica - Modulo Bioinformatica - 3 CFU - ore frontali 32 - corso di Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Laboratorio Interdisciplinare - Bioinformatica - 2 CFU - ore frontali 32 - Laurea Triennale in Biotecnologie Industriali e Ambientali - Università di Milano

Biologia Molecolare e Bioinformatica - Modulo Bioinformatica - 3 CFU - ore frontali 24 - corso di laurea triennale in Scienze Biologiche - Università di Milano

##### **Anni accademici dal 2005-2006 al 2008-2009**

Laboratorio di Bioinformatica Molecolare - 5 CFU -ore frontali 80 - Laurea triennale in Biotecnologie Industriali ed Ambientali, Università di Milano

Bioinformatica - laurea magistrale in Informatica, Università di Milano - 3 CFU - ore frontali 24

Informatica e Laboratorio di Informatica - 4 CFU - 32 ore - primo anno interfacoltà corso di laurea in Biotecnologie, Università di Milano

##### **Anno accademico 2004-2005**

Laboratorio di Bioinformatica Molecolare - 3 CFU - ore frontali 24 - laurea specialistica in Biologia molecolare della cellula, Università di Milano

Informatica e Laboratorio di Informatica - 4 CFU - 32 ore - primo anno interfacoltà corso di laurea in Biotecnologie, Università di Milano

Corso FSE “Laboratorio Professionalizzante di Bioinformatica”, terzo anno corso di laurea in Scienze Biologiche e Biotecnologie Industriali e Ambientali, 2 moduli di 32 ore ciascuno (3 CFU)

##### **Corsi a contratto**

Negli ultimi 4 anni (anni accademici da 2017-18 a 2020-21) ho tenuto 30 ore di lezione a contratto per anno nell'ambito del corso di “Algoritmi per la Bioinformatica (Bioinformatic Algorithms)”, Laurea Triennale in Ingegneria Informatica, Politecnico di Milano.

## ATTIVITÀ DI DIDATTICA INTEGRATIVA E DI SERVIZIO AGLI STUDENTI

### ATTIVITÀ DI RELATORE DI ELABORATI DI LAUREA, DI TESI DI LAUREA MAGISTRALE, DI TESI DI DOTTORATO E DI TESI DI SPECIALIZZAZIONE

(inserire anno accademico, ateneo, corso laurea, ecc.)

#### **Tesi di laurea**

A partire dalla mia presa di servizio come ricercatore a tempo indeterminato nell'anno accademico 2004-2005 sono stato relatore di tesi di laurea triennale in Biotecnologia (precedentemente Biotecnologie Industriali e Ambientali), in media uno studente per anno accademico. Relatore di tesi di laurea magistrale in bioinformatica, nell'ambito di diverse Lauree Magistrali di area Biologica o Biotecnologica, in media 1 studente per anno accademico.

Elenco di seguito i tesisti magistrali di cui sono stato relatore a partire dalla mia presa di servizio come ricercatore universitario. Sono indicati con (\*) i tesisti per cui il lavoro è stato svolto principalmente in un laboratorio esterno, nell'ambito di mie collaborazioni scientifiche, in cui ho seguito puntualmente il lavoro del tesista e le tecniche bioinformatiche sviluppate o utilizzate. Non sono riportati i tesisti di cui sono stato relatore interno "formale" senza partecipare alla supervisione lavoro quotidiano del tesista stesso. L'elenco è incompleto, in quanto purtroppo l'archivio elettronico delle tesi di laurea di cui sono stato relatore da me predisposto non è più accessibile per problemi tecnici. Alcuni buchi temporali coincidono con i periodi in cui, non avendo incarichi didattici nelle lauree magistrali, mi era difficile trovare studenti interessati ai miei progetti di tesi o viceversa con conoscenze di base adeguate al lavoro da me proposto.

#### **Tesi di Laurea Magistrale**

Dmytro Kladnytskyi (\*) - 2020 - Molecular Biotechnology and Bioinformatics  
Alessio Murgia - 2020 - Molecular Biotechnology and Bioinformatics  
Fabio Bedin (\*) - 2020 - Molecular Biotechnology and Bioinformatics  
Carolina Bezzi (\*) - 2020 - Molecular Biotechnology and Bioinformatics  
Matteo Milani - 2017 - Biotecnologie Molecolari e Bioinformatica  
Enrico Massignani (\*) - 2017 - Biotecnologie Molecolari e Bioinformatica  
Mattia Giomi - 2015 - Biotecnologie Molecolari e Bioinformatica  
Maurizio Pedrazzoli - 2014 - Biotecnologie Molecolari e Bioinformatica  
Francesco Ilardi (\*) - 2014 - Genomica Funzionale e Bioinformatica  
Gian Marco Prazzoli - 2010 - Genomica Funzionale e Bioinformatica  
Claudio Peri (\*) - 2010 - Genomica Funzionale e Bioinformatica  
Francesca Pozzi - 2009 - Biologia Molecolare della Cellula  
Massimiliano Borsani - 2009 - Genomica Funzionale e Bioinformatica  
Federico Zambelli - 2006 - Biologia Molecolare della Cellula

#### **Tesi di Laurea Triennale**

Nell'ambito del corso di Laurea Triennale in Biotecnologia sono stato relatore dei seguenti studenti:

Federico Rossignoli - 2020 - Biotecnologia  
Roberto Rossini - 2018 - Biotecnologia

Negli anni accademici precedenti, i "tesisti" delle lauree triennali in cui sono stato relatore (una decina in tutto) svolgevano per l'elaborato finale un lavoro essenzialmente compilativo, che consisteva nella lettura/riassunto di due/tre articoli, e nella stesura di una tesina riassuntiva valutata dal relatore. Non trattandosi di tesi sperimentali non ne ho tenuto traccia, per cui non è disponibile l'elenco dettagliato.

#### **Tesi di dottorato**

Sono stato relatore delle seguenti tesi di dottorato:

Jacopo Lucci, XXIV Ciclo (Relatore interno, dottorato svolto presso l'EMBL-Heidelberg) - Esame finale aprile 2012. Dottorato in Scienze Genetiche e Biomolecolari. Università degli Studi di Milano  
Massimiliano Borsani, XXV Ciclo (Relatore interno, dottorato svolto presso il Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università di Milano-Bicocca) - Esame finale maggio 2013. Dottorato in Scienze Biologiche e Molecolari. Università degli Studi di Milano.  
Gian Marco Prazzoli, XXVII Ciclo - Esame finale aprile 2015. Dottorato in Scienze Biologiche e Molecolari. Università degli Studi di Milano.  
Elena Ceccacci, XXIX Ciclo - Esame finale settembre 2017. Dottorato in Biologia Molecolare e Cellulare. Università degli Studi di Milano

#### **ATTIVITÀ DI TUTORATO DEGLI STUDENTI DI CORSI DI LAUREA E DI LAUREA MAGISTRALE E DI TUTORATO DI DOTTORANDI DI RICERCA**

*(inserire anno accademico, corso laurea, ecc.)*

2006-oggi: Membro del Collegio dei Docenti della Scuola di Dottorato in Scienze Genetiche e Biomolecolari, successivamente rinominata Scienze Biologiche e Molecolari e attualmente Biologia Molecolare e Cellulare. Università degli Studi di Milano.

Sono stato tutor dei seguenti studenti di dottorato (si veda la sezione precedente):

Jacopo Lucci, XXIV Ciclo (Relatore interno, dottorato svolto presso l'EMBL-Heidelberg) - Esame finale aprile 2012.

Massimiliano Borsani, XXV Ciclo (Relatore interno, dottorato svolto presso il Dipartimento di Informatica, Sistemistica e Comunicazione dell'Università di Milano-Bicocca) - Esame finale maggio 2013.

Gian Marco Prazzoli, XXVII Ciclo - Esame finale aprile 2015.

Elena Ceccacci, XXIX Ciclo - Esame finale settembre 2017.

Sono stato tutor (relatore esterno) dei seguenti studenti di Atenei diversi dall'Università degli Studi di Milano, che hanno svolto nel mio laboratorio il tirocinio di tesi di laurea:

Gaia Sambruni - 2017 - Laurea Magistrale in Biologia - Università di Milano-Bicocca

Andrea Corneo - 2017 - Laurea Magistrale in Ingegneria Informatica - Politecnico di Milano

#### **SEMINARI**

*(inserire titolo del seminario, luogo, data, ecc.)*

A partire dal 2015 (6 anni in tutto) ho tenuto con cadenza annuale un seminario di 4 ore per gli studenti della Scuola di Dottorato in Medicina Molecolare (SEMM) all'Istituto Europeo di Oncologia di Milano.

Titolo anni 2015-2016-2017 (febbraio): "Annotating Regulatory Elements in Genomic Sequences"

Titolo anni 2018-2019-2020 (febbraio): "Introduction to single cell transcriptomics"

Inoltre, sempre nell'ambito dello stesso programma di Dottorato, negli ultimi tre anni ho tenuto un ulteriore seminario, di due ore (sempre nel mese di febbraio): "Introduction to Computational approaches for the analysis of next-generation sequencing data".

A partire dal 2015 e fino al 2019 (novembre) ho tenuto un seminario con cadenza annuale all'Università di Milano-Bicocca su "Computational approaches for the annotation of regulatory elements in genomic sequences" per gli studenti delle lauree magistrali di area biologica e biotecnologica.

Singoli seminari recenti su invito in centri di ricerca/Università italiane ed estere:

“Single cell transcriptomics, as of today” - febbraio 2019 - Istituto Italiano di Tecnologia (IIT) - Genova.

“Quando i geni non bastano: genetica ed epigenetica” - maggio 2013 - Istituto Lombardo Accademia di Scienze e Lettere - Milano.

“In silico prediction of transcription factor binding sites: from Chips to ChIPs” - marzo 2010 - Centre for Genomic Regulation (CRG), Barcellona (Spagna).

“Next-Generation Sequencing and Next-Generation Bioinformatics” - agosto 2009 - Kasetsart University, Bangkok, Thailandia.

“Playing ‘tag’ with next-generation sequencing data” - luglio 2009 - Fondazione Filerete - Milano.

“In silico identification of regulatory elements in genome sequences” - settembre 2008 - Norwegian University of Science and Technology (NTNU), Trondheim (Norvegia).

## ATTIVITÀ DI RICERCA SCIENTIFICA

### PUBBLICAZIONI SCIENTIFICHE

Articoli indicizzati in Scopus (escludendo duplicati ed errata): **102**

Di questi:

articoli indicizzati in PubMed: **74**

articoli indicizzati in DBLP (Computer Science Bibliography Database): **42**  
(alcuni indicizzati da entrambi)

Citazioni Scopus (luglio 2021): **6666**, h-index **30**

Citazioni Google Scholar (luglio 2021): **9923**, h-index **36**

Tutte le mie pubblicazioni a partire dal 2010 sono apparse esclusivamente in riviste indicizzate in PubMed, a testimonianza dello spostamento dei miei interessi di ricerca dall'ambito algoritmico a quello applicativo/biomolecolare. Si elencano di seguito solo le 74 pubblicazioni di Pubmed. Evidenziate in grassetto quelle di cui sono primo autore o “corresponding author”.

1: Bocchi VD, Conforti P, Vezzoli E, Besusso D, Cappadona C, Lischetti T, Galimberti M, Ranzani V, Bonnal RJP, De Simone M, Rossetti G, He X, Kamimoto K, Espuny-Camacho I, Faedo A, Gervasoni F, Vuono R, Morris SA, Chen J, Felsenfeld D, Pavesi G, Barker RA, Pagani M, Cattaneo E. The coding and long noncoding single-cell atlas of the developing human fetal striatum. *Science*. 2021 May 7;372(6542):eabf5759. doi: 10.1126/science.abf5759. PMID: 33958447.

2: Marzorati F, Wang C, Pavesi G, Mizzi L, Morandini P. Cleaning the Medicago Microarray Database to Improve Gene Function Analysis. *Plants (Basel)*. 2021 Jun 18;10(6):1240. doi: 10.3390/plants10061240. PMID: 34207216; PMCID: PMC8234645.

3: Zambelli F, Pavesi G. Using RNentropy to Detect Significant Variation in Gene Expression Across Multiple RNA-Seq or Single-Cell RNA-Seq Samples. *Methods Mol Biol*. 2021;2284:77-96. doi: 10.1007/978-1-0716-1307-8\_6. PMID: 33835439.

4: Pagin M, Pernebrink M, Giubbolini S, Barone C, Sambruni G, Zhu Y, Chiara M, Ottolenghi S, Pavesi G, Wei CL, Cantù C, Nicolis SK. Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network. *Stem Cells*. 2021 Mar 19. doi: 10.1002/stem.3373. Epub ahead of print. PMID: 33739574.

5: Zambelli F, Chiara M, Ferrandi E, Mandreoli P, Tangaro MA, Pavesi G, Pesole G. aScan: A Novel Method for the Study of Allele Specific Expression in Single Individuals. *J Mol Biol*. 2021 May 28;433(11):166829. doi: 10.1016/j.jmb.2021.166829. Epub 2021 Jan 26. PMID: 33508309.

6: Ronzio M, Bernardini A, Pavesi G, Mantovani R, Dolfini D. On the NF-Y regulome as in ENCODE (2019). *PLoS Comput Biol*. 2020 Dec 28;16(12):e1008488. doi: 10.1371/journal.pcbi.1008488. PMID: 33370256; PMCID: PMC7793273.

- 7: Chiara M, D'Erchia AM, Gissi C, Manzari C, Parisi A, Resta N, Zambelli F, Picardi E, Pavesi G, Horner DS, Pesole G. Next generation sequencing of SARS- CoV-2 genomes: challenges, applications and opportunities. *Brief Bioinform.* 2021 Mar 22;22(2):616-630. doi: 10.1093/bib/bbaa297. PMID: 33279989; PMCID: PMC7799330.
- 8: Ravasio R, Ceccacci E, Nicosia L, Hosseini A, Rossi PL, Barozzi I, Fornasari L, Zuffo RD, Valente S, Fioravanti R, Mercurio C, Varasi M, Mattevi A, Mai A, Pavesi G, Bonaldi T, Minucci S. Targeting the scaffolding role of LSD1 (KDM1A) poises acute myeloid leukemia cells for retinoic acid-induced differentiation. *Sci Adv.* 2020 Apr 8;6(15):eaax2746. doi: 10.1126/sciadv.aax2746. PMID: 32284990; PMCID: PMC7141832.
- 9: Ronzio M, Zambelli F, Dolfini D, Mantovani R, Pavesi G. Integrating Peak Colocalization and Motif Enrichment Analysis for the Discovery of Genome-Wide Regulatory Modules and Transcription Factor Recruitment Rules. *Front Genet.* 2020 Feb 21;11:72. doi: 10.3389/fgene.2020.00072. PMID: 32153638; PMCID: PMC7046753.
- 10: Forestan C, Farinati S, Zambelli F, Pavesi G, Rossi V, Varotto S. Epigenetic signatures of stress adaptation and flowering regulation in response to extended drought and recovery in *Zea mays*. *Plant Cell Environ.* 2020 Jan;43(1):55-75. doi: 10.1111/pce.13660. Epub 2019 Nov 12. PMID: 31677283.
- 11: Durano D, Di Felice F, Caldarelli F, Lukacs A, D'Alfonso A, Saliola M, Sciubba F, Miccheli A, Zambelli F, Pavesi G, Bianchi ME, Camilloni G. Histone acetylation landscape in *S. cerevisiae* nhp6ab mutants reflects altered glucose metabolism. *Biochim Biophys Acta Gen Subj.* 2020 Jan;1864(1):129454. doi: 10.1016/j.bbagen.2019.129454. Epub 2019 Oct 30. PMID: 31676292.
- 12: Bertolini JA, Favaro R, Zhu Y, Pagin M, Ngan CY, Wong CH, Tjong H, Vermunt MW, Martynoga B, Barone C, Mariani J, Cardozo MJ, Tabanera N, Zambelli F, Mercurio S, Ottolenghi S, Robson P, Creighton MP, Bovolenta P, Pavesi G, Guillemot F, Nicolis SK, Wei CL. Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell Stem Cell.* 2019 Mar 7;24(3):462-476.e6. doi: 10.1016/j.stem.2019.02.004. PMID: 30849367; PMCID: PMC6506828.
- 13: Massignani E, Cuomo A, Musiani D, Jammula S, Pavesi G, Bonaldi T. hmSEEKER: Identification of hmSILAC Doublets in MaxQuant Output Data. *Proteomics.* 2019 Mar;19(5):e1800300. doi: 10.1002/pmic.201800300. Epub 2019 Feb 18. PMID: 30656827. (co-corresponding author)
- 14: Onelli E, Scali M, Caccianiga M, Stroppa N, Morandini P, Pavesi G, Moscatelli A. Microtubules play a role in trafficking prevacuolar compartments to vacuoles in tobacco pollen tubes. *Open Biol.* 2018 Oct 31;8(10):180078. doi: 10.1098/rsob.180078. PMID: 30381363; PMCID: PMC6223213.
- 15: Cerrato V, Mercurio S, Leto K, Fucà E, Hoxha E, Bottes S, Pagin M, Milanese M, Ngan CY, Concina G, Ottolenghi S, Wei CL, Bonanno G, Pavesi G, Tempia F, Buffo A, Nicolis SK. Sox2 conditional mutation in mouse causes ataxic symptoms, cerebellar vermis hypoplasia, and postnatal defects of Bergmann glia. *Glia.* 2018 Sep;66(9):1929-1946. doi: 10.1002/glia.23448. Epub 2018 May 6. PMID: 29732603.
- 16: Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, Cappelletti E, Gozzo F, Harman RM, Antczak DF, Miller D, Scharfe M, Pavesi G, Raimondi E, Sullivan KF, Giulotto E. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Res.* 2018 Jun;28(6):789-799. doi: 10.1101/gr.231159.117. Epub 2018 Apr 30. PMID: 29712753; PMCID: PMC5991519.
- 17: Zambelli F, Mastropasqua F, Picardi E, D'Erchia AM, Pesole G, Pavesi G. RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res.* 2018 May 4;46(8):e46. doi: 10.1093/nar/gky055. PMID: 29390085; PMCID: PMC5934672.
- 18: Soldi M, Mari T, Nicosia L, Musiani D, Sigismondo G, Cuomo A, Pavesi G, Bonaldi T. Chromatin proteomics reveals novel combinatorial histone modification signatures that mark distinct subpopulations of macrophage enhancers. *Nucleic Acids Res.* 2017 Dec 1;45(21):12195-12213. doi: 10.1093/nar/gkx821. PMID: 28981749; PMCID: PMC5716071.
- 19: Chiara M, Pavesi G. Evaluation of Quality Assessment Protocols for High Throughput Genome Resequencing Data. *Front Genet.* 2017 Jul 7;8:94. doi: 10.3389/fgene.2017.00094. PMID: 28736571; PMCID: PMC5500642.
- 20: Paci P, Colombo T, Fiscon G, Gurtner A, Pavesi G, Farina L. SWIM: a computational tool to unveiling crucial nodes in complex biological networks. *Sci Rep.* 2017 Mar 20;7:44797. doi: 10.1038/srep44797. Erratum in: *Sci Rep.* 2017 Jun 16;7:46843. PMID: 28317894; PMCID: PMC5357943.
- 21: Pavesi G. ChIP-Seq Data Analysis to Define Transcriptional Regulatory Networks. *Adv Biochem Eng Biotechnol.* 2017;160:1-14. doi: 10.1007/10\_2016\_43. PMID: 28070596.
- 22: Galardi S, Savino M, Scagnoli F, Pellegatta S, Pisati F, Zambelli F, Illi B, Annibali D, Beji S, Orecchini E, Alberelli MA, Apicella C, Fontanella RA, Michienzi A, Finocchiaro G, Farace MG, Pavesi G, Ciafrè SA, Nasi S. Resetting cancer

stem cell regulatory nodes upon MYC inhibition. *EMBO Rep.* 2016 Dec;17(12):1872-1889. doi: 10.15252/embr.201541489. Epub 2016 Nov 16. PMID: 27852622; PMCID: PMC5283599.

**23: Zambelli F, Pavesi G. Genome wide features, distribution and correlations of NF-Y binding sites. *Biochim Biophys Acta Gene Regul Mech.* 2017 May;1860(5):581-589. doi: 10.1016/j.bbagr.2016.10.007. Epub 2016 Oct 18. PMID: 27769808.**

**24: Dolfini D, Zambelli F, Pedrazzoli M, Mantovani R, Pavesi G. A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. *Nucleic Acids Res.* 2016 Jun 2;44(10):4684-702. doi: 10.1093/nar/gkw096. Epub 2016 Feb 20. PMID: 26896797; PMCID: PMC4889920.**

**25: Bragato C, Gaudenzi G, Blasevich F, Pavesi G, Maggi L, Giunta M, Cotelli F, Mora M. Zebrafish as a Model to Investigate Dynamin 2-Related Diseases. *Sci Rep.* 2016 Feb 4;6:20466. doi: 10.1038/srep20466. PMID: 26842864; PMCID: PMC4740890.**

**26: Mihailovich M, Bremang M, Spadotto V, Musiani D, Vitale E, Varano G, Zambelli F, Mancuso FM, Cairns DA, Pavesi G, Casola S, Bonaldi T. miR-17-92 fine-tunes MYC expression and function to ensure optimal B cell lymphoma growth. *Nat Commun.* 2015 Nov 10;6:8725. doi: 10.1038/ncomms9725. PMID: 26555894; PMCID: PMC4667639.**

**27: Onorati MC, Arancio W, Cavaliere V, Ingrassia AM, Pavesi G, Corona DF. Trans-Reactivation: A New Epigenetic Phenomenon Underlying Transcriptional Reactivation of Silenced Genes. *PLoS Genet.* 2015 Aug 20;11(8):e1005444. doi: 10.1371/journal.pgen.1005444. PMID: 26292210; PMCID: PMC4546373.**

**28: Zambelli F, Pavesi G. RIP-Seq data analysis to determine RNA-protein associations. *Methods Mol Biol.* 2015;1269:293-303. doi: 10.1007/978-1-4939-2291-8\_18. PMID: 25577386.**

**29: Zambelli F, Pavesi G. De novo secondary structure motif discovery using RNAPProfile. *Methods Mol Biol.* 2015;1269:49-62. doi: 10.1007/978-1-4939-2291-8\_4. PMID: 25577372.**

**30: D'Erchia AM, Atlante A, Gadaleta G, Pavesi G, Chiara M, De Virgilio C, Manzari C, Mastropasqua F, Prazzoli GM, Picardi E, Gissi C, Horner D, Reyes A, Sbisà E, Tullo A, Pesole G. Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion.* 2015 Jan;20:13-21. doi: 10.1016/j.mito.2014.10.005. Epub 2014 Nov 1. PMID: 25446395.**

**31: Zambelli F, Pesole G, Pavesi G. Using Weeder, Pscan, and PscanChIP for the Discovery of Enriched Transcription Factor Binding Site Motifs in Nucleotide Sequences. *Curr Protoc Bioinformatics.* 2014 Sep 8;47:2.11.1-31. doi: 10.1002/0471250953.bi0211s47. PMID: 25199791.**

**32: Bellini E, Pavesi G, Barbiero I, Bergo A, Chandola C, Nawaz MS, Rusconi L, Stefanelli G, Strollo M, Valente MM, Kilstrup-Nielsen C, Landsberger N. MeCP2 post-translational modifications: a mechanism to control its involvement in synaptic plasticity and homeostasis? *Front Cell Neurosci.* 2014 Aug 13;8:236. doi: 10.3389/fncel.2014.00236. PMID: 25165434; PMCID: PMC4131190.**

**33: Gregis V, Andrés F, Sessa A, Guerra RF, Simonini S, Mateos JL, Torti S, Zambelli F, Prazzoli GM, Bjerkan KN, Grini PE, Pavesi G, Colombo L, Coupland G, Kater MM. Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. *Genome Biol.* 2013 Jun 11;14(6):R56. doi: 10.1186/gb-2013-14-6-r56. PMID: 23759218; PMCID: PMC3706845.**

**34: Zambelli F, Pesole G, Pavesi G. PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments. *Nucleic Acids Res.* 2013 Jul;41(Web Server issue):W535-43. doi: 10.1093/nar/gkt448. Epub 2013 Jun 7. PMID: 23748563; PMCID: PMC3692095.**

**35: Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Res.* 2013 Aug;23(8):1195-209. doi: 10.1101/gr.148080.112. Epub 2013 Apr 17. PMID: 23595228; PMCID: PMC3730095.**

**36: Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignanò T, D'Erchia AM, Picardi E, Zambelli F, Principato G, Pavesi G, Pesole G. SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D125-31. doi: 10.1093/nar/gks997. Epub 2012 Oct 30. PMID: 23118479; PMCID: PMC3531144.**

**37: Baroni S, Milani M, Pandini V, Pavesi G, Horner D, Aliverti A. Is renalase a novel player in catecholaminergic signaling? The mystery of the catalytic activity of an intriguing new flavoenzyme. *Curr Pharm Des.* 2013;19(14):2540-51. doi: 10.2174/1381612811319140005. PMID: 23116393.**

**38: Dolfini D, Minuzzo M, Pavesi G, Mantovani R. The short isoform of NF-YA belongs to the embryonic stem cell transcription factor circuitry. *Stem Cells.* 2012 Nov;30(11):2450-9. doi: 10.1002/stem.1232. PMID: 22969033.**

- 39: Zambelli F, Prazzoli GM, Pesole G, Pavesi G. Cscan: finding common regulators of a set of genes by using a collection of genome-wide ChIP-seq datasets. *Nucleic Acids Res.* 2012 Jul;40(Web Server issue):W510-5. doi: 10.1093/nar/gks483. Epub 2012 Jun 4. PMID: 22669907; PMCID: PMC3394253.
- 40: Zambelli F, Pesole G, Pavesi G. Motif discovery and transcription factor binding sites before and after the next-generation sequencing era. *Brief Bioinform.* 2013 Mar;14(2):225-37. doi: 10.1093/bib/bbs016. Epub 2012 Apr 19. PMID: 22517426; PMCID: PMC3603212.
- 41: Chowdhary R, Tan SL, Pavesi G, Jin J, Dong D, Mathur SK, Burkart A, Narang V, Glurich I, Raby BA, Weiss ST, Wong L, Liu JS, Bajic VB. A database of annotated promoters of genes associated with common respiratory and related diseases. *Am J Respir Cell Mol Biol.* 2012 Jul;47(1):112-9. doi: 10.1165/rcmb.2011-0419OC. Epub 2012 Mar 1. PMID: 22383585.
- 42: Martynova E, Pozzi S, Basile V, Dolfini D, Zambelli F, Imbriano C, Pavesi G, Mantovani R. Gain-of-function p53 mutants have widespread genomic locations partially overlapping with p63. *Oncotarget.* 2012 Feb;3(2):132-43. doi: 10.18632/oncotarget.447. PMID: 22361592; PMCID: PMC3326644.
- 43: D'Onorio de Meo P, D'Antonio M, Griggio F, Lupi R, Borsani M, Pavesi G, Castrignanò T, Pesole G, Gissi C. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D1168-72. doi: 10.1093/nar/gkr1144. Epub 2011 Nov 28. PMID: 22123747; PMCID: PMC3245153.
- 44: Mihailovic M, Wurth L, Zambelli F, Abaza I, Militti C, Mancuso FM, Roma G, Pavesi G, Gebauer F. Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR. *RNA.* 2012 Jan;18(1):53-64. doi: 10.1261/rna.029603.111. Epub 2011 Nov 18. Erratum in: *RNA.* 2013 Apr;19(4):586. Mihailovich, Marija [corrected to Mihailovic, Marija]. PMID: 22101243; PMCID: PMC3261744.
- 45: Gatta R, Dolfini D, Zambelli F, Imbriano C, Pavesi G, Mantovani R. An acetylation-mono-ubiquitination switch on lysine 120 of H2B. *Epigenetics.* 2011 May;6(5):630-7. doi: 10.4161/epi.6.5.15623. PMID: 21739721.
- 46: Martelli PL, D'Antonio M, Bonizzoni P, Castrignanò T, D'Erchia AM, D'Onorio De Meo P, Fariselli P, Finelli M, Licciulli F, Mangiulli M, Mignone F, Pavesi G, Picardi E, Rizzi R, Rossi I, Valletti A, Zauli A, Zambelli F, Casadio R, Pesole G. ASPicDB: a database of annotated transcript and protein variants generated by alternative splicing. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D80-5. doi: 10.1093/nar/gkq1073. Epub 2010 Nov 4. PMID: 21051348; PMCID: PMC3013677.
- 47: Zambelli F, Pavesi G, Gissi C, Horner DS, Pesole G. Assessment of orthologous splicing isoforms in human and mouse orthologous genes. *BMC Genomics.* 2010 Oct 1;11:534. doi: 10.1186/1471-2164-11-534. PMID: 20920313; PMCID: PMC3091683. (co-first author)
- 48: Dolfini D, Zambelli F, Pavesi G, Mantovani R. A perspective of promoter architecture from the CCAAT box. *Cell Cycle.* 2009 Dec 15;8(24):4127-37. doi: 10.4161/cc.8.24.10240. Epub 2009 Dec 5. PMID: 19946211.
- 49: Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* 2010 Jan;38(Database issue):D75-80. doi: 10.1093/nar/gkp902. Epub 2009 Oct 30. PMID: 19880380; PMCID: PMC2808995.
- 50: Horner DS, Pavesi G, Castrignanò T, De Meo PD, Liuni S, Sammeth M, Picardi E, Pesole G. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief Bioinform.* 2010 Mar;11(2):181-97. doi: 10.1093/bib/bbp046. Epub 2009 Oct 27. PMID: 19864250.
- 51: Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.* 2009 Jul;37(Web Server issue):W247-52. doi: 10.1093/nar/gkp464. Epub 2009 May 31. PMID: 19487240; PMCID: PMC2703934.
- 52: Pozzi S, Zambelli F, Merico D, Pavesi G, Robert A, Maltère P, Gidrol X, Mantovani R, Vigano MA. Transcriptional network of p63 in human keratinocytes. *PLoS One.* 2009;4(3):e5008. doi: 10.1371/journal.pone.0005008. Epub 2009 Mar 25. PMID: 19390658; PMCID: PMC2672041.
- 53: Ré M, Pavesi G. Detecting conserved coding genomic regions through signal processing of nucleotide substitution patterns. *Artif Intell Med.* 2009 Feb-Mar;45(2-3):117-23. doi: 10.1016/j.artmed.2008.07.015. Epub 2008 Sep 21. PMID: 18809300.
- 54: Intra J, Pavesi G, Horner DS. Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family. *BMC Evol Biol.* 2008 Jul 22;8:214. doi: 10.1186/1471-2148-8-214. PMID: 18647384; PMCID: PMC2492878.

- 55: Pasini ME, Intra J, Pavesi G. Expression study of an alpha-L-fucosidase gene in the Drosophilidae family. *Gene*. 2008 Aug 15;420(1):23-33. doi: 10.1016/j.gene.2008.04.021. Epub 2008 May 15. PMID: 18556148.
- 56: Intra J, Cenni F, Pavesi G, Pasini M, Perotti ME. Interspecific analysis of the glycosidases of the sperm plasma membrane in Drosophila. *Mol Reprod Dev*. 2009 Jan;76(1):85-100. doi: 10.1002/mrd.20932. PMID: 18484570.
- 57: Pavesi G, Pesole G. Using Weeder for the discovery of conserved transcription factor binding sites. *Curr Protoc Bioinformatics*. 2006 Oct;Chapter 2:Unit 2.11. doi: 10.1002/0471250953.bi0211s15. PMID: 18428764.
- 58: Pavesi G, Zambelli F, Caggese C, Pesole G. Exalign: a new method for comparative analysis of exon-intron gene structures. *Nucleic Acids Res*. 2008 May;36(8):e47. doi: 10.1093/nar/gkn153. Epub 2008 Apr 8. PMID: 18397951; PMCID: PMC2377436.
- 59: Castrignanò T, D'Antonio M, Anselmo A, Carrabino D, D'Onorio De Meo A, D'Erchia AM, Licciulli F, Mangiulli M, Mignone F, Pavesi G, Picardi E, Riva A, Rizzi R, Bonizzoni P, Pesole G. ASPicDB: a database resource for alternative splicing analysis. *Bioinformatics*. 2008 May 15;24(10):1300-4. doi: 10.1093/bioinformatics/btn113. Epub 2008 Apr 3. PMID: 18388144.
- 60: Pavesi G, Siccardi A, Viale G, Grazioli C, Calciolari T, Tenchini ML, Plevani P. Hedgehogs, humans and high-school science. The benefits of involving high-school students in university research. *EMBO Rep*. 2008 Mar;9(3):208-11. doi: 10.1038/embor.2008.25. PMID: 18311167; PMCID: PMC2267378.
- 61: Ceribelli M, Dolfini D, Merico D, Gatta R, Viganò AM, Pavesi G, Mantovani R. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol*. 2008 Mar;28(6):2047-58. doi: 10.1128/MCB.01861-07. Epub 2008 Jan 22. PMID: 18212061; PMCID: PMC2268390.
- 62: Galbiati M, Simoni L, Pavesi G, Cominelli E, Francia P, Vavasseur A, Nelson T, Bevan M, Tonelli C. Gene trap lines identify Arabidopsis genes expressed in stomatal guard cells. *Plant J*. 2008 Mar;53(5):750-62. doi: 10.1111/j.1365-313X.2007.03371.x. Epub 2007 Nov 23. PMID: 18036199.
- 63: Pavesi G, Pesole G. Finding conserved gene order across multiple genomes. *Methods Mol Biol*. 2007;396:111-9. doi: 10.1007/978-1-59745-515-2\_8. PMID: 18025689.
- 64: Menges M, Pavesi G, Morandini P, Bögre L, Murray JA. Genomic organization and evolutionary conservation of plant D-type cyclins. *Plant Physiol*. 2007 Dec;145(4):1558-76. doi: 10.1104/pp.107.104901. Epub 2007 Oct 19. PMID: 17951462; PMCID: PMC2151690.
- 65: Pavesi G, Zambelli F, Pesole G. WeederH: an algorithm for finding conserved regulatory motifs and regions in homologous sequences. *BMC Bioinformatics*. 2007 Feb 7;8:46. doi: 10.1186/1471-2105-8-46. PMID: 17286865; PMCID: PMC1803799.
- 66: Intra J, Perotti ME, Pavesi G, Horner D. Comparative and phylogenetic analysis of alpha-L-fucosidase genes. *Gene*. 2007 May 1;392(1-2):34-46. doi: 10.1016/j.gene.2006.11.002. Epub 2006 Nov 11. PMID: 17175120.
- 67: Pavesi G, Mereghetti P, Zambelli F, Stefani M, Mauri G, Pesole G. MoD Tools: regulatory motif discovery in nucleotide sequences from co-regulated or homologous genes. *Nucleic Acids Res*. 2006 Jul 1;34(Web Server issue):W566-70. doi: 10.1093/nar/gkl285. PMID: 16845071; PMCID: PMC1538899.
- 68: Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, ... Pavesi G, ..... Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. *Science*. 2005 Sep 2;309(5740):1559-63. doi: 10.1126/science.1112014. Erratum in: *Science*. 2006 Mar 24;311(5768):1713. PMID: 16141072.
- 69: Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Régnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005 Jan;23(1):137-44. doi: 10.1038/nbt1053. PMID: 15637633.
- 70: Pavesi G, Mauri G, Pesole G. In silico representation and discovery of transcription factor binding sites. *Brief Bioinform*. 2004 Sep;5(3):217-36. doi: 10.1093/bib/5.3.217. PMID: 15383209.
- 71: Pavesi G, Mereghetti P, Mauri G, Pesole G. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res*. 2004 Jul 1;32(Web Server issue):W199-203. doi: 10.1093/nar/gkh465. PMID: 15215380; PMCID: PMC441603.

72: Pavese G, Mauri G, Stefani M, Pesole G. RNAProfile: an algorithm for finding conserved secondary structure motifs in unaligned RNA sequences. *Nucleic Acids Res.* 2004 Jun 15;32(10):3258-69. doi: 10.1093/nar/gkh650. PMID: 15199174; PMCID: PMC434454.

73: Pavese G, Mauri G, Iannelli F, Gissi C, Pesole G. GeneSyn: a tool for detecting conserved gene order across genomes. *Bioinformatics.* 2004 Jun 12;20(9):1472-4. doi: 10.1093/bioinformatics/bth102. Epub 2004 Feb 19. PMID: 14976031.

74: Pavese G, Mauri G, Pesole G. An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics.* 2001;17 Suppl 1:S207-14. doi: 10.1093/bioinformatics/17.suppl\_1.s207. PMID: 11473011.

#### Libri di Testo

M Helmer Citterich, F Ferrè, G Pavese, G Pesole, C Romualdi  
Fondamenti di Bioinformatica  
Zanichelli, Bologna 2018

Testo introduttivo alla bioinformatica per corsi di laurea triennale o magistrale di area biologica/biotecnologica.

#### ORGANIZZAZIONE, DIREZIONE E COORDINAMENTO DI CENTRI O GRUPPI DI RICERCA NAZIONALI E INTERNAZIONALI O PARTECIPAZIONE AGLI STESSI

Coordino dal 2005 le attività del gruppo di ricerca in “Bioinformatica, Evoluzione e Genomica Comparata” del Dipartimento di Bioscienze dell’Università degli Studi di Milano.

Le mie attività di ricerca sono sempre state incentrate sulla bioinformatica in generale e sullo sviluppo e applicazione di algoritmi per l’analisi di sequenze genomiche in particolare. Mentre agli inizi (tesi di laurea e dottorato) ho approfondito prevalentemente gli aspetti più algoritmici e computazionali, a partire dalla presa di servizio come Ricercatore presso il Dipartimento di Scienze Biomolecolari e Biotecnologie mi sono focalizzato principalmente sugli aspetti applicativi, curando l’implementazione degli algoritmi sviluppati in pacchetti software o interfacce Web liberamente accessibili ai ricercatori di tutto il mondo (si veda la sezione “Software” più avanti). I metodi sono principalmente dedicati all’analisi dell’espressione genica e della sua regolazione, e la loro ampia diffusione è testimoniata anche dall’elevato numero di citazioni delle pubblicazioni che li descrivono - che costituisce almeno un terzo delle mie citazioni complessive (si vedano ad esempio tra le altre le pubblicazioni allegate numero 1 - citazioni Scopus 257; numero 3 - 371 citazioni; numero 6 - 264 citazioni).

Altra linea parallela è quella che si può considerare ricerca applicata, in collaborazione con gruppi di ricerca in biologia molecolare o genetica del mio Dipartimento o di altre istituzioni italiane o estere. In questo caso io e il mio gruppo forniamo il supporto per analisi bioinformatiche e soprattutto per l’interpretazione dei risultati ottenuti. Le collaborazioni in questo senso si sono intensificate nel corso degli anni in seguito all’introduzione e alla diffusione delle pratiche sperimentali generalmente dette “omiche”, e in particolare del sequenziamento “next-generation”. L’enfasi di queste ricerche è chiaramente sul risultato biologico o genetico, piuttosto che sullo sviluppo di metodi.

Tuttavia, l’apporto fornito per l’analisi dei dati non è limitato alla semplice e diretta applicazione di programmi o pipeline già disponibili, ma richiede solitamente lo sviluppo ad hoc di opportuni approcci algoritmici e statistici, poiché la relativa novità delle tecnologie utilizzate per la produzione dei dati e il loro costante aggiornamento comporta spesso la mancanza di protocolli e strumenti standard per le analisi più avanzate. Si vedano ad esempio le pubblicazioni allegate, in cui oltre alla pianificazione delle analisi bioinformatiche e all’interpretazione dei risultati, sono stati risolti problemi specifici con l’introduzione di nuovi metodi di analisi: numero 10 (identificazione di diffusa poliadenilazione alternativa in trascritti di cellule tumorali che porta a un generale tendenza all’accorciamento delle 3’UTR per sfuggire alla degradazione tramite miRNA), numero 14 (analisi integrata di dati di ChIA-PET, CHIP-Seq e RNA-Seq per l’identificazione a livello genomico, epigenomico e trascrizionale degli effetti della rimozione del fattore di trascrizione Sox2 in cellule staminali neurali), numero 15 (i lncRNA identificati nello studio nel cervello fetale dimostrano un livello di conservazione in primati e mammiferi significativamente inferiore rispetto ai lncRNA delle cellule adulte).

La possibilità di lavorare su dati di “prima mano” prodotti con nuove tecnologie è d’altra parte di fondamentale importanza nella bioinformatica moderna, in quanto l’interazione con gruppi in grado di fornire validazioni e verifiche sperimentali atte a confermare la validità dei risultati delle analisi in silico permette di stabilire quali possano essere le formalizzazioni più adatte ai diversi casi di studio, nonché le migliori strategie algoritmiche per la risoluzione dei relativi problemi. Questo è anche il motivo per cui parecchi miei tesisti hanno svolto tirocini essenzialmente bioinformatici ma “esterni” a contatto con il gruppo che produceva i dati da analizzare.

Negli ultimi tempi questa linea di ricerca ha assunto una nuova prospettiva, in cui è il mio gruppo di ricerca a scegliere il dominio da studiare e le metodologie da utilizzare, in collaborazione con gruppi “wet” che producono i dati e le validazioni sperimentali dei risultati ottenuti. Si veda ad esempio la pubblicazione allegata numero 11, in cui sono stati identificati, partendo dai dati di ChIP-Seq del progetto ENCODE, numerosi fattori di trascrizione che compongono moduli regolatori “genome wide” in associazione con il fattore NF- $\kappa$ B, successivamente validati tramite ChIP e western blot.

Il progressivo spostamento della mia attività di ricerca bioinformatica dagli aspetti più “informatici” a quelli più “biologici” è sostanzialmente alla base dell’esito negativo della valutazione ASN a professore di prima fascia di informatica, in cui la mia produzione scientifica degli ultimi 10 anni e oltre è stata giudicata all’unanimità “non congruente” con il settore scientifico disciplinare INF/01.

### Software di Analisi Bioinformatica

I seguenti software, pubblicati e/o aggiornati nel corso degli ultimi anni, sono basati su algoritmi da me sviluppati, e sono liberamente accessibili via web su server ospitati nel Laboratorio di Bioinformatica del Dipartimento di Bioscienze dell’Università di Milano.

Ascan (software): identificazione di trascrizione “allele specifica” o “allele biased” da esperimenti di RNA-Seq. Prima pubblicazione 2021. <https://github.com/Federico77z/aScan/>

RNentropy (software): identificazione di geni differenzialmente espressi da esperimenti di RNA-Seq. Prima pubblicazione 2018. <http://www.beaconlab.it/RNentropy>

PscanChIP (versione web e software): motif finding per l’individuazione di siti di legame per fattori di trascrizione in esperimenti di ChIP-Seq. Prima pubblicazione 2013, ultimo aggiornamento gennaio 2018. [http://www.beaconlab.it/pscan\\_chip\\_dev](http://www.beaconlab.it/pscan_chip_dev)

Cscan (versione web): analisi genome-wide basata su ChIP-Seq per l’individuazione di fattori regolatori comuni a geni co-espressi o co-regolati. Prima pubblicazione 2012, ultimo aggiornamento gennaio 2016. <http://www.beaconlab.it/cscan>

Pscan (versione web e software): motif finding per individuazione di siti di legame per fattori di trascrizione arricchiti in promotori di geni co-regolati. Prima pubblicazione 2009, ultimo aggiornamento gennaio 2018. <http://www.beaconlab.it/pscan>

Weeder (software): de novo motif finding in sequenze nucleotidiche. Prima pubblicazione 2004, ultimo aggiornamento 2016. <http://www.beaconlab.it/modtools>

Exalign (versione web e software): confronto e analisi evolutiva di strutture introni-esoni di geni. Prima pubblicazione 2008. <http://www.beaconlab.it/exalign>

WeederH (software): de novo motif finding in sequenze nucleotidiche omologhe. Prima pubblicazione 2007. <http://www.beaconlab.it/modtools>

EPIGEN@GALAXY (web server): web server per la gestione, la condivisione, e l’analisi on-line con diversi strumenti di dati di esperimenti di “next-generation sequencing”, sviluppato nell’ambito del progetto EPIGEN, e accessibile tramite registrazione ai gruppi di ricerca coinvolti nel progetto. <http://www.beaconlab.it/epigalaxy>

### **Finanziamenti e Progetti di Ricerca**

Insieme agli altri componenti del mio gruppo di ricerca sono attualmente coinvolto in diversi progetti nazionali e internazionali, che garantiscono fondi sufficienti alle necessità di strumentazione e personale. In particolare, sono stato responsabile di Unità di Ricerca nei seguenti progetti:

PRIN 2009 (Ottobre 2011 - Ottobre 2013). “Sviluppo ed applicazione di strumenti bioinformatici per l'identificazione e la caratterizzazione di splicing alternativi e della rispettiva regolazione”. Finanziamento ottenuto 35.567 Euro.

PRIN 2012 (Marzo 2014 - Marzo 2017). “Alterazioni dell'RNA Editing A-to-I nelle leucemie mieloidi acute: nuove strategie diagnostiche e terapeutiche”. Finanziamento ottenuto 107.750 Euro.

Progetto Bandiera CNR Epigenomica “EPIGEN” (Gennaio 2012 - Dicembre 2018), Sottoprogetto 2 “Development and application of Bioinformatic tools for the characterization of chromatin states”. Finanziamento ottenuto 240.000 Euro.

### **Assegni di ricerca**

Sono stato responsabile scientifico del seguenti assegni di ricerca, su fondi di ricerca propri:

MONTICOLO FRANCESCO (01/09/2018 - 31/08/2019)  
DOLFINI DILETTA (01/01/2016 - 31/12/2016)  
PEDRAZZOLI MAURIZIO (01/07/2015 - 30/09/2015)  
CHIARA MATTEO (01/03/2015 - 30/06/2017)  
PRAZZOLI GIAN MARCO (01/11/2014 - 30/06/2015)  
CHIARA MATTEO (01/06/2013 - 31/05/2014)  
PRAZZOLI GIAN MARCO (01/11/2011 - 31/10/2014)  
ZAMBELLI FEDERICO (01/10/2008 - 30/09/2009)  
ZAMBELLI FEDERICO (01/10/2007 - 30/09/2008)

E del seguente assegno (con rinnovo) su fondi di Ateneo:

ZAMBELLI FEDERICO (01/05/2010 - 30/04/2012)  
ZAMBELLI FEDERICO (01/05/2012 - 30/04/2014)

### **ATTIVITÀ QUALI LA DIREZIONE O LA PARTECIPAZIONE A COMITATI EDITORIALI DI RIVISTE SCIENTIFICHE** (per ciascuna voce inserire anno, ruolo, rivista scientifica, ecc.)

--

### **TITOLARITÀ DI BREVETTI**

(per ciascun brevetto, inserire autori, titolo, tipologia, numero brevetto, ecc.)

--

### **PREMI E RICONOSCIMENTI NAZIONALI E INTERNAZIONALI PER ATTIVITÀ DI RICERCA** (inserire premio, data, ente organizzatore, ecc.)

--

## PARTECIPAZIONE IN QUALITÀ DI RELATORE A CONGRESSI E CONVEGNI DI INTERESSE INTERNAZIONALE

*(inserire titolo congresso/convegno, data, ecc.)*

Si riportano le partecipazioni più recenti su invito degli organizzatori.

Relatore invitato al XV Congresso della Società Italiana di Genetica Umana (SIGU). Sorrento, novembre 2012. Titolo: "Next Generation Sequencing in Epigenetics: Applications and Perspectives".

Relatore invitato a "Workshop BIO-Info-Bio". Pisa, settembre 2015. Titolo "On the entropy of the human transcriptome"

Keynote speaker. "Opportunities and challenges in single cell transcriptomics". 12th International Horse Genome Workshop. Pavia, settembre 2018. <https://havpav.com/>

Relatore invitato a Workshop Single Cell Revolution 2.0 - Opportunities and challenges in Single-Cell Biology. Titolo: "Single cell transcriptomics as of today". Torino, giugno 2019.

Relatore invitato a "4th International & 16th Iranian Genetics Congress". Titolo: "Painting the Regulatory Landscape of Neural Stem Cells". Online (in origine previsto a Teheran), ottobre 2020.

## ATTIVITÀ GESTIONALI, ORGANIZZATIVE E DI SERVIZIO

### INCARICHI DI GESTIONE E AD IMPEGNI ASSUNTI IN ORGANI COLLEGIALI E COMMISSIONI, PRESSO RILEVANTI ENTI PUBBLICI E PRIVATI E ORGANIZZAZIONI SCIENTIFICHE E CULTURALI, OVVERO PRESSO L'ATENEO O ALTRI ATENEI

*(inserire incarico/impegno, ente, data, ecc.)*

Da ottobre 2019 (attualmente ancora in carica), sono **Presidente del Collegio Didattico Interateneo** in "Bioinformatics for Computational Genomics" dell'Università di Milano, che coordina l'omonimo corso di Laurea Magistrale Interateneo in collaborazione con il Politecnico di Milano. Sono stato tra i principali promotori del corso di laurea stesso, sia per quanto riguarda l'idea iniziale, sia per quanto riguarda la preparazione della documentazione per l'approvazione del progetto in sede locale e ministeriale. Il Corso di Laurea è stato attivato a partire dall'anno accademico 2019-2020.

Dal 2013 al 2017, **Presidente della Commissione Paritetica Docenti-Studenti** del Dipartimento di Bioscienze, Università di Milano.

Dal 2006 al 2017, **membro della Commissione per l'orientamento della Facoltà di Scienze MM.FF.NN.** (attualmente Scienze e Tecnologie) dell'Università di Milano. Ho partecipato all'organizzazione degli Open Day di Facoltà e di Ateneo, e in modo particolare all'allestimento della presentazione e divulgazione del corso di Laurea Triennale in Biotecnologie e della Laurea Magistrale in Biotecnologie Molecolari e Bioinformatica.

Dal 2006 al 2012 sono stato **membro della Commissione Scientifica del Dipartimento di Scienze Biomolecolari e Biotecnologie.**

Dal 2017 (attualmente in carica) **co-direttore** (con prof. Rosalba Giugno, Università di Verona) del **Laboratorio Nazionale "InfoLife"**. Il Laboratorio, che afferisce al Consorzio Interuniversitario Nazionale per l'Informatica (CINI), riunisce e coordina 35 gruppi di ricerca di area informatica

appartenenti ad Atenei italiani che svolgono attività di ricerca nell'ambito della Bioinformatica (<https://www.conorzio-cini.it/index.php/it/labinfolife-home>)

Dal 2017 (attualmente in carica) **Rappresentante dell'Università di Milano presso la Joint Research Unit "Elixir-IIB"**, nodo italiano di "Elixir Europe" per la costruzione di un'infrastruttura bioinformatica europea (<http://www.elixir-europe.org/>).

Dal 2018 (attualmente in carica) **rappresentante italiano (per Elixir-IIB) alla piattaforma "Tools" di Elixir Europe.**

### **ATTIVITÀ CLINICO ASSISTENZIALI**

*(indicare, data, durata, ruolo, ente presso il quale si è prestata attività assistenziale, ecc.)*

--

Data

8 luglio 2021

Luogo

Milano