## Curriculum vitae Matteo Rè

*Last update Dec 16$^{th}$ 2015*
*e-mail: re@di.unimi.it*

*home page: http://homes.di.unimi.it/re/*

## Personal information:

- **Date of birth:** 16/07/1974

- **Place of birth:** Codogno

- **Military obligations:** Fulfilled

- **Work address:** Dept. of Computer Science - Università degli Studi di Milano - Via Comelico 39/41 - 20135 Milano (MI), Italy, phone 02.503.16224/16410, fax 02503.16373

## Sudies:

- **2007** PhD in Cellular and Molecular Biology (research areas: bioinformatics, comparative genomics)

- **2004** Master degree in Biology

## Professional experience

- **2007 - 2008 (Oct):** Chief bioinformatician for the analysis of genetics variants data (Human common genetics diseases Unit) at the dept of Biotechnologies (DiBIT) of Vita e Salute University (S. Raffaele del Monte Tabor hospital) - Milan, IT.

- **2008 (Nov) - 2011:** Post doc researcher dept. of Computer Science - Università degli Studi di Milano

- **2011 - 2012 (Oct):** Bioinformatician at Fondazione Filarete (Business accelerator in biomaterials, healthcare and advanced technology), Milan, Italy. Genomics and BioInformatics technology platform. Role: development and application of tools for the analysis of Next Generation Sequencing data in plant genetics selection projects.

- **2012 (Nov) - 2015 (Oct):** Assistant Professor (RTD-A), dept. of Computer Science - Università degli Studi di Milano.

## Scientific collaborations

- **Department of Computer Science, Aristotle University of Thessaloniki:** Development of hierarchical and multilabel classification methods for automated prediction of gene functions.

- **Institute for Medical Genetics and Human Genetics, Charitè-Universitatsmedizin, Berlin:** Development and application of machine learning methods for the automated prediction of human genes - pathological phenotypes associations.

- **Centre for Systems and Synthetic Biology, dept. of Computer Science, Royal Holloway University, London:** Development of networks integration based machine learning methods for gene and protein function prediction. Development of diseases prognosis, onset and susceptibility predictors based on patient networks built using complex molecular, genetics and clinical data.

- **Bioinformatics and Systems Biology Research Unit, Institute of Molecular Life Sciences, Zurich University:** Development and application of graph kernel based methods for the analysis of big data in bioinformatics.

## Editorial activity and organization on international conferences/workshops:

Matteo Rè serves as reviewer for several international bioinformatics and machine learning journals (*Bioinformatics*, *BMC Bioinformatics*, *Nucleic Acids Research*, *Artificial Intelligence in Medicine*, *Advances in Bioiformatics*, *Neurocomputing*) and for some international conferences (*European Conference on Artificial Intelligence* (ECAI), *Multiple Classifier Systems* (MCS) e *Supervised and Unsupervised Ensemble Methods and their Applications* (SUEMA)). He has been co-chair of the SUEMA 2010 international workshop (European COnference on Machine Learning, ECML), dedicated to supervised and unsupervised ensemble methods and their applications and is editor of a Springer volume containing the best papers presented at SUEMA 2010[19].
He is author of a review article on ensemble systems[21] and is executive editor of the International Journal of Neural Networks (ISSN: 2249-2763 print version, E-ISSN: 2249-2771 electronic version).
His research activity is documented by 18 papers published in international peer reviewed journals and 16 book chapters.

## Scopus record:

ELSEVIER SCOPUS (`www.scopus.com`, last update 05/12/2015):
Author ID: 35305844500
Indexed documents: 27 (Article: 15, Conference paper: 11, Editorial: 1)
Time frame: 2006 - 2015
N. citations: 94
Indice h: 7

# Research Interests:

My research interests belong to the areas of Machine Learning, Computational Biology and Bioinformatics. In specific, I have worked in the past/I am still working on the following topics:

I. **Bioinformatics**

   A) Analysis development and application of supervised machine learning methods

   A1. Ontology based functional classification of genes and proteins. [30, 28, 22, 23, 9, 29, 4]

   A2. Heterogeneous data integration for gene function prediction. [12, 35, 34, 32, 10, 31, 11, 1]

   A3. Complex bio-molecular data integration for the supervised classification of co-expressed genes. [33, 13]

   B) Analysis, development and application of semi-supervised node ranking and classification methods for the anlaysis of genes and drugs networks.

   B1. Cancer gene modules prediction and ranking in gene networks based on kernelized score functions. [25, 5, 6]

   B2. Development of drug networks integration methods based on bipartite networks projection and application in drugs repositioning problems. [24, 7, 3]

   B3. Development and analysis of Big data analysis methods in bioinformatics. Multi organism gene and protein function prediction in large bio-molecular networks. [2]

   C) Other research topics in Bioinformatics

   C1. Automated identification of protein-coding genomic regions and protein-coding mRNAs. [18, 17, 15, 36, 14, 26, 27]

   C2. DNA microarray data analisys for the study of molecular mechanisms controlling the onset and progression of acute mieloid leukemya. [16, 8]

II. **Machine Learning**

   A) Development and analysis of multiclass, multilabel and multipath ensemble methods[21] for hierarchical classification problems. [30, 28, 30, 9, 23, 22]

   B) Development and analysis of semi-supervised methods for graph nodes ranking and classification. [25, 3, 5, 6, 1]

   C) Development and analysis of ensemble methods for the integration of heterogeneous data. [10, 11, 9, 33, 1]

   D) Development of semi-supervised methods based on graph kernels and kernelized score functions for the analysis of large graphs. [2]

   E) Development of machine learning software libraries

   F) Construction and publication of biological datasets

## Teaching experience:

**A.a. 2015/16:**

- Architettura degli Elaboratori I - Laboratorio, dept. of Computer Science, Università degli Studi di Milano.

**A.a. 2014/15:**

- Bioinformatica, dept. of Computer Science, Università degli Studi di Milano.
- Sistemi operativi II (Laboratorio di sistemi operativi), dept. of Computer Science, Università degli Studi di Milano.

**A.a. 2013/14:**

- Bioinformatica, dept. of Computer Science, Università degli Studi di Milano.
- Sistemi operativi II (Laboratorio di sistemi operativi), dept. of Computer Science, Università degli Studi di Milano.

**A.a. 2012/13:**

- Bioinformatica, dept. of Computer Science, Università degli Studi di Milano.
- Bioinformatica, dept. of Pharmaceutical Sciences, Università degli Studi di Milano.

**A.a. 2011/12:**

- Bioinformatica, dept. of Pharmaceutical Sciences, Università degli Studi di Milano.
- Biologia Computazionale, dept. of Biosciences, Università degli Studi di Milano.
- Bioinformatica (4 lessons), dept. of Computer Science, Università degli Studi di Milano.

**A.a. 2010/11:**

- Biologia Computazionale, dept. of Biosciences, Università degli Studi di Milano.

**A.a. 2009/10:**

- Bioinformatica (8 lessons), dept. of Computer Science, Università degli Studi di Milano.
- "Metodi Bioinformatici" dept. of Biosciences, Università degli Studi di Milano.
- "Informatica Avanzata", dept. of Biosciences, Università degli Studi di Milano.

## Talks in international conferences

- Identification of promoter regions in genomic sequences by 1-dimensional constraint clustering, WIRN 2011, Vietri sul mare, Salerno, Italy
- Functional Inference in FunCat through the Combination of Hierarchical Ensembles with Data Fusion Methods. Second International Workshop on Learning from Multi-Label Data, in conjunction with ICML/COLT 2010. June 25, 2010 - Haifa, Israel
- International Symposium on Integrative Bioinformatics, 6th annual meeting: Noise tolerance of multiple classifier systems in data integration-based gene function prediction, 22nd to 24th March 2010, Cambridge, UK.

- Predicting Gene Expression from Heterogeneous Data, Sixth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB), 15-17 October 2009 - Genova (Italy)

- MLD09, ECML workshop on Learning from Multi-Label Data, Sept. 7, 2009: Weighted True Path Rule: a multilabel hierarchical algorithm for gene function prediction. ECML PKDD 2009, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Bled, Slovenia.

- MLSB09, the Third International Workshop on Machine Learning in Systems Biology, Ljubljana, Slovenia,Sept. 5-6 2009. Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction.

- Comparing early and late data fusion methods for gene function prediction. WIRN 2009, May 28-30, 2009, Vietri sul Mare, Salerno, Italy

- Signal Processing in Comparative Genomics. International Workshop on Fuzzy Logic and Applications, WILF 2007. Portofino Vetta - Ruta di Camogli, Genova (Italy) - July 7-10, 2007

## Talks in domestic conferences

- Analysis of large bio-molecular networks through semi-supervised graph-based learning methods ,PRIN workshop 2015. May 29, 2015, Naples, Italy

- Data fusion based gene function prediction using ensemble methods. Sixth Annual Meeting of the Bioinformatics Italian Society. March 18-20, 2009, Genoa, Italy.

## Publications:

## International journals with peer-review

[1] G. Valentini, A. Paccanaro, H. Caniza, A.E. Romero, and **M. Re**. An estensive analysis of disease-gene associations using network integration and fast kernel-based gene prioritization methods. *Artificial Intelligence in Medicine*, 61:63–78, 2014. ISSN: 0933-3657. Impact factor: 2.019.

[2] M. Mesiti, **M. Re**, and G. Valentini. Think globally and solve locally: secondary memory-based network learning for automated multi-species functon prediction. *Gigascience*, 3:1–14, 2014. ISSN: 2047-217X.

[3] **M. Re** and G. Valentini. Network-based drug ranking and repositioning with respect to drugbank therapeutic categories. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10:1359–1371, 2013. ISSN: 1545-5963. Impact factor: 1.438.

[4] M. Frasca, A. Bertoni, **M. Re**, and G. Valentini. A neural network algorithm for semi-supervised node label learning from unbalanced data. *Neural Networks*, 43:84–98, 2013. ISSN: 0893-6080. Impact factor: 2.708.

[5] **M. Re** and G. Valentini. Cancer module genes ranking using kernelized score functions. *BMC Bioinformatics*, 13:S3–S3, 2012. ISSN: 1471-2105. Impact factor: 2.580.

[6] **M. Re**, M. Mesiti, and G. Valentini. A fast ranking algorithm for predicting gene functions in biomolecular networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9:1812–1818, 2012. ISSN: 1545-5963. Impact factor: 1.438.

[7] **M. Re**, M. Mesiti, and G. Valentini. Drug repositioning through pharmacological spaces integration based on networks projection. *EMBnet.journal*, 18:30–31, 2012. ISSN:2226-6089.

[8] A. Beghini, F. Corlazzoli, L. Del Giacco, **M. Re**, F. Lazzaroni, M. Brioschi, G. Valentini, F. Ferrazzi, A. Ghilardi, , M. Righi, M. Turrini, M. Mignardi, C. Cesana, V. Bronte, M. Nilsson, E. Morra, and R. Cairoli. Regeneration-associated wnt signaling is activated in long-term reconstituting ac133bright acute myeloid leukemia cells. *Neoplasia*, 14:1236–1248, 2012. ISSN: 1522-8002. Impact factor: 4.252.

[9] N. Cesa-Bianchi, **M. Re**, and G. Valentini. Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference. *Machine Learning*, pages 1–33, December 2011. http://dx.doi.org/10.1007/s10994-011-5271-6 ISSN: 0885-6125. Impact factor: 1.889.

[10] **M. Ré** and G. Valentini. Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction. *Journal of Machine Learning Research - Machine Learning in Systems Biology*, 8:98–111, 2010. ISSN: 1938-7228.

[11] **M. Ré** and G. Valentini. Noise tolerance of multiple classifier systems in data integration-based gene function prediction. *Journal of Integrative Bioinformatics*, 7(3), 2010. ISSN: 1613-4516.

[12] **M. Re** and G. Valentini. Integration of heterogeneous data sources for gene function prediction using decision templates and ensembles of learning machines. *Neurocomputing*, 73(7-9):1533–1537, 2010. ISSN: 0925-2312. Impact factor: 2.083.

[13] **M. Ré**. Comparing early and late data fusion methods for gene expression prediction. *Soft Comput.*, 15(8):1497–1504, March 2010. ISSN: 1432-7643. Impact factor: 1.271.

[14] **M. Re**, G. Pesole, and D.S. Horner. Accurate discrimination of conserved coding and non-coding regions through multiple indicators of evolutionary dynamics. *BMC Bioinformatics*, 10:282, 2009. ISSN: 1471-2105. Impact factor: 2.580.

[15] **M. Ré** and G. Pavesi. Detecting conserved coding genomic regions through signal processing of nucleotide substitution patterns. *Artificial Intelligence in Medicine*, 45(2-3):117–123, 2009. ISSN: 0933-3657. Impact factor: 2.019.

[16] R. Avogadri, M. Brioschi, F. Ferrazzi, **M. Re**, A. Beghini, and G. Valentini. A stability-based algorithm to validate hierarchical clusters of genes. *IJKESDP*, 1(4):318–330, 2009.

[17] P. D'Onorio De Meo, D. Carrabino, N. Sanna, T. Castrignano, G. Grillo, F. Licciulli, S. Liuni, **M. Re**, F. Mignone, and G. Pesole. A high performance grid-web service framework for the identification of 'conserved sequence tags'. *Future Generation Comp. Syst.*, 23(3):371–381, 2007. ISSN: 0167-739X. Impact factor: 2.786.

[18] **M. Ré**, F. Mignone, M. Iacono, G. Grillo, S. Liuni, and G. Pesole. A new strategy to identify novel genes and gene isoforms: Analysis of human chromosomes 15, 21 and 22. *Gene*, (365):35–40, 2006. ISSN: 0378-1119. Impact factor: 2.138.

## Books editing

[19] O. Okun, G. Valentini, and **M. Re**, editors. *Ensembles in Machine Learning Applications*, volume 373 of *Studies in Computational Intelligence*. Springer-Verlag Berlin Heidelberg, 2011. ISBN: 978-3-642-22909-1.

[20] O. Okun, **M. Re**, and G. Valentini, editors. *Ensembles in Machine Learning Applications*. Proceedings of the the Third Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications (SUEMA), European Conference on Machine Learning, Barcelona, Spain. 2010. Available as `http://suema10.dsi.unimi.it/suemafiles/SUEMA10_proceedings.pdf`.

## Review papers

[21] **M. Re** and G. Valentini. Ensemble methods: A review. In M.J Way, J.D. Scargle, K.M. Ali, and A.N. Srivastava, editors, *Advances in Machine Learning and Data Mining for Astronomy*, Data Mining and Knowledge Discovery, pages 563–582. Chapman and Hall an imprint of CRC Press (a division of Taylor and Francis), 2012. ISBN: 9781439841730.

## Peer reviewed book chapters

[22] G. Valentini, S. Köhler, **M. Re**, M. Notaro, and P.N. Robinson. Prediction of human gene-phenotype associations by exploiting the hierarchical structure of the human phenotype ontology. In *Bioinformatics and biomedical engineering : third international conference, IWBBIO 2015, Granada, Spain, April 15-17, 2015: proceedings.*, volume 9043 of *Lecture Notes in Computer Science*, pages 66–77. Springer Berlin / Heidelberg, 2015. ISBN: 978-3-319-16482-3.

[23] P.N. Robinson, M. Frasca, S. Köhler, M. Notaro, **M. Re**, and G. Valentini. A hierarchical ensemble method for dag-structured taxonomies. In *Multiple Classifier Systems : 12th international workshop MCS 2015, Günzburg, Germany, june 29 - july 1, 2015. Proceedings.*, volume 9132 of *Lecture Notes in Computer Science*, pages 15–26. Springer Berlin, 2015. ISBN: 978-3-319-20247-1.

[24] **M. Re** and G. Valentini. Large scale ranking and repositioning of drugs with respect to drugbank therapeutic categories. In Leonidas Bleris, Ion Mandoiu, Russell Schwartz, and Jianxin Wang, editors, *Bioinformatics*

*Research and Applications*, volume 7292 of *Lecture Notes in Computer Science*, pages 225–236. Springer Berlin / Heidelberg, 2012.

[25] **M. Re** and G. Valentini. Genes prioritization with respect to cancer gene modules using functional linkage network data. In R. Bellazzi and P. Romano, editors, *11th International Workshop, NETTAB 2011, Network Tools and Application in Biology, Pavia, Italy, October 12-14, 2011, Proceedings*, pages 124–125, 2011.

[26] A. Rozza, G. Lombardi, **M. Re**, E. Casiraghi, G. Valentini, and P. Campadelli. A novel ensemble technique for protein subcellular location prediction. In O. Okun, G. Valentini, and M. Re, editors, *Ensembles in Machine Learning Applications*, volume 373 of *Studies in Computational Intelligence*, pages 151–167. Springer-Verlag Berlin Heidelberg, 2011.

[27] A. Bertoni, **M. Re**, F. Sacca, and G. Valentini. Identification of promoter regions in genomic sequences by 1-dimensional constraint clustering. In B. Apolloni, S. Bassis, A. Esposito, and C.F. Morabito, editors, *Neural Nets WIRN11 - Proceedings of the 21st Italian Workshop on Neural Nets, Vietri sul Mare, Salerno, Italy, 2011*, Frontiers in Artificial Intelligence and Applications, pages 162–169. IOS Press, 2011.

[28] **M. Re** and G. Valentini. An experimental comparison of hierarchical bayes and true path rule ensembles for protein function prediction. In N. El Gayar, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems, 9th International Workshop, MCS 2010, Cairo, Egypt, April 7-9, 2010. Proceedings*, volume 5997 of *Lecture Notes in Computer Science*, pages 294–303, 2010.

[29] N. Cesa-Bianchi, **M. Re**, and G. Valentini. Functional inference in funcat through the combination of hierarchical ensembles with data fusion methods. In M. Zhang, G. Tsoumakas, and Z. Zhou, editors, *ICML/COLT Workshop on learning from Multi-Label Data MLD'10 Working Notes, Jun 25, Haifa, Israel*, pages 13–20, 2010. Available as `http://cse.seu.edu.cn/conf/mld10/files/MLD'10.pdf`.

[30] G. Valentini and **M. Re**. Weighted true path rule: a multilabel hierarchical algorithm for gene function prediction. In G. Tsoumakas, M. Zhang, and Z. Zhou, editors, *MLD-ECML 2009, 1st International Workshop on learning from Multi-Label Data, Sept 7, Bled, Slovenia*, pages 132–145, 2009. Available as `http://lpis.csd.auth.gr/workshops/mld09/mld09.pdf`.

[31] **M. Re** and G. Valentini. Simple ensemble methods are competitive with state-of-the-art data integration methods for gene function prediction. In S. Dzeroski, P. Geurts, and J. Rousu, editors, *Machine Learning in Systems Biology, Proceedings of the Third international workshop, Sept 5-6, Ljubljana, Slovenia*, pages 95–104, 2009.

[32] **M. Re** and G. Valentini. Prediction of gene function using ensembles of svms and heterogeneous data sources. In O. Okun and G. Valentini, editors, *Applications of Supervised and Unsupervised Ensemble Methods*, volume 245 of *Studies in Computational Intelligence*, pages 79–91. Springer, 2009.

[33] **M. Re** and G. Valentini. Predicting gene expression from heterogeneous data. In *Computational Intelligence Methods for Bioinformatics and Biostatistics - 6th International Meeting Proceedings, CIBB 2009, Genoa, Italy, October 15-17, 2009*, 2009.

[34] **M. Re** and G. Valentini. Ensemble based data fusion for gene function prediction. In J.A. Benediktsson, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems, 8th International Workshop, MCS 2009, Reykjavik, Iceland, June 10-12, 2009. Proceedings*, volume 5519 of *Lecture Notes in Computer Science*, pages 448–457. Springer, 2009.

[35] **M. Re** and G. Valentini. Comparing early and late data fusion methods for gene function prediction. In B. Apolloni, S. Bassis, and F.C. Morabito, editors, *Neural Nets WIRN09 - Proceedings of the 19th Italian*

*Workshop on Neural Nets, Vietri sul Mare, Salerno, Italy, May 28-30 2009*, Frontiers in Artificial Intelligence and Applications, pages 197–207. IOS Press, 2009.

[36] **M. Ré** and G. Pavesi. Signal processing in comparative genomics. In F. Masulli, S. Mitra, and G. Pasi, editors, *Applications of Fuzzy Sets Theory, 7th International Workshop on Fuzzy Logic and Applications, WILF 2007, Camogli, Italy, July 7-10, 2007, Proceedings*, volume 4578 of *Lecture Notes in Computer Science*, pages 544–550. Springer, 2007.

[37] **M. Re** and G. Valentini. Random walking on functional interaction networks to rank gene involved in cancer. In *Artificial Intelligence Applications and Innovations: AIAI 2012 International workshops: AIAB, AIeIA, CISE, COPA, IIVC, ISQL, MHDV and WADTMB, Halkidiki, Greece. September 27-30, 2012 : proceedings, Part II.*, volume 7292 of *Lecture Notes in Computer Science*, pages 225–236. Springer Berlin / Heidelberg, 2012.

## Domestic conferences proceedings

[38] A. Rozza, G. Lombardi, **M. Re**, E. Casiraghi, G. Valentini, and P. Campadelli. A novel ensemble approach for the subcellular localization of proteins. In *BITS 2011, Bioinformatics Italian Society Annual Meeting, Pisa, Italy, 2011. Proceedings*, 2011.

[39] D. Malchiodi, **M. Re**, and G. Valentini. Uso di mathematica per la classificazione di dati di qualit variabile. In *Mathematica Italia User Group Meeting - Atti del Convegno 2010*. Adalta, 2010.

[40] **M. Re** and G. Valentini. Data fusion based gene function prediction using ensemble methods. In *BITS 2009, Bioinformatics Italian Society Annual Meeting, Genova, Italy, 2009. Proceedings*, 2009.

[41] **M. Re**, C. Nasi, G. Pesole, and D.S. Horner. Efficient detection of conserved coding regions through a comparative genomic approach. In *BITS 2007, Bioinformatics Italian Society Annual Meeting, Napoli, Italy, 2007. Proceedings*, 2007.

[42] V. Piccolo, **M. Re**, G. Pesole, and S.D. Horner. Towards an integrated pipeline for the in-silico prediction of plant micrornas and their precursors. In *Nono Congresso annuale FISV (Federazione Italiana Scienze della Vita), Riva del Garda*, 2007.

[43] **M. Re**, S.D. Horner, C. Nasi, and G. Pesole. Improving the capacity of the cstminer algorithm to correctly classify conserved sequences. In *Ottavo Congresso annuale FISV (Federazione Italiana Scienze della Vita), Riva del Garda*, 2006.

[44] F. Mignone, **M. Re**, D.S. Horner, and G. Pesole. A new strategy to identify novel genes and genes isoforms: whole genome comparison of human and mouse. In *BITS 2006, Bioinformatics Italian Society Annual Meeting, Bologna, Italy, 2006. Proceedings*, 2006.

[45] D.S. Horner, **M. Re**, C. Nasi, and G. Pesole. Improving the cstminer algorithm to correctly classify conserved sequences. In *BITS 2006, Bioinformatics Italian Society Annual Meeting, Bologna, Italy, 2006. Proceedings*, 2006.

[46] **M. Re**, M. Iacono, F. Mignone, T. Castrignano, S. Liuni, G. Grillo, F. Licciulli, D.S. Horner, and G. Pesole. Identification of novel genes and genes isoforms in the human genome using cst miner, a novel algorithm for the differentiazion of coding and non-coding conserved sequence tags. In *BITS 2005, Bioinformatics Italian Society Annual Meeting, Milan, Italy, 2005. Proceedings*, 2005.