

BIOGRAPHICAL SKETCH

NAME: Daniele Ramazzotti (PI)

eRA COMMONS USER NAME: N/A

POSITION TITLE: Tenure Track Assistant Professor

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
2012–2016, Università degli Studi di Milano-Bicocca, Milan (Italy).	Ph.D.	02/2016	Computer Science
2009–2012, Università degli Studi di Milano-Bicocca, Milan (Italy).	Master's degree	07/2012	Computer Science
2004–2009, Politecnico di Milano, Milan (Italy).	Bachelor's degree	02/2009	Computer Engineering
1999–2004, Liceo Scientifico Paolo Frisi, Monza (Italy).	High school scientific diploma	07/2004	High school scientific Lyceum

A. Personal Statement

Daniele Ramazzotti earned his Ph.D. in Computer Science from the University of Milano-Bicocca in February 2016. Since then, he has held Research Associate positions at several prestigious research centers and universities, including the Massachusetts Institute of Technology, New York University, and Stanford University, where he spent over four years. Currently, he is an Assistant Professor at the School of Medicine and Surgery at the University of Milano-Bicocca. His research interests include Bioinformatics with a particular emphasis on Cancer Evolution, Statistics, Bayesian Learning, Machine Learning, Algorithmics, and Theories of Causality.

B. Positions, Scientific Appointments and Honors

2022, March - Present. University of Milano-Bicocca, Milano, Italy. Assistant Professor.

2019, July - 2022, February. University of Milano-Bicocca, Milano, Italy. Postdoctoral scholar.

2020, May - 2020, August. Google Summer of Code 2020.

Mentor for the Organization National Resource for Network Biology (NRNB).

2016, May - 2019, June. Stanford University, Stanford, California, USA. Postdoctoral scholar.

2018, May - 2018, August. Google Summer of Code 2018. Mentor for the Organization National Resource for Network Biology (NRNB).

2017, May - 2017, August. Google Summer of Code 2017. Mentor for the Organization National Resource for Network Biology (NRNB).

2016, January - 2016, April. Università degli Studi di Milano-Bicocca, Milan, Italy. Postdoctoral researcher.

2015, June - 2015, August. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. Intern at the Trotman Laboratory. Hosted by Lloyd Trotman.

2014, July - 2014, August. Yahoo! Labs Barcelona, Barcelona, Spain. Intern at the Web Mining Research group. Hosted by Francesco Bonchi.

2014, January - 2014, June. New York University, New York, USA. Visiting Scholar at the Bioinformatics Laboratory, Courant Institute of Mathematical Sciences. Hosted by Bud Mishra.

2012, September - 2013, December. Freelance. Freelance Mobile Application Developer.
2012, May - 2012, June. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Visitor Student at the AnyScale Learning For All (ALFA) group. Hosted by Una-May O'Reilly.
2011, October - 2012, February. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Visitor Student at the AnyScale Learning For All (ALFA) group. Hosted by Una-May O'Reilly.
2008, October - 2008, November. Run Time Solutions S.r.l., Cinisello Balsamo, Italy. Intern as a Software Developer.

C. Contributions to Science

Modeling the evolution of diseases driven by cumulative progressions

Several diseases related to cell proliferation are characterized by the accumulation of somatic DNA changes, with respect to wildtype conditions. Cancer and HIV are two common examples of such diseases, where the mutational load in the cancerous/viral population increases over time. In these cases, selective pressures are often observed along with competition, cooperation, and parasitism among distinct cellular clones.

Starting from my PhD studies, I have been developing mathematical frameworks to model these phenomena, based on a combination of Bayesian inference and Suppes' theory of probabilistic causation, depicted in graphical structures dubbed Suppes-Bayes Causal Networks (SBCNs). SBCNs are generative probabilistic graphical models that recapitulate the potential ordering of accumulation of such DNA changes during the progression of the disease. Such models can be inferred from data by exploiting likelihood-based model-selection strategies and regularization.

I have widely applied such methods to genomics data derived from cancer patients with results showing the value of this approach.

Selected publications:

- 1) Ramazzotti, D., Caravagna, G., Olde Loohuis, L., Graudenzi, A., Korsunsky, I., Mauri, G., ... & Mishra, B. (2015). CAPRI: efficient inference of cancer progression models from cross-sectional data. *Bioinformatics*, 31(18), 3016-3026.
- 2) Caravagna, G., Graudenzi, A., Ramazzotti, D., Sanz-Pamplona, R., De Sano, L., Mauri, G., ... & Mishra, B. (2016). Algorithmic methods to infer the evolutionary trajectories in cancer progression. *Proceedings of the National Academy of Sciences*, 113(28), E4025-E4034.
- 3) Caravagna, G., Giarratano, Y., Ramazzotti, D., Tomlinson, I., Graham, T. A., Sanguinetti, G., & Sottoriva, A. (2018). Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nature methods*, 15(9), 707.
- 4) Ramazzotti, D., Graudenzi, A., De Sano, L., Antoniotti, M., & Caravagna, G. (2019). Learning mutational graphs of individual tumour evolution from single-cell and multi-region sequencing data. *BMC bioinformatics*, 20(1), 210.
- 5) Ramazzotti, D., Angaroni, F., Maspero, D., Ascolani, G., Castiglioni, I., Piazza, R., ... & Graudenzi, A. (2022). Lace: inference of cancer evolution models from longitudinal single-cell sequencing data. *Journal of Computational Science*, 58, 101523.
- 6) Angaroni, F., Chen, K., Damiani, C., Caravagna, G., Graudenzi, A., & Ramazzotti, D. (2022). PMCE: efficient inference of expressive models of cancer evolution with high prognostic power. *Bioinformatics*, 38(3), 754-762.
- 7) Fontana, D., Crespiatico, I., Crippa, V., Malighetti, F., Villa, M., Angaroni, F., ... & Ramazzotti, D. (2023). Evolutionary signatures of human cancers revealed via genomic analysis of over 35,000 patients. *Nature Communications*, 14(1), 5982.

Understanding heterogeneity in heterogeneous populations

The recent development of high-resolution single-cell RNA-seq technologies increased the availability of high throughput gene expression measurements of individual cells. This may allow us to dissect previously unknown

heterogeneity and functional diversity among cell populations. In this line of work recent efforts have demonstrated that de novo cell type discovery of functionally distinct cell sub-populations is possible via unbiased analysis of all transcriptomic information provided by scRNA-seq data.

However, such analysis heavily relies on the accurate assessment of pairwise cell-to-cell similarities, which poses unique challenges such as outlier cell populations, transcript amplification noise, and dropout events (i.e., zero expression measurements due to sampling or stochastic transcriptional activities).

To address these challenges, I have been contributing to a novel framework that learns an appropriate similarity metric from the input data. The learned similarities enable effective dimension reduction, clustering, and visualization tasks.

More recently, I have been interested in extending such approach to the task of defining clinically different subtypes of cancer. In this case, the input data consists of multi-omics data and the task also requires effective data integration of the different sources.

Selected publications:

- 1) Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., & Batzoglou, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nature methods*, 14(4), 414.
- 2) Ramazzotti, D., Lal, A., Wang, B., Batzoglou, S., & Sidow, A. (2018). Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. *Nature communications*, 9(1), 4453.
- 3) Wang, B., Ramazzotti, D., De Sano, L., Zhu, J., Pierson, E., & Batzoglou, S. (2018). SIMLR: A tool for large-scale genomic analyses by multi-kernel learning. *Proteomics*, 18(2), 1700232.
- 4) Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., ... & Marbach, D. (2019). Assessment of network module identification across complex diseases. *Nature methods*, 16(9), 843-852.

Mutational signature discovery in tumor genomes

Cancer is a genetic disease caused by somatic mutations in genes controlling key biological functions such as cellular growth and division. Such mutations may arise both through cell-intrinsic and exogenous processes, generating characteristic mutational patterns over the genome named mutational signatures. The study of mutational signatures has become a standard component of modern genomics studies since it can reveal which (environmental and endogenous) mutagenic processes are active in a tumor and may highlight markers for therapeutic response.

Mutational signatures computational analysis presents many pitfalls. First, the task of determining the number of signatures is very complex and depends on heuristics. Second, several signatures have no clear etiology, casting doubt on them being computational artifacts rather than due to mutagenic processes. Last, approaches for signatures assignment are greatly influenced by the set of signatures used for the analysis.

To overcome some of these limitations, I have developed novel computational frameworks that allows (i) the efficient extraction, (ii) exposure estimation, and (iii) confidence assessment during the computational inference of mutational patterns.

Selected publications:

- 1) Lal, A., Ramazzotti, D., Weng, Z., Liu, K., Ford, J. M., & Sidow, A. (2019). Comprehensive genomic characterization of breast tumors with BRCA1 and BRCA2 mutations. *BMC Medical Genomics*, 12, 1-13.
- 2) Lal, A., Liu, K., Tibshirani, R., Sidow, A., & Ramazzotti, D. (2021). De novo mutational signature discovery in tumor genomes using SparseSignatures. *PLoS computational biology*, 17(6), e1009119.
- 3) Hernández-Verdin, I., Akdemir, K. C., Ramazzotti, D., Caravagna, G., Labreche, K., Mokhtari, K., ... & Alentorn, A. (2022). Pan-cancer landscape of AID-related mutations, composite mutations, and their potential role in the ICI response. *npj Precision Oncology*, 6(1), 89.
- 4) Heide, T., Househam, J., Cresswell, G. D., Spiteri, I., Lynn, C., Mossner, M., ... & Sottoriva, A. (2022). The co-evolution of the genome and epigenome in colorectal cancer. *Nature*, 1-11.

Recent publications (since 2023)

- 1) Tumor evolution metrics predict recurrence beyond 10 years in locally advanced prostate cancer. J. Fernandez-Mateos, G.D. Cresswell, N. Trahearn, K. Webb, C. Sakr, et al. *Nature Cancer*, 2024; doi:10.1038/s43018-024-00787-0.
- 2) Clonal Lineage Tracing with Somatic Delivery of Recordable Barcodes Reveals Migration Histories of Metastatic Prostate Cancer. R.N. Serio, A. Scheben, B. Lu, D.V. Gargiulo, L. Patruno, et al. *Cancer Discovery*, 2024; doi:10.1158/2159-8290.CD-23-1332.
- 3) Control-FREEC viewer: a tool for the visualization and exploration of copy number variation data. V. Crippa, E. Fina, D. Ramazzotti, R. Piazza *BMC Bioinformatics*, 2024; doi:10.1186/s12859-024-05694-w.
- 4) Differential Expression of NOTCH-1 and Its Molecular Targets in Response to Metronomic Followed by Conventional Therapy in a Patient with Advanced Triple-Negative Breast Cancer. A. Ilari, V. Cogliati, N. Sherif, E. Grassilli, D. Ramazzotti, et al. *Biomedicines*, 2024; doi:10.3390/biomedicines12020272.
- 5) Idiopathic erythrocytosis: a germline disease?. E. M. Elli, M. Mauri, D. D'Aliberti, I. Crespiatico, D. Fontana, et al. *Clinical and Experimental Medicine*, 2024; doi:10.1007/s10238-023-01283-y.
- 6) First-hit SETBP1 mutations cause a myeloproliferative disorder with bone marrow fibrosis. I. Crespiatico, M. Zaghi, C. Mastini, D. D'Aliberti, M. Mauri, et al. *Blood*, 2024; doi:10.1182/blood.2023021349.
- 7) Evaluating the performance of large language models in haematopoietic stem cell transplantation decisionmaking†. I. Civettini, A. Zappaterra, B. M. Granelli, G. Rindone, A. Aroldi, et al. *British Journal of Haematology*, 2023; doi:10.1111/bjh.19200.
- 8) Contribution of pks+ *E. coli* mutations to colorectal carcinogenesis†. B. Chen, D. Ramazzotti, T. Heide, I. Spiteri, J. Fernandez-Mateos, et al. *Nature Communications*, 2023; doi:10.1038/s41467-023-43329-5.
- 9) Evolutionary signatures of human cancers revealed via genomic analysis of over 35,000 patients. D. Fontana, I. Crespiatico, V. Crippa, F. Malighetti, M. Villa, et al. *Nature Communications*, 2023; doi:10.1038/s41467-023-41670-3.
- 10) Effects of blocking CD24 and CD47 'don't eat me' signals in combination with rituximab in mantle-cell lymphoma and chronic lymphocytic leukaemia. A. Aroldi, M. Mauri, D. Ramazzotti, M. Villa, F. Malighetti, et al. *Journal of Cellular and Molecular Medicine*, 2023; doi:10.1111/jcmm.17868.
- 11) Characterization of cancer subtypes associated with clinical outcomes by multi-omics integrative clustering†. V. Crippa, F. Malighetti, M. Villa, A. Graudenzi, R. Piazza, et al. *Computers in Biology and Medicine*, 2023; doi:10.1016/j.combiomed.2023.107064.
- 12) LACE 2.0: an interactive R tool for the inference and visualization of longitudinal cancer evolution. G. Ascolani, F. Angaroni, D. Maspero, F. Craighero, N. L. S. Bhavesh, et al. *BMC Bioinformatics*, 2023; doi:10.1186/s12859-023-05221-3.
- 13) DNA Damage Response (DDR) Is Associated With Treatment-free Remission in Chronic Myeloid Leukemia Patients. F. Malighetti, G. Arosio, C. Manfroni, M. Mauri, M. Villa, et al. *HemaSphere*, 2023; doi:10.1097/HS9.0000000000000852.
- 14) Targeting the immune microenvironment in Waldenström Macroglobulinemia via halting the CD40/CD40ligand axis. A. Sacco, V. Desantis, J. Celay, V. Giustini, F. Rigali, et al. *Blood*, 2023; doi:10.1182/blood.2022019240.