

Curriculum Vitae

Date Prepared: March 25, 2015
Name: **Luca Pinello**
Office Address: 450 Brookline Ave, Boston, MA 02215, LC 1062
Home Address: 97 Strathmore Rd #11, Brighton, MA 02135
Work Phone: (617) 582-9781
Work Email: lpinello@jimmy.harvard.edu
Work FAX: (617) 632-2444
Place of Birth: Palermo, Italy

Education

2004	Bachelor Degree <i>summa cum laude</i>	Computer Science (Prof. Vito Di Gesù)	University of Palermo, Palermo, Italy
2007	Master Degree <i>summa cum laude</i>	Computer Science (Prof. Vito Di Gesù)	University of Palermo
2011	Doctor of Philosophy (Ph.D.)	Mathematics and Computer Science (Prof. Vito Di Gesù/ Prof. Domenico Tegolo)	University of Palermo

Postdoctoral Training

04/2011 - present	Research Fellow	Computational Biology (Prof. Guo-Cheng Yuan)	Dana-Farber Cancer Institute and Harvard TH Chan School of Public Health, Boston, MA
----------------------	-----------------	---	---

Professional Societies

2011-
present "International Association for Pattern Recognition" (IAPR) Member

Editorial Activities

Ad hoc Reviewer

- *Genomics*
- *BMC Bioinformatics*
- *Bioinformatics*
- *Briefings in Bioinformatics*

Report of Funded and Unfunded Projects

Funding Information

Past

2011-2015 1R01 - Toward a systemic understanding of targeting mechanisms for epigenetic factors
1R01 HG005085-01A2 (PI: Yuan, Guo-Cheng)
Role: Postdoctoral Fellow
Granting Organization: National Institutes of Health
The major goals of this project are to develop a wavelet analysis-based computational approach to predict genome-wide epigenetic patterns using DNA sequence information, predict tissue-specific epigenetic changes, and validate computational predictions.

Current

2015-2020 **K99 Pathway to Independence Award**: Discovering Disease-Causal Variants By Linking Genetic And Epigenetic Variation
1K99HG008399-01
Role: Principle Investigator
Granting Organization: National Institutes of Health (NHGRI)
The major goal of this project is to develop novel computational strategies that reveal causal genetic variants underlying complex traits and diseases, thereby aiding the development of personalized therapeutic approaches.

Report of Local Teaching and Training

Teaching of Students in Courses

2006	Programming 35-40 undergraduate students 1 st year Computer Science major (semester long course)	University of Palermo, Italy Teaching assistant Prepared and monitored bi-weekly lab sessions, graded homework
2007	Programming 35-40 undergraduate students 1 st year Computer Science major (semester long course)	University of Palermo, Italy Teaching assistant Prepared and monitored bi-weekly lab sessions, graded homework
2008	Operating System I 35-40 undergraduate students 2 nd year Computer Science major (semester long course)	University of Palermo, Italy Teaching assistant Prepared and monitored bi-weekly lab sessions, graded homework

2013	Computational Methods for Epigenetics 10 PhD students, (week-long course)	University of Palermo, Italy Guest Lecturer
2015	Programming (BIO 510) Efficient computation in Python 25-30 undergraduate students (single guest lecture)	Harvard TH Chan School of Public Health, Boston, MA Guest Lecturer on “Efficient computation in Python”

Formally Supervised Trainees and Faculty

09/2015- present	Research supervisor for Huidong Chen, Phd student, for the development of lineage differentiation trajectories from single cell gene expression data.
12/2014- present	Research supervisor for Yihao Li, a visiting graduate student from Zhejiang University Department of Mathematics Hangzhou, Zhejiang, China for the development of a new tree-clustering algorithm for single cell gene expression data based on bifurcation analysis.

Formal Teaching of Peers (e.g., CME and other continuing education courses)

No presentations below were sponsored by outside entities.

2015	Best practice training session for the department of biostatistics and computational biology on: “Efficient computation in Python - Big Data processing and Parallel Computing”	Dana-Farber Cancer Institute / Harvard TH Chan School of Public Health. Three hours lecture
2016	Best practice training session for the department of biostatistics and computational biology on: “Efficient Computation in Python: Reproducible Research & Big Data Processing”	Dana-Farber Cancer Institute / Harvard TH Chan School of Public Health. Three hours lecture

Local Invited Presentations

No presentations below were sponsored by outside entities.

2007	“A software for modeling the decisional behavior of website users”, Center for Technology and Knowledge (CITC), University of Palermo, Palermo, Italy (short talk)
2007	“Parallel Multi Layer Model P-MLM”, GRID Open Days”, University of Palermo and COMETA Consortium, Palermo, Italy (short talk)
2008	“Multi-Layer Method to Study Genome-Scale Positions of Nucleosome”, International Italy Israel Project in Pattern Discovery Algorithms, Italian Ministry of Scientific Research, Firb. University of Palermo (short talk)
2011	“Motif Independent Metric”, Dana-Farber Cancer Institute, Biostatics and Computational Biology Departmental (seminar)
2013	"Analysis of Chromatin State Plasticity Identifies Cell-type Specific Regulators of

- H3K27me3 Patterns", Brain Lunch Seminar Series, Postdoc and Graduate Students Affairs Office, Dana-Farber Cancer Institute (seminar)
- 2014 "Prediction of epigenome variability from DNA sequences identifies cell-type-specific regulation", Dana-Farber Cancer Institute Biostatics and Computational Biology Departmental Retreat (short talk)
- 2014 "A systematic approach to studying chromatin state variations", Harvard TH Chan School of Public Health, Department of Biostatistics, Principles in Quantitative Genomics Training Grant Retreat (short talk)
- 2015 "IPython, Parallel Computing and Big Data", How is done seminar series, Harvard TH Chan School of Public Health/ Dana-Farber Cancer Institute (seminar)

Report of Regional, National and International Invited Teaching and Presentations

Invited Presentations and Courses

Regional

- 2015 "Data analysis from genome editing experiments", Genome Engineering workshop 3.0, BROAD Institute (workshop)

Abstract Oral Presentations

- 2015 "Analysis of chromatin state plasticity identifies cell-type specific regulators", Cell Circuits And Epigenomics Program Seminar Series, BROAD Institute, Boston (seminar)

National

- 2015 "Learning Chromatin States from ChIP-seq data ", ENCODE Research Applications and Users Meeting (invited talk)

Abstract Oral Presentations

- 2015 "Genome-wide characterization of chromatin state plasticity", ENCODE Research Applications and Users Meeting (short talk from selected abstracts)

International

- 2012 "Epigenetic plasticity: Why DNA sequence matters", Lipari International Summer School on Bioinformatics and Computational Biology, Lipari Island (ME), Italy (invited talk)

Abstract Oral Presentations

- 2009 "A Fuzzy One Class Classifier for Multi Layer Model", International Workshop on Fuzzy Logic and Applications, 2009 Palermo, Italy (short talk from selected

abstracts)
2013 "Prediction of Chromatin State Variability", International Society for Stem Cell Research (ISSCR), 11th Annual meeting, Boston (short talk from selected abstracts)

Report of Scholarship

Peer reviewed publications in print or other media

Research investigations

1. Beyaz S, Mana MD, Roper J, Kedrin D, Saadatpour A, Hong SJ, Bauer-Rowe KE, Xifaras ME, Akkad A, Arias E, **Pinello L**, Katz Y, Shinagare S, Abu-Remaileh M, Mihaylova MM, Lamming DW, Dogum R, Guo G, Bell GW, Selig M, Nielsen GP, Gupta N, Ferrone CR, Deshpande V, Yuan GC, Orkin SH, Sabatini DM, Yilmaz ÖH. High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature*. 2016 Mar 3;531(7592):53-8. doi: 10.1038/nature17173. PubMed PMID: 26935695.
2. Guo G*, **Pinello L***, Han X, Lai S, Shen L, Lin TW, Zou K, Yuan GC, Orkin SH. Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. *Cell Rep*. 2016 Feb 2;14(4):956-65. doi: 10.1016/j.celrep.2015.12.089. Epub 2016 Jan 21. PubMed PMID: 26804902; PubMed Central PMCID: PMC4740311.
3. Wu JN*, **Pinello L***, Yissachar E, Wischhusen JW, Yuan GC, Roberts CW. Functionally distinct patterns of nucleosome remodeling at enhancers in glucocorticoid-treated acute lymphoblastic leukemia. *Epigenetics Chromatin*. 2015 Dec 2;8:53. doi: 10.1186/s13072-015-0046-0. eCollection 2015. PubMed PMID:26633995; PubMed Central PMCID: PMC4667523.
4. Canver MC*, Smith EC*, Sher F*, **Pinello L***, Sanjana NE*, Shalem O, Chen DD, Schupp PG, Vinjamur DS, Garcia SP, Luc S, Kurita R, Nakamura Y, Fujiwara Y, Maeda T, Yuan G-C, Zhang F, Orkin SH & Bauer DE. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. 2015 Sep 16. doi: 10.1038/nature15521. [Epub ahead of print]
5. Huang J, Marco E, **Pinello L**, Yuan GC. Predicting chromatin organization using histone marks. *Genome Biol*. 2015 Aug 14;16(1):162.
6. Hu T, Yeh JE, **Pinello L**, Jacob J, Chakravarthy S, Yuan GC, Chopra R, Frank DA. Impact of the N-terminal domain of STAT3 in STAT3-dependent transcriptional activity. *Mol Cell Biol*. 2015 Jul 13. pii: MCB.00060-15.
7. Hatchi E, Skourti-Stathaki K, Vents S, **Pinello L**, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, Kellis M, Hill SJ, Parmigiani G, Proudfoot NJ, Livingston DM. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell*. 2015 Feb 19;57(4):636-47.

8. Xu J, Shao Z, Li D, Xie H, Kim W, Huang J, Taylor JE, **Pinello L**, Glass K, Jaffe JD, Yuan GC, Orkin SH. Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. Mol Cell. 2015 Jan 22;57(2):304-16.
9. Wu G, Cai J, Han Y, Chen J, Huang ZP, Chen C, Cai Y, Huang H, Yang Y, Liu Y, Xu Z, He D, Zhang X, Hu X, **Pinello L**, Zhong D, He F, Yuan GC, Wang DZ, Zeng C. LincRNA-p21 Regulates Neointima Formation, Vascular Smooth Muscle Cell Proliferation, Apoptosis and Atherosclerosis by Enhancing p53 Activity. Circulation. 2014 Aug 25. pii: CIRCULATIONAHA.114.011675.
10. **Pinello L***, Xu J*, Orkin SH, Yuan GC. Analysis of chromatin state plasticity identifies cell-type specific regulators of H3K27me3 patterns, PNAS 2014 Jan 6; 10.1073/pnas.1322570111.
11. Das PP, Shao Z, Beyaz S, Apostolou E, **Pinello L**, De Los Angeles A, O'Brien K, Atsma JM, Fujiwara Y, Nguyen M, Ljuboja D, Guo G, Woo A, Yuan GC, Onder T, Daley G, Hochedlinger K, Kim J, Orkin SH. Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity. Mol Cell. 2014 Jan 9;53(1):32-48.
12. **Pinello L**, Lo Bosco G, Yuan GC. Applications of alignment-free methods in epigenomics. Brief Bioinform. 2014 May;15(3):419-30. doi: 10.1093/bib/bbt078.
13. Pulakanti K *, **Pinello L ***, Stelloh C, Blinka S, Allred J, Milanovich S, Kiblawi S, Peterson J, Wang A, Yuan GC, Rao S. Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. Epigenetics. 2013 Oct 17; 8(12).
14. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, **Pinello L**, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science. 2013 Oct 11; 342(6155):253-7.
15. Walker SR, Nelson EA, Yeh JE, **Pinello L**, Yuan GC, Frank DA. STAT5 outcompetes STAT3 to regulate the expression of the oncogenic transcriptional modulator BCL6. Mol Cell Biol. 2013 Aug; 33(15):2879-90.
16. Giancarlo R, Lo Bosco G, **Pinello L**, Utro F. A methodology to assess the intrinsic discriminative ability of a distance function and its interplay with clustering algorithms for microarray data analysis. BMC Bioinformatics. 2013 Jan 14; (Suppl 1), S6, BioMed Central Ltd.
17. Xu J, Shao Z, Glass K, Bauer DE, **Pinello L**, Van Handel B, Hou S, Stamatoyannopoulos JA, Mikkola HK, Yuan GC, Orkin SH. Combinatorial Assembly of Developmental Stage-Specific Enhancers Controls Gene Expression Programs during Human Erythropoiesis. Developmental Cell, 2012 Oct 16; 23(4):796-811.
18. Sala A*, Toto M*, **Pinello L***, Gabriele A, Di Benedetto V, Ingrassia AM, Lo Bosco G, Di Gesù V, Giancarlo R, Corona DFV. Genome-wide characterization of chromatin binding and nucleosome spacing activity of the nucleosome remodelling ATPase ISW. EMBO Journal, 2011 May 4; 30(9): 1766–1777.

19. **Pinello L**, Lo Bosco G, Hanlon B, Yuan GC. A motif-independent metric for DNA sequence specificity. *BMC Bioinformatics*. 2011 Oct 21; 12(1):408+.
20. Di Gesù V, Lo Bosco G, **Pinello L**, Yuan GC, Corona DFV. A multi-layer method to study genome-scale positions of nucleosomes. *Genomics* 2009; 93(2):140-5.
21. Di Gesù V, Lo Bosco G, **Pinello L**. A one class KNN for Signal identification: a biological case study. *International Journal of Knowledge Engineering and Soft Data Paradigms (IJKESDP)* 2009; 1(4): N.4.

*** These authors contributed equally to this work**

Research publications without named authorship

1. Meyer P, Siwo G, Zeevi D, Sharon E, Norel R; DREAM6 Promoter Prediction Consortium, Segal E, Stolovitzky G. Inferring gene expression from ribosomal promoter sequences, a crowdsourcing approach. *Genome Research*. 2013 Oct 4; 23: 1928-1937. (*member of the investigative team cited in the appendix of the manuscript)

[Non-peer reviewed scientific or medical publications/materials in print or other media](#)

Proceedings of meetings or other non-peer reviewed research publications

1. Giancarlo R, Lo Bosco G, Pinello L, Utro F. The Three Steps of Clustering in the Post-Genomic Era: A Synopsis. In *Computational Intelligence Methods for Bioinformatics and Biostatistics, Lecture Notes in Computer Science Vol. 6685 (2011)*, pp. 13-30, doi:10.1007/978-3-642-21946-7_2
2. Benvegna F, D'Alessandro A, Lo Bosco G, Luzio D, Pinello L, Tegolo D. A New Dissimilarity Measure for Clustering Seismic Signals. *Image Analysis and Processing – ICIAP 2011, Lecture Notes in Computer Science Volume 6979, 2011*, pp. 434-443
3. Giancarlo R, Lo Bosco G, Pinello L. Distance Functions, Clustering Algorithms and Microarray Data Analysis. *Lecture Notes in Computer Science vol. 6073, Learning and Intelligent Optimization LION 4, Venice, Italy, January 2010*.
4. Lo Bosco G, Pinello L. A Fuzzy One Class Classifier for Multi Layer Model. *Lecture Notes in Artificial Intelligence vol. 5571, 8th International workshop on Fuzzy Logic and application WILF2009, Palermo, Italy, June 2009*, pp. 124-131, V. Di Gesù, S.K. Pal, A. Petrosino (Eds).
5. Di Gesù V, Lo Bosco G, Pinello L. Interval Length Analysis in Multi Layer Model. *Lecture Notes in Bioinformatics vol. 5488, 5th International Meeting on Computational Intelligence for Bioinformatics and Biostatistics CIBB 2008, Vietri sul Mare, Salerno, Italy, 3-4 October 2008*, pp.114-122, F. Masulli, R. Tagliaferri, G.M. Verkhivker (Eds.).
6. Di Gesù V, Lo Bosco G, Pinello L. A one class classifier for Signal identification: a biological case study, *Lecture Notes in Artificial Intelligence vol. 5179, 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES-2008, Zagreb, Croatia, September 2008*, pp.747-754, I. Lovrek, R.J. Howlett, J.C. Lakhmi (Eds.).

7. Corona D, Di Gesù V, Lo Bosco G, Pinello L, Collesano M, Yuan GC. A Multi Layer Method to study Genome-Scale Positions of Nucleosomes. In Proc. of the Ettore Majorana International Center for the Advancement of Science, "Modeling and Simulation in Science", Erice, Italy, April 2007, pp.169-177.
8. Corona D, Di Gesù V, Lo Bosco G, Pinello L, Yuan GC. A new Multi-Layers Method to Analyze Gene Expression. Lecture Notes in Artificial Intelligence vol. 4694, 11th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES-2007 in conjunction with XVII Italian Workshop on Neural Networks WIRN-2007, Vietri sul Mare, Italy, September 2007, pp. 862-869, B. Apolloni, R.J. Howlett, L.C. Jain (Eds.)

Reviews, chapters, monographs and editorials

1. Giancarlo R, Lo Bosco G, Pinello L, Utro F. The Three Steps of Clustering in the Post-Genomic Era, Biological Knowledge Discovery Handbook: Preprocessing, Mining and Post-processing of Biological Data (M. Elloumi and A.Y. Zomaya), John Wiley and Son, ISBN: 978-1-118-13273-9.

Professional educational materials or reports, in print or other media

1. CRISPResso (SOFTWARE): CRISPResso is a software pipeline for the analysis of targeted CRISPR-Cas9 deep sequencing data. This algorithm allows for the quantification of both non-homologous end joining (NHEJ) and homologous directed repair (HDR) occurrences.
<https://github.com/lucapinello/CRISPResso>
2. Haystack (SOFTWARE): Haystack is a suite of computational tools to study epigenetic variability, cross-cell-type plasticity of chromatin states and transcription factors (TFs) motifs providing mechanistic insights into chromatin structure, cellular identity and gene regulation.
<https://github.com/lucapinello/Haystack>
3. Motif Independent Metric (SOFTWARE): software that calculates, given a set of DNA sequences, a measure of sequence specificity independent from the presence of DNA motifs.
<https://github.com/lucapinello/mim>
4. Python Bioutilites (SOFTWARE): free and open source Python library to streamline analysis of genomic data. In the current version the library provides:
<https://github.com/lucapinello/bioutilites>
5. Multi-Layer Model (MLM) (SOFTWARE): analysis of nucleosome position data obtained with microarray-based approaches.
<http://www.math.unipa.it/pinello/mlm/>

Thesis

Pinello L, Multi Layer Analysis (2011), <http://arxiv.org/abs/1107.5349>.

This thesis presents a new methodology to analyze one-dimensional signals through a new approach called Multi Layer Analysis, for short MLA. It also provides some new insights on the relationship between one-dimensional signals processed by MLA and tree kernels, test of randomness and signal processing techniques. The MLA approach has a wide range of application to the fields of pattern discovery and matching, computational biology and many other areas of computer science and signal processing. This thesis includes also some applications of this approach to real problems in biology and seismology.

Narrative Report

I am a computational biologist and currently a research fellow in Prof. Guocheng Yuan's lab at the Harvard TH Chan School of Public Health/Dana-Farber Cancer Institute and an Instructor in Medicine at Harvard Medical School. My postdoc training has provided me with an extraordinarily stimulating environment, interactions with key investigators, and has allowed me to perform and publish high quality research.

I have a background in Computer Science and extensive experience in machine learning, data mining and with high-performance computing technologies, such as parallel computing and the new cloud oriented architectures. I have applied these technologies to solve computationally challenging and Big Data problems associated with next generation sequencing data analysis. Over the course of my academic career, my major focus has been to unravel the role of epigenetic mechanisms that influence gene expression, such as nucleosome positioning, DNA methylation and histone modification. My research has emphasized the critical interplay between epigenetic mechanisms and linear DNA sequences. For example, I have shown that epigenetic patterns can be characterized and predicted from DNA sequences. In the past two years I have also worked on the characterization of chromatin state variability and have developed a new method to identify cell-type specific regulators through the variation of chromatin states. This work has made important contributions to the understanding of epigenetic variations and gene expression regulation. In addition, it has also led to the development of a computational pipeline called HAYSTACK.

I have recently been awarded a NIH K99 transition award where I proposed a general framework for investigating genetic and epigenetic variations in order to generate insights and testable hypotheses for the study of causal genetic variants underlying complex traits and human diseases. My long-term goal is to use innovative computational approaches and cutting-edge experimental assays to systematically analyze sources of variation that affect gene regulation in cancer and other complex diseases. This includes epigenetic variation, genetic variation and (single-cell) gene expression variability. My main objective in these efforts will be to further our understanding of disease etiology and provide a foundation for the development of new drugs and more targeted treatments.

I enjoy sharing my knowledge with students and I am passionate about teaching. My previous teaching experience and background will allow me to teach a wide range of courses and mentor students.

I am a highly collaborative researcher with a broad computational skill-set. Maintaining and fostering prosperous collaborations, mentoring postdocs/students and writing joint grants with other labs and investigators will always be an important part of my daily work because communication with molecular biologists, physicians and statisticians is essential for the interdisciplinary field of computational biology.

I have a strong desire to answer questions about the interplay of genetics, epigenetics and gene expression in development and common diseases such as cancer. Importantly, I am confident my experience, determination and skills will allow me to accomplish my research goals.