

C.A.L.R. (Center for accelerating Leukemia/Lymphoma Research) at Humanitas University

Multi-omics Bioinformatician/Data Scientist

Job Description

We are looking for a Bioinformatician/Data scientist with proven background in cancer and genomics integrative analyses with strong expertise of at least 2/3 years in computational biology, bioinformatics and machine learning related areas, to join our research group.

As part of our technological team, we expect the candidate to help us discover the hidden information underlying complex data (clinical information, NGS, RNAseq, single-cell analysis, and imaging), developing innovative multi-modal integrative models able to translate research findings into personalized healthcare strategies.

The common purpose leading the research activities is the progression toward a data-driven precision medicine with a main focus on hematological cancers.

Main Activities

- Processing and analysis of clinical, multi-omics data and harmonization of complex and highly fragmented data;
- Investigate, define and support the implementation of scalable computational models in order to extract relevant features for the hematological care process improvement;
- Collaborate in research and development of innovative techniques for understanding disease-specific patterns from multi-modal and heterogeneous data;
- Contribute to solutions design and establishment of requirements;
- Data visualization and effective results reporting;
- Self and team-management on individual and team-sized studies' deadlines;
- Collaborate with international partners on cross-academic research projects.

Candidate profile:

We are looking for highly motivated candidates who appreciate challenging tasks, objectives and are strongly oriented to both scientific excellence, high performances computing and human relationships.

Mandatory Requirements

- PhD in Bioinformatics, Computational Biology, Biomedical Engineering, Computer Science or equivalent professional experience (more than 2 years in the same field). Having a PhD in one of these areas is preferred.
- Fluent in written and spoken English and Italian

Hard Skills

- Strong working knowledge of main programming languages (R, Python, bash)
- Strong knowledge of the main bioinformatics data types and formats
- Proven experience processing and analyzing genome/transcriptome sequencing data with appropriate publication and track record
- Ability to design and implement algorithms for data integration investigating disease-specific relevant markers, exploring dimensionality reduction methods to identify key hematological features for clinical process improvement
- Working knowledge of Machine Learning techniques (e.g. k-NN, SVM, Random Forests, etc.) and applied statistics
- Working knowledge of containers technologies (Docker and/or Singularity)
- Experience with cloud (GCP, AWS) and/or distributed computing (HPC)
- Experience in pipeline development, of reproducible research (e.g. git) and/or reproducible software development is a plus

We can offer a fixed-term contract. Contract duration, salary, as well as employment level, will be defined based on candidate's profile.

To apply, please send your CV and a cover letter highlighting your relevant experience to Prof Matteo G Della Porta, matteo.della_porta@hunimed.eu

Selected References

- D'Amico S et al. Synthetic Data Generation by Artificial Intelligence to Accelerate Research and Precision Medicine in Hematology. *J Clin Oncol CCI* 2023, in press
- Sauta E et al. Real-world validation of Molecular International Prognostic Scoring System (IPSS-M) for myelodysplastic syndromes. *J Clin Oncol* 2023 Mar 17;JCO2201784. doi: 10.1200/JCO.22.01784.
- Maggioni G et al. A sex-informed approach to improve the personalised decision making process in myelodysplastic syndromes: a multicentre, observational cohort study. *Lancet Haematol.* 2022 Nov 24:S2352-3026

- Bernard, E et al. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. *NEJM Evid* 2022; 1 (7); DOI:<https://doi.org/10.1056/EVIDoa220000>
- Bernard E, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med.* 2020;26:1549-1556.
- Bersanelli M, et al. Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. *J Clin Oncol.* 2021 Apr 10;39(11):1223-123
- Rossi M, Clinical relevance of clonal hematopoiesis in persons aged ≥ 80 years. *Blood.* 2021 Nov 25;138(21):2093-2105.