umberto.perron@fht.org | +39 3516576080 | | | | |
Fondazione Human Technopole - Viale Rita Levi-Montalcini, 1 - Area MIND – Milano, IT

Umberto Perron

Profile

- Postdoctoral Researcher in Bioinformatics with 5+ years of experience.
- Experienced in building, processing, annotating, integrating, analysing large biomedical datasets and knowledge bases using statistical, machine learning methods.
- Development of tools and approaches for flexible, scalable, and reproducible data science and predictive modelling.
- Experience leading multidisciplinary biomedical research projects, interacting with diverse stakeholders; planning and coordinating data processing, and analysis.
- Experience working in and collaborating with large international research organizations, research hospitals, and pharma R&D teams.

Research experience

Aug 2020 - present Fondazione Human Technopole, Milano, IT
Postdoctoral researcher
Team leader: Dr. Francesco Iorio,

- I processed and integrated multiomics data from a colorectal cancer PDX cohort and developed an ensemble ML classifier of Cetuximab sensitivity which outperforms SOTA biomarker signatures.
- I designed and developed a cancer dependency biomedical knowledge graph integrating multiple data sources. This will be served interactively via a web app as well as being used to train graph ML models for drug repurposing.

Oct 2016 - June 2020 European Bioinformatics Institute (EMBL-EBI), Cambridge, UK
Predoc fellow / PhD student
Supervisor: Dr. Nick Goldman, Dr. Iain Moal

- I developed a novel substitution model for amino acid sequence evolution that makes use of both amino acid identity and side chain rotamer configuration information.
- I designed scalable approaches for annotating protein sequences with structural features computed from X-ray crystallography data.

May 2019 - Jul 2019 Medicines Discovery Catapult, Alderley Edge, UK
Visiting PhD student
Supervisor: Prof. John P. Overington

- I investigated amino acid side chain rotamer configuration stability in different structural contexts.
- I prototyped a protein variant effect prediction ML model combining sequence conservation and structural features.

Oct 2014 - Jul 2016 Universitá degli Studi di Torino, Torino, IT
Graduate research intern
Supervisor: Prof. Paolo Provero

- I developed a pipeline to ingest public RNAseq data, compute a multi-species and multi-tissue co-epression network using Python, R, and Bash.
- I used this co-expression network to annotate uncharacterized lncRNAs with ontology terms associated with protein-coding genes.

umberto.perron@fht.org | +39 3516576080 | | | |
Fondazione Human Technopole - Viale Rita Levi-Montalcini, 1 - Area MIND – Milano, IT

Education

2016-2020 *PhD* in Bioinformatics - University of Cambridge, UK / EMBL International PhD Programme

2014-2016 *MSc* in Molecular Biotechnology - Universitá degli Studi di Torino, IT

2011-2014 *BSc* in Biotechnology - Universitá degli Studi di Torino, IT

Skillset

Programming: strong scientific software development, data analysis and visualization in Python, R. Basic C/C++ and Julia.

Domain knowledge: trained in molecular biology, biochemistry and pharmacology.

Bioinformatics: highly experienced in processing, integrating and analysing large multiomics/clinical datasets. Development of Python/R pipelines for protein sequence and structure analysis, NGS and gene expression analysis, variant and transcript annotation, biomarker and target discovery, drug repositioning, phylogenetics, molecular evolution

Data Engineering: ingestion, processing, and retrieval of biomedical and chemical resources. Backend, internal REST API, Neo4j graph database design and development. Basic SQL, MongoDB, Docker, Kubernetes, Unix/Linux, HPC.

Biostatistics / Machine learning: predictive modelling, inference, testing, statistical computing with NumPy and SciPy, machine learning with scikit-learn, Keras/TensorFlow. Basic PyTorch.

Project lead: experienced in coordinating computational researchers, clinicians, software engineers, data curators, facility staff; communicating results and priorities, drafting analytical plans, scientific grant and manuscript writing and editing.

Selected publications

Published

Vinceti, A., R. R. De Lucia, P. Cremaschi, Perron, U., E. Karakoc, L. Mauri, C. Fernandez, K. H. Kluczynski, D. S. Anderson, and F. Iorio (2022). "CRISPRcleanR WebApp: an interactive web application for processing genome-wide pooled CRISPR-Cas9 viability screen". *bioRxiv*.

Vinceti, A., Perron, U., L. Trastulla, and F. Iorio (2022). "Reduced gene templates for supervised analysis of scale-limited CRISPR-Cas9 fitness screens". *bioRxiv*.

Kalkauskas, A., Perron, U., Y. Sun, N. Goldman, G. Baele, S. Guindon, and N. De Maio (2021). "Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk". *PLoS computational biology* 17.1, e1008561.

Najgebauer, H., Perron, U., and F. Iorio (2021). "Redefining false discoveries in cancer data analyses". *Nature Computational Science* 1.1, 22–23.

Vinceti, A., E. Karakoc, C. Pacini, Perron, U., R. R. De Lucia, M. J. Garnett, and F. Iorio (2021). "CoRe: a robustly benchmarked R package for identifying core-fitness genes in genome-wide pooled CRISPR-Cas9 screens". *BMC genomics* 22.1, 1–16.

Perron, U., A. M. Kozlov, A. Stamatakis, N. Goldman, and I. H. Moal (2019). "Modeling Structural Constraints on Protein Evolution via Side-Chain Conformational States". *Mol. Biol. Evol.* 36.9, 2086–2103.

Perron, U., P. Provero, and I. Molineris (2017). "In silico prediction of lncRNA function using tissue specific and evolutionary conserved expression". *BMC Bioinformatics* 18.Suppl 5, 144.

In preparation

Perron, U., A. Chatzipli, E. Grassi, M. Viviani, E. Karakoc, E. Zanella, C. Isella, I. Molineris, H. Klett, P. Penk, J. Schueler, A. Bertotti, L. Trusolino, N. Conte, U. McDermott, and F. Iorio (-). "Integrative ensemble modelling of Cetuximab sensitivity in Colorectal Cancer PDXs".

Perron, U., R. R. De Lucia, and F. Iorio (-). "A cancer dependency knowledge graph".