# VINYL: Variant prIoritizatioN bY survivaL analysis

Pietro Mandreoli[1,2], Marco Antonio Tangaro[2], David S. Horner [1,2], Federico Zambelli [1,2], Graziano Pesole[2,3], Matteo Chiara[1,2]

[1] Department of Biosciences, University of Milan, via Celoria 26, 20133 Milano, Italy Email: matteo.chiara@unimi.it

[2] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR), Via Giovanni Amendola 122/O, 70126 Bari, Italy

[3] Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, Via Orabona 4, 70126 Bari, Italy

**Project Website**: http://beaconlab.it/VINYL
**Source Code**: https://github.com/matteo14c/VINYL/


**Main Text of Abstract**

Applications of modern NGS sequencing technologies in healthcare and clinical practice are driving a revolution in medical science. The capacity to link genetic variants with phenotypic traits and pathological conditions, is of instrumental importance for the development of informed approaches to medical science, such as precision medicine [1]. However, the need to handle, analyze and interpret large collections of "big" genomic data is posing major challenges which at present remain unresolved. A typical NGS assay can identify even millions of genetic variants, all of which need to be carefully annotated and interpreted to identify genetic traits associated with a pathological condition. This process, which is known as "variant prioritization" typically requires manual curation by an expert clinician, and represents a major bottleneck for the application of large scale genotyping assays in clinical settings [2]. In this contribution, we present VINYL, a novel fully automated Galaxy-based workflow for the prioritization of genetic variants in clinical studies. VINYL incorporates a collection of functional annotations of human genetic variants to derive a pathogenicity score. By comparing a population of affected individuals and a matched control population, a procedure based on survival analysis is applied to optimize the scoring system and to delineate the ideal threshold for the prioritization of pathogenic variants. Extensive simulations based on publicly available human genetic data demonstrate that VINYL can detect different types of genetic variants associated with pathological conditions. Moreover re-analysis of a publicly available dataset of cardiomyopathies patients show that VINYL achieves higher levels of sensitivity and specificity than equivalent state of the art tools [3]. VINYL is available at beaconlab.it/VINYL as a dedicated Galaxy instance- made available through the Laniakea@ReCaS Galaxy [4] on demand service- which incorporates a highly curated collection of tools and resources for the functional annotation of genetic variants.

**References**

[1]https://doi.org/10.3109/10408363.2014.997930
[2] https://www.doi.org/10.1038/nrg.2017.52
[3] https://doi.org/10.1101/2020.01.23.917229/
[4] https://doi.org/10.1093/gigascience/giaa033