

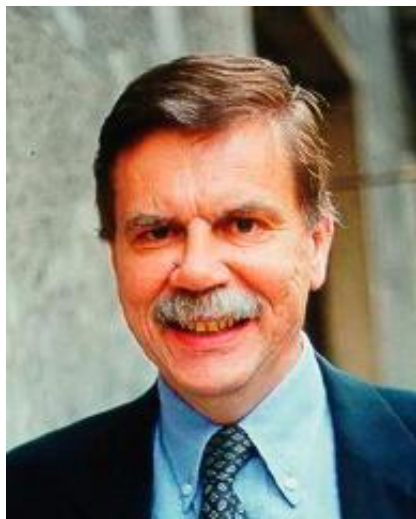
Bioinformatics Seminar

Mon 19 Nov 2018 | 12 – 1 PM

Duke-NUS Medical School

Meeting Room 7C Level 7

Host | Steve Rozen, Professor
Programme in Cancer & Stem Cell Biology
Director, Centre for Computational Biology
Duke-NUS Medical School



Stefano Ceri

Professor of Database Systems
Dipartimento di Elettronica, Informazione e
Bioingegneria (DEIB) of Politecnico di
Milano, Italy

The GeCo Project: extraction, analysis and integration of NGS genomic datasets

Geco is an Advanced ERC Project for data-driven genomic computing. In this talk, I will describe the resources developed during the first two years of the project:

- GMQL, a language and system for querying processed genomic datasets on the cloud
- PyGMQL, a Python library for data analysis embedding GMQL, and its wrapping as Jupyter notebook
- The GMQL Repository, which integrates processed datasets from IGC (TCGA), Encode, RoadMap Epigenomics and many useful annotations, all represented in the same data model and with a common core of semantically aligned metadata.

I will then show GMQL applications to cancer genomics which use the GMQL repository. GMQL uses the Spark big data engine and can be accessed at CINECA in Italy and at the Broad Institute in Cambridge (US).

All are welcome

Questions? Contact Barbara Levy
barbara.levy@duke-nus.edu.sg