

# **Ph.D. in Information Technology: Thesis Defense**

**March 8th, 2019**

**Villa Del Grumello - Como – 5.00 pm**

**Stefano PERNA – XXXI Cycle**

“Data-Driven Methods for Knowledge Discovery in Regulomics”

Advisors: **Prof. Stefano Ceri, Prof. Limsoon Wong**

## **Abstract:**

In this thesis, novel computational methods for transcription factor-transcription factor interactions are developed, tested and validated. All methods are strongly data-driven and aim at providing interactions hypotheses for wet-lab verification and classifying existing or supposed TF-TF interactions in cooperations and competitions, without the need of a priori biological expertise. These methods have wide applicability for medical and academic purposes, as the algorithms of inferring and classifying protein-protein interactions are still in their infancy. The proposed framework consists of three methods: a foundational TF-TF interaction detector, TICA, that leverages on both the Genometric Query Language (GMQL) and sound statistical inference to detect whether two transcription factors interact based on positional information of their binding sites; and two complementary follow-ups, NAUTICA and ESTETICA, that tackle the classification problem based on Protein-protein Interaction Networks (PPI) analysis for the former and the use of pattern recognition on the signal enrichment feature. These methods have been tested and validated both against known protein complex and PPI databases (respectively CORUM and BioGRID) and by manually investigating the known literature. All methods have shown very good performance and highlighted novel, previously unknown TF-TF interactions. The overall framework is shown to be cohesive and functional, and the model to be sound as well.

## **PhD Committee:**

Prof. **Stefano Ceri**, DEIB

Prof. **Raffaele Calogero**, Università di Torino

Prof. **Colin Logie**, Radboud University, Nijmegen, NL