

# Terzo Meeting PRIN 2017-March 21, 2021

## Unipa

---

- **Persone**

- **Strutturati**

- R. Giancarlo, G. Lo Bosco, S.E. Rombo

- **Dottorandi**

- D. Amato, M. Bonomo, **G. Grimaudo, A. La Placa**



# Tasks T1 e T3

---

- Dr. Domenico Amato- a breve
- A Big Data Approach to Sequence Indexing on the Cloud Via Burrows and Wheeler Transform
  - Primo algoritmo che distribuisce il calcolo della trasformata su un cluster Hadoop, piuttosto che solo i dati
  - Vantaggi: uso pieno delle risorse di calcolo, rispetto a metodi già ben ricevuti in letteratura
  - Applicazioni: Precision Medicine
  - M. Randazzo e Simona E. Rombo, Proc. AAI4H 2020



# Task T2

---

- Corenup
  - Modello compatto basato su Convolutional e Recurrent Neural Nets per la l'Individuazione di Nucleosomi in Genomica
    - Modelli piccoli ed accurati in questo dominio sono rari
  - Amato, D., Lo Bosco, G., & Riccardo, R. (2020). CORENup: a combination of convolutional and recurrent deep neural networks for nucleosome positioning identification. BMC BIOINFORMATICS, 21(8), 326.



# Task T2

---

- Prediction of lncRNA-disease associations from tripartite graphs
  - Un grafo compatto che modella associazioni tra malattie, lncRNAs e miRNA
  - Metodi di Mining Efficienti
    - Migliora in maniera sostanziale modelli esistenti
- Mariella Bonomo, Armando La Placa, Simona E Rombo, DMAH, 2020.



# Task T2

---

- Dizionari di K-mers in piccolo spazio per classificazione Alignment Free
  - FADE: Piattaforma Spark che utilizzando rappresentazioni compatte di k-mers permette il primo benchmarking su grandissima scala di misure di similarità alignment-free in genomica
  - U. Ferraro Petrillo, F. Palini, G. Cattaneo, R. Giancarlo, Alignment-free Genomic Analysis via a Big Data Spark Platform Bioinformatics, 2021.



# Task T4

---

- Multicriteria Compression per Hadoop
  - Primo metodo per selezionare compressori genomici ed importarli in Hadoop, scrivendo poche linee di codice
  - Abilita la compressione specialistica di dati genomici su Hadoop, dando ampia scelta
  - Compressione specialistica consente forti guadagni in tempo/spazio di elaborazione di dati genomici su Hadoop
    - U. Ferraro Petrillo, F. Palini, G. Cattaneo, R. Giancarlo, FASTA/Q Data Compressors for MapReduce-Hadoop Genomics: Space and Time Savings Made Easy, To appear BMC Bioinformatics
- Lossy Graph Compression
- PPM vs Neural Nets for Compression- Multi-Sede