Ishak Yusuf

Computational Biologist

via Trento 17, Cologno Monzese ishakyusuf18@outlook.com 3894205359

Yusuf I. isky20 0000-0003-1635-1350

ABOUT

Experienced Computational Biologist with 5+ years expertise in computational analysis of diverse biological data. Yusuf is a skilled bioinformatician with a strong academic background. He holds a Master of Science degree in Bioinformatics and Molecular Genetics from the University of Leicester, where he excelled in various modules related to gene analysis, bioinformatics programming, and algorithms. His research dissertation focused on optimizing an NLP algorithm for extracting table data from GWAS publications. Prior to his master's, Yusuf completed a Bachelor's degree in Biotechnology at UCSI, where he gained experience in microbiology, pharmacology, and biochemistry. Currently, he works as a Research Fellow at Istituto di Tecnologie Biomediche, CNR, in Lombardy, Italy, focusing on evaluating bioinformatics tools for omics data analysis, particularly in proteomics and transcriptomics. The aim is to understand complex biological systems and advance research in crucial cellular processes.

EXPERIENCE

Research Fellow (Computational Biologist),
Istituto di Tecnologie Biomediche, CNR
Nov 2022 – present
Milan, Italy

I focus my attention on evaluating bioinformatics tools for analyzing omics data, mainly in the fields of proteomics and transcriptomics. Regarding proteomics, I start using and learning software such as Proteome Discoverer 2.5, MaxQuant, and Peaks. After characterizing protein profiles, I begin applying statistical methods to evaluate them and identify differentially expressed proteins using label-free quantitative approaches.

For the analysis of RNAseq data, I utilise STAR (Spliced Transcripts Alignment to a Reference) to determine the origin of reads on the reference genome. STAR is an aligner specifically designed to address the challenges of RNA-seq data mapping, taking spliced alignments into account. I install and use it through Conda, an open-source package management and environment system that is compatible with Windows, macOS, and Linux.

In this context, I am also devoting my efforts to process omics data at a holistic level, applying data-derived systems biology approaches primarily based on protein-protein interaction (PPI) network models. To achieve this, I am delving deeper into the use of software such as Cytoscape and its Apps, including String, BINGO, CenGscape, ClusterOne, etc.

Bioinformatics Scientist, *Genoks Genetics Centre* Jun 2021 – Oct 2022
Ankara, Turkey

Experienced in analyzing multi-omics data with skills in R, bash, and Python for data manipulation and visualization. Proficient in running and optimizing pipelines on HPC systems, ensuring efficient analysis of large sample sets. Effective communication of results through reports and visualizations.

Research Internship, *Qatar Medical Genetic Center* Sep 2018 – Jun 2019
Doha, Qatar

I conducted a study investigating therapy-induced changes in gene expression among Chronic Myeloid Leukemia (CML) patients. By collecting blood samples and extracting DNA, I assessed the impact of treatment on microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Utilizing reverse transcription polymerase chain reaction (RT-PCR), I analyzed the gene expression data. This work sheds light on the molecular dynamics of CML treatment response and its influence on miRNA and lncRNA expression profiles.

EDUCATION

MSc Molecular Genetics and Bioinformatics, Sep 2019 – Oct 2020
University of Leicester Leicester, England

Grade: Merit

MSc project:

- Optimising an NLP algorithm to extract table data from GWAS publications

BSc Biotechnology, *UCSI University* Oct 2014 – Aug 2018
Kuala Lumpur, Malaysia

GPA: 2.81

Research Project:

- Determination of sub-cytotoxic dose range of fungal immunomodulatory protein for potential neurodegenerative treatment.

PROJECTS

Transcriptomics in Multiple Sclerosis (MS):

I analyzed data from peripheral and induced T cell from patients affected by multiple sclerosis (MS). In particular, I evaluated the genes expression between induced T cells and peripheral T cells following three different treatments: EVs from healthy cases, EVs from MS patients and empty EVs (vehicle). The initial step involved assessing the quality of fastq reads and the alignment percentage using various tools such as FastQC, STAR, and MultiQC. In addition, in this context, the Tflink database was utilized to extract the relationship between the transcription factors (TF) and targeted genes. Starting from FoxP3, a TF of our interest, we extracted more than 7500 genes. The aim of this data processing was to compare this list of genes targeted by FoxP3 with those we found differentially expressed by comparing induced and peripheral t cells.

Proteomic Analysis in Neuroblastoma:

The project specifically involves the analysis of proteomic data from extracellular vesicles (EVs) released by neuroblastoma cells after treatment with OLIGP GM1 peptides. These data were compared with those obtained from untreated neuroblastoma cell lines. Three biological replicates were considered, analyzed in two technical replicates each. Differentially expressed proteins (DEPs) between treated and untreated neuroblastoma cell lines were identified using Linear Discriminant Analysis (LDA) (n=103, $P<0.05$). Subsequently, an enrichment analysis was applied to each sample using an R code created by rbioapi library. Significant enrichment terms were selected, and the results were visualized using ggplots. Furthermore, the characterized profiles and DEPs were used to create a protein-protein interaction network model. These models were processed at functional and topological levels for the selection of functional modules and proteins defined as hubs.

Proteomic Analysis in Prostate Cancer:

The project focused on studying proteins identified in urine samples from healthy controls (HC), patients with low-risk prostate cancer (LRP), and patients with high-risk prostate cancer (HRP). Characterized proteomic profiles (2491 total proteins) for each group were quantitatively compared for the selection of Differentially Expressed Proteins (DEPs) (n=85 through LDA ($P<0.05$); 56 through identification frequency). As mentioned earlier, an enrichment analysis was applied to each sample using an R code created by myself and based on the rbioapi library. Significant enrichment terms were selected, and the results were visualized using ggplots. Furthermore, the characterized profiles and DEPs were used to create a protein-protein interaction network model. These models were processed at functional and topological levels for the selection of functional modules and proteins defined as hubs. In this scenario, I finally created an R code to generate a tree plot representing the selected proteins/genes as hubs and the main corresponding GO terms. For this purpose, I utilized the ggtree library, the GOchord library, and the geom_fruit function.

Cardio Regeneration by miRNA Therapy:

A project where I analyzed transcriptomics and proteomics data concerns the cardio regeneration by miRNAs therapy. Data were obtained by cardiomyocytes overexpressing miRNA199, miRNA590, miR373, miRNA1825, miRNA33B and miRNA302 that induce proliferation. The aim of this project is to identify genes and proteins targeted and most affected by the overexpressed miRNAs to shed light on mechanisms inducing cardiomyocytes proliferation. The analysis included the extraction of genes and proteins targeted by miRNAs. Based on experimental data and the information provided by multiMiR, TargetScan and miRTarBase that provide information about predicted and experimentally validated miRNA targets. Cytoscape software and its APPs, including STRING and CentiScape were utilized to reconstruct the PPI network models, and analyzed at topological level, respectively. Specifically, it allowed us to identify genes and proteins we define hubs.

PUBLICATIONS

The published papers will be uploaded through the link

REFERENCES

References are available and can be provided upon request.