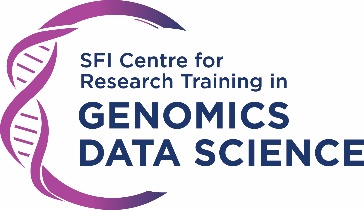
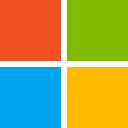
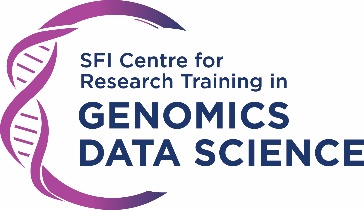
**[](https://careers.microsoft.com/us/en/job/1498955/Research-Intern-Computer-Science-for-Health?jobsource=indeed&utm_source=indeed&utm_medium=indeed&utm_campaign=indeed-feed)[](https://www.benevolent.com/careers/internships/)**

**Annual Report (max 5 pages)**

|  |  |
| --- | --- |
| Name | **Karen Guerrero Vazquez** |
| Email address | **K.GuerreroVazquez1@nuigalway.ie** |
| Institution | University of Galway |
| Supervisor | Dr. Katarzyna Goljanek-Whysall, Dr. Pilib O’Broin |
| Start date on programme | September 1st 2021 |
| Year of PhD (Year 1, 2, 3 or 4) | 2 |
| Reporting period | September 1st 2022 - September 1st 2023 |
| **Content of the report** | |
| **Research** | |
| Computational approaches for therapeutic target discovery to ameliorate  muscle wasting during aging and disease | |
| **Background**  Progressive muscle wasting is a natural consequence of aging (sarcopenia) and an important consequence of illness. The number of people affected by sarcopenia is predicted to reach 18 mln in Europe by 2045 [1]. This number will increase substantially in the light of the COVID-19 pandemic; over 50% hospitalised COVID-19 patients display significantly reduced muscle function [2] . Muscle loss is therefore becoming a health care priority. This project aligns with SFI strategic priorities and is in the area of muscle wasting during aging and disease. Both aging and critical illness lead to frailty and eventually increased morbidity and mortality. There is no cure for muscle loss. Target identification and validation is a pressing challenge, with many targets failing for efficacy reasons or showing poor association with the disease. Computational prediction of therapeutic targets could significantly decrease the attrition rates in the drug discovery pipeline by significantly reducing the initial search phase. microRNAs (miRs) are robust gene expression regulators which have been demonstrated to play key roles in regulating muscle homeostasis [3]. Multi-omics analyses on the muscle samples from adult and older humans and critically ill patients, which included proxemics, transcriptomics (including splicing) and small RNA-Seq (bionomics) have been previously performed by us and others. This project will use these data to generate a model of common epigenetic mechanisms regulating changes in muscle atrophy during aging and disease with the focus on microRNAs. The changes in human muscle will be compared with changes from in vitro and in vivo models to allow for future target(s) validation. This computational model will be next used to predict novel therapeutic targets: miRs, for muscle wasting in aging and disease. | |
| **Objectives of the Project**  The main objectives of this project include:  1. Create independent models and a common model of miR:target interactions in muscle loss during aging and critical illness based on differentially expressed (DE) miRs and genes.  2. Fine-tune the model(s) based on conserved changes between humans and in vitro and in vivo models. This objective will include our existing and analysed RNA-Seq, small RNA-Seq and proteomics datasets from human primary myoblasts and mice, as well as published data.  3. Test the model(s) for new therapeutic targets for muscle loss by in silico simulation in selected miR(s) affected by aging and critical illness to determine potential therapeutic targets. Overexpression/inhibiton of selected miR(s) will be simulated to determine the effects on expression of genes associated with muscle hypertrophy, atrophy and regeneration pathways.  4. Test the selected therapeutic target in an in vitro model of muscle loss through overexpression and inhibition of a selected miR predicted to regulate muscle loss in aging and disease.  **Significance**. Steady-state mRNA levels measured in transcriptomic studies provide important information on the modulation of cellular phenotype under different conditions, however there is often a discord between mRNA and protein abundance. Often these studies fail to fully inform us about the dynamic post-transcriptional control of mRNA turnover and translation This project will be advantageous to ongoing studies using animal models and will reduce the number of animals used. Moreover, the large-scale analyses will comprehensively investigate the common changes in muscle aging and disease. Ultimately, these targets can be tested in vitro and in vivo for further therapeutic development. The focus on a global population of muscle atrophy (sarcopenia) and critical illness provides an opportunity for a significant impact in our aging population. | |
| **New methodologies learned over the past year**  MicroRNA database handling   * Known and predicted Seeds of the microRNAs * MicroRNAs targets predicted, putative and verified from different sources   Differential Analysis   * Use of Lugh server for differential expression analyses * Sequence read Quality reports (Fastqc) * Sequence pseudo alignment (Kallisto) * Adapter removal (Cutadapt) * DESeq2 for differential analysis * Pathway analysis using GSEApy   Network construction   * NetworkX * Network evaluation using PageRank and Closeness centralities * Cytoscape | |
| Description of work completed during this period  Differential Expression  From RNAseq and microarray data from five studies, with a total of 246 samples of skeletal muscle from healthy participants with ages ranging from 19 to 85 years old, we obtained differentially expressed genes from groups of young, middle age and older adults. 251 genes were identified, 125 that increase with age and 126 that decrease with age.  We separated them into 6 subsets: young vs old, young vs middle age, middle age vs old, and for each of them, up and down-regulated.  Software development  Using Python language, we develop a program that takes JSON formatted networks, adds the microRNAs from miRTarbase, miRDB, and TargetScan, and filters the result based on closeness and PageRank centralities. We had called this program miRKat. And incorporating existing code to evaluate the networks based on random walk with restart.  Network creation  From the selected genes, we include the protein-protein interactions with IntAct and a combination of co-expression, protein-protein interactions, literature, and others from Genemania. This network was subjected to the in-home program and the evaluation.  The was one network generated by each subset of genes, and from there, from using IntAct and closeness centrality, we got a total of 30 microRNA, 23 detected in Young vs Old, 3 in Young vs Middle Age, and 6 in Middle Age vs All. The selection of this preliminary list was obtained getting the top 5% centrality nodes, however, further analysis needs to be done in order to declare any of these microRNAs as part of the shortlisted.  Database  In the previous review, we showed the first glance, the development of a database that includes the microRNA data and their targets. I had successfully included more targets, add the functionality of seeds and will keep working on adding the tissue expression. This work includes the curation of the data, formatting, and parsing to be included. | |
| Discussion  Differential Expression  The experiments where the genes were extracted are incomparable among each other, finding common genes. For example, when we evaluate Young vs Old and Middle age vs Old, we encountered 8 genes in common. The probability of this happening is around 0.004% (calculated with a quick hypergeometric test), however, more tests are needed to calculate the probability of obtaining the observed overlap between two groups by chance in each group.  However, having the same genes appearing as differentially expressed in multiple experiments can add value to the overall analysis. It indicates that these genes may be more biologically relevant and potentially important in the context of the tissue and experimental conditions being studied.  Software development  miRKat seems potential to be a key tool in the selection of genes and microRNAs given the initial set. However, more tests are needed to ensure the quality of the microRNAs, this is, that the selected microRNAs are relevant to muscle biology and indeed are the most relevant within the network.  In terms of the code itself, a test suite is integrated to ensure the code works as expected. As miRKat grows in complexity, this suit gets more value.  We are still left to integrate the tissue presence and the type of interactions alongside Transcription Factor data.  Finally, the scoring of the selected nodes does not discriminate among original differentially expressed genes, genes with their interactions, or microRNAs, it could improve the selection if we consider these elements in the selection of the genes and miRNAs.  Network creation  Until now, we have two ways to extend from the original differentially expressed genes; IntAct and GeneMania. The first uses protein-protein interactions and the second a combination of 5 parameters. The first approach gives more relevant microRNAs than the second, it is contemplated that the amount of extra information acts like noise for the model.  More work is necessary to compare the resulting microRNAs from the different genes subsets (young vs old, young vs middle age, middle age vs old, and for each of them, up and down-regulated)  Database  The database miRKat is starting to get the shape and had already shown utility outside of this project. More refinement is vital before it can be published to a broad public.  Documentation is missing and usability is limited by MySQL knowledge. Plans to take advantage of Natural Language Processing Models are present to allow anybody to ask complex queries more intuitively. | |
| **Communications to other scientists**  **09 September 2022**  EMBL-EBI: Slack and handbook closure for Mathematics of life: modelling molecular mechanisms 2022 Poster presentation  **02 December 2022**  Irish Computational Biology in Genomics Poster presentation and organization  **20,21 April 2023**  European Mathematical Genetics Meeting. Poster presentations  Mathematics Research Day. Poster Presentation  **12 May 2023**  The Virtual Institute of Bioinformatics & Evolution. Poster Presentation and lightning talk  **08 – 11 June 2023**  American Aging Association anual meeting 2023 Poster presentation | |
| **Future workplan/ schedule**   * Get a concise way to evaluate the networks, measure by the probability of the microRNA to be found in muscle and the effect on the disease * Integrate the obtained six networks from three different ages comparisons points into a single network, and evaluate the performance of the integrated network by comparing it to known biological pathways and performing functional enrichment analysis * Select a few microRNAs based on their relevance in the networks for in vitro experimentation. * Define and clean up miRKat database to be deploy to public servers. * Generate model with selected microRNAs and genes selected and tested. * Publish miRKat database * Publish miRKat * Publish model | |
| **Training** | |
| EMBL-EBI: Mathematics of life: modelling molecular mechanisms 2022 course  I learned it the usage of softwares:   * Cytoscape:   + Esceincial for the visual manipulation of networks * Copasi   + Tool for the develop of biological models * NetworkX   + Library containing the base for network analysis * Cutadap   + Tool for trimming and cleanning of RNA sequences * DESeq2   + Library for the differential expression analysis | |
|  | |
| **Factors affecting progress** | |
| The placement selection has blockers due to the nature of the type of placement I am aiming for | |
| **Placement** | |
| Placement hadn’t been agreed yet. There had been applications sended to the ones showed in the following table:   |  |  |  |  | | --- | --- | --- | --- | | Company | Program | Application date | Status | | Jonson and Jonson | R&D Data Sciences Intern Summer 2023 | 14 March 2023 | Rejected | | Calico | Data Science - Statistical Genetics (PhD) | 13 March 2023 | Rejected | | Genentech | 2023 Summer Intern - Data Engineering | 13 March 2023 | Rejected | | Bio-Techne | Computational Biologist Intern | 13 March 2023 | Rejected | | Myriad | - | February 2023 | Rejected | | United Health | Research Scientist PhD or Masters Intern - Optum Genomics | 14 March 2023 | Waiting | | Microsoft | Research Intern - Computer Science for Health | 20 March 2023 | Waiting | | BenevolentAI | 2023 Internship Talent Pool | 16 ene 2023 | Waiting | | |
| **Education and Public Engagement** | |
| **22 May 2023**  Pint of Science presentation | |
| **Awards, Prizes and Distinctions** | |
| vitaDao hackathon Lonhack: Second place  team AdStella  AdStella is focused on developing sarcopenia detection software, initially for pre-surgical patients. The software will help measure sarcopenia pre-operationally to provide better patient care and save time and costs.  More information in the Longhack site: [https://longhack.org/vitadao#!/tproduct/365899158-1628509562635](https://longhack.org/vitadao" \l "!/tproduct/365899158-1628509562635) | |
| **Exploitation** | |
|  | |
|  | |
|  | |

Date & Signature of Student: Date & Signature of supervisor

****

**Personal Career Developme****nt Plan (1-3 pages)**

**Please develop your personal career development plan (PCDP) using this template**. Please get input from your supervisor. The PCDP will have to be updated every year to incorporate any changes to the plan that are required. For help to explore potential long term career options suiting your interest and skills, you can complete the self-assessment tool provided by ScienceCareers here: <https://myidp.sciencecareers.org/Overview/Summary>

|  |  |
| --- | --- |
| Name | **Karen Guerrero Vazquez** |
| Email address | [**K.GuerreroVazquez1@nuigalway.ie**](mailto:K.GuerreroVazquez1@nuigalway.ie) |
| Institution | University of Galway |
| Supervisor | Dr. Katarzyna Goljanek-Whysall |
| Start date on programme | September 1st 2021 |
| Reporting period |  |
|  | |
| **Abstract of your research project**  Progressive muscle wasting is a natural consequence of aging (sarcopenia) and an important consequence of illness. The number of people affected by sarcopenia is predicted to reach 18 mln in Europe by 2045 [1]. This number will increase substantially in the light of the COVID-19 pandemic;  over 50% hospitalised COVID-19 patients display significantly reduced muscle function [2] . Muscle loss is therefore becoming a health care priority. This project aligns with SFI strategic priorities and is in the area of muscle wasting during aging and disease. Both aging and critical illness lead to frailty and eventually increased morbidity and mortality. There is no cure for muscle loss. Target  identification and validation is a pressing challenge, with many targets failing for efficacy reasons or showing poor association with the disease. Computational prediction of therapeutic targets could significantly decrease the attrition rates in the drug discovery pipeline by significantly reducing the initial search phase. microRNAs (miRs) are robust gene expression regulators which have been  demonstrated to play key roles in regulating muscle homeostasis [3]. Multi-omics analyses on the muscle samples from adult and older humans and critically ill patients, which included proxemics, transcriptomics (including splicing) and small RNA-Seq (bionomics) have been previously performed by us and others. This project will use these data to generate a model of common epigenetic mechanisms regulating changes in muscle atrophy during aging and disease with the focus on microRNAs. The changes in human muscle will be compared with changes from in vitro and in vivo models to allow for future target(s) validation. This computational model will be next used to predict novel therapeutic targets: miRs, for muscle wasting in aging and disease. | |
| **Long-term career Goals (over 5 years): (2-4 bullet points)** | |
| 1. Get a software manager position in an aging-related company 2. Have tools published and widely used for researchers. | |
| **What further research activity or other training is needed to attain these goals?**   1. Get experience inside the tech or health industry in the shape of internships 2. From the material I have, polish and publish the tools that I am creating | |
| **Short-term Goals (1-2 years) in terms of** | |
| 1. Research results e.g.    * Finalise microRNAs that are going to be part of the final model    * Combine the different dichotomy networks into one network with evolution across time    * Have miRKat publicly available 2. Research skills and techniques:    * Comprehensive manage of COPASI    * Natural Language Models 3. Research management:    * Getting placements inside the industry 4. Communication skills training    * Improve conversational English level taking Conversational English classes at the university 5. Other professional training (course work, teaching activity)    * Get involved in science communication for lay audience | |

Date & Signature of Student: Date & Signature of supervisor