*National University of Ireland, Galway*

**Annual Student Report to GRC**

*To be completed by all research (PhD, MD and Research Masters) students every year and submitted to the GRC in advance of the annual review meeting with the GRC.*

|  |  |
| --- | --- |
| **Name of Student** | Karen Guerrero Vazquez |
| **Student ID** | 21124191 |
| **Year of Study** | 2 |
| **PhD / MD / Research Masters** | PhD |
| **Discipline / School** | School of Maths |
| **Full or Part Time** | Full time |
| **Name of Supervisor(s)** | Goljanek-Whysall, Katarzyna and Pilib Ó Broin |
| **Period covered by report** | September 1st 2022 - September 1st 2023 |

|  |
| --- |
| **Description of work completed during this period:**  **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.  **Compleated work:**  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. |

|  |
| --- |
| **Indicate any communications of your work or relevant articles submitted for publication or published during this period:**  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 |

|  |
| --- |
| **Completion Plan**  Students nearing completion must provide a completion plan  i.e.all full-time PhD/MD students in years 3, 4 and later (part-time students in years 4, 5, 6 and later) and all full-time Research Masters students in all years (part-time students in years 2 and later).  Plan must include tasks to be completed during the next year with timeframe.  A Gantt chart is appropriate.  Thesis writing should also be included along with publication plans.  *Enlarge this box as necessary.* |

|  |
| --- |
| When do you plan to submit your thesis? |

***Note:*** *Research students past their T****ime Limit*** *(i.e.* ***after*** *4 years for a full-time PhD, 6 year part-time PhD,* ***after*** *2 years for full-time Masters and 3 years part-time Research Masters students) should meet more frequently with their GRC e.g. quarterly.*

**For students on structured research programmes**

Complete the tables below indicating the taught modules you have taken this academic year.

|  |  |  |
| --- | --- | --- |
| **GS Modules to be assessed by supervisor**  (e.g. GS501) | | |
| **Code** | **Module Title** | **ECTS** |
| MA653 | Foundations of Genomics Data Science | 10 |
| MA654 | Advanced Genomics Data Science | 10 |
| MA655 | Applications of Genomics Data Science | 10 |

|  |  |  |
| --- | --- | --- |
| **GS Modules (with module owners)**  (e.g. GS506) | | |
| **Code** | **Module Title** | **ECTS** |
|  |  |  |
|  |  |  |
|  |  |  |

List of GS modules can be found at <http://www.nuigalway.ie/graduatestudies/module_table.html>

|  |  |  |
| --- | --- | --- |
| **Advanced Specialised Modules**  (generally discipline-specific, e.g. CH503) | | |
| **Code** | **Module Title** | **ECTS** |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
| Have you successfully completed any module in another Irish university during this period?  If yes, provide details here and attach evidence of successful completion.  *It is important that you have registered for such modules at NUI Galway. College offices can advise on these procedures.* |

|  |
| --- |
| Have you completed the minimum number of taught modules required in your structured research programme?  If not, what modules do you plan to take next year?  *Module selection must be agreed with your supervisor.* |

Please attach to this report:

1. Your registration statement for this academic year indicating the modules for which you have registered (available from <http://www.nuigalway.ie/registration/>) and
2. For students in year 2 and later, copies of transcripts for earlier years of your research programme indicating the modules you have completed successfully (available from the Examinations Office).

|  |
| --- |
| **Student’s signature:** |

|  |
| --- |
| **Date:** |

**University Guidelines for Research Degree Programmes**

<http://www.nuigalway.ie/graduatestudies/documents/university_guidelines_for_research_degree_programmes.pdf>