L-Università ta' Malta

M.Sc. in Bioinformatics

Project Proposals for MMB5010 – Dissertation

2022-2023 Intake

February 2023

Table of Contents

Title (ID.)	Isolation of unique cancer-generated methylations (CanPep)
Principal Supervisor	Dr Byron Baron
Contact Email	byron.baron@um.edu.mt
Co-Supervisor (if any)	Mr Joseph Bonello
Project Description	Protein methylation alters many cellular functions and the process of carcinogenesis brings about numerous changes in the cell so as to overcome apoptosis, ensure survival to stressful conditions, grow and metastasise. Mass spectrometric analysis of matched healthy and cancer proteomes reveal the presence of unique modifications in the cancer proteome but these are not always easily detectable, measured and validated because the length and therefore peptide sequence is not always identical after the addition of the methylation (an event referred to as missed cleavage). The aim of this project is therefore to query proteomic datasets for unique cancer-generated peptides with methylations to be used as diagnostic or prognostic markers
Outcomes	■ A query system to search for cancer-generated peptides with specific parameters ■ A validation tool to measure relative abundance of cancer-generated peptides ■ A shortlist of cancer-generated peptides with diagnostic or prognostic potential
Keywords / Main Areas	Comprehensive mass spectrometry; shotgun proteomics; post-translational modifications; protein methylation, methyloproteomics; label-free analysis
References	Hart-Smith, G., Yagoub, D., Tay, A.P., Pickford, R. and Wilkins, M.R., 2016. Large scale mass spectrometry-based identifications of enzyme-mediated protein methylation are subject to high false discovery rates. *Molecular & Cellular Proteomics*, *15*(3), pp.989-1006. Tay, A.P., Geoghegan, V., Yagoub, D., Wilkins, M.R. and Hart-Smith, G., 2017. MethylQuant: A Tool for Sensitive Validation of Enzyme-Mediated Protein Methylation Sites from Heavy-Methyl SILAC Data. *Journal of proteome research*, *17*(1), pp.359-373.

Title (ID.)	Investigating RNA editing in human monocytes during community-acquired pneumonia (RNAEdit)
Principal Supervisor	Brendon P. Scicluna
Contact Email	brendon.scicluna@um.edu.mt
Co-Supervisor (if any)	N/A
Project Description *Not more than 200 words*	RNA editing is a process that modifies the primary RNA in post-transcriptional mechanism mediated by adenosine or cytidine deaminases.[1] RNA editing has been implicated in various diseases including cancer and neurological diseases. Moreover, RNA editing has also been related to cancer heterogeneity, the onset of carcinogenesis, and the response to treatment. Whether RNA editing processes have a role in the host response during community-acquired pneumonia (CAP) is unknown. The hypothesis of the study is that **RNA editing modifies the functionality of human monocytes during CAP**. To test this hypothesis, the following aims will be addressed: ■ Identify RNA editing processes in RNA of human monocytes obtained from patients during the acute and recovery stage of CAP ■ Test the robustness of identified base conversions in publicly available transcriptomics data Primary RNA-sequencing data is available through the ELDERBIOME project[2,3] with circulating monocytes obtained from a longitudinal, observational cohort study, **75 CAP patients** (> 18 years) were included between October 2016, and June 2018, at the Internal Medicine ward or Intensive Care Unit of the Academic Medical Center and BovenIJ Hospital (Amsterdam, the Netherlands) within 16 h of hospital admission and revisited after 1 month. In addition, 41 age and sex-matched control participants were included in the study. All data is readily available.
Outcomes	■ Identification of A-to-I and/or C-to-U conversions in monocytes obtained from CAP patients relative to control (age, sex-matched) participants ■ Validation of the conversions in external, publicly available data of human monocytes from other CAP cohorts.
Keywords / Main Areas	Pneumonia; monocytes; RNA editing; epitranscriptomics
References	1) Christofi, T., Zaravinos, A. RNA editing in the forefront of

Additional or suggested reading	epitranscriptomics and human health. J Transl Med 17, 319 (2019). 2) Brands X. et al. Genome Med. 2021 Aug 16;13(1):131. 3) Schuurman A. et al. EBioMedicine. 2022 Jul;81:104082.

Title (ID.)	Variants of Unknown Significance (VUS) Curation (VUSCur)
Principal Supervisor	Dr J. P. Ebejer
Contact Email	jean.p.ebejer@um.edu.mt
Co-Supervisor (if any)	Dr Jeanesse Scerri
Project Description	Clinical classification of variants generated from next-generation sequencing (gene panels/exomes) is done following the ACMG guidelines (Richards et al. 2015) and utilising a number of databases and *in silico* tools, such as: ■ ClinVar: A database of variants together with interpretations/classifications submitted by international genomic clinics and laboratories ■ Population databases such as GnomAD: the lower the variant frequency (VF) in a population, the more likely a variant is to be pathogenic ■ *In silico* prediction algorithms: predict the effect of the variant on the structure & function of the protein, including the creation or destruction of splice sites (e.g. MutationTaster, Polyphen-2, SIFT) ■ Other gene annotation & conservation scores: the more an amino acid site is conserved along evolutionary lines, the more important it is and thus the more likely a variant creating a change in this amino acid is pathogenic. ■ Peer-reviewed literature ■ ACMG classification is in 5 tiers: ■ Pathogenic ■ Likely Pathogenic ■ Variant of Unknown Significance (VUS) ■ Likely Benign ■ Benign VUS, for which not enough evidence exists to suggest pathogenicity or otherwise, or the existing evidence is conflicting, are a "by-product" of high-throughput sequencing and a challenge for genetic diagnostics. The current practice is not to report these variants, but to record them and periodically review and curate them. If additional evidence suggesting pathogenicity of a VUS emerges, leading to its re-classification into pathogenic or likely pathogenic, the genetic consultants caring for the patients and their family members carrying such a variant must

	be notified. This practice results in an ever-increasing VUS list which has to be regularly manually curated. This is a very laborious and tedious activity which most of the time does not generate any re-classification of variants.
Outcomes	The development of a solution that enables: a) storage of variants in a database, including easy access to the individuals harbouring the variants for recalling purposes and quality assurance/audit trail capabilities (e.g. recorded date of review and next review date); b) automatic curation of the variants with flagging of new evidence, by the utilisation of all the tools mentioned above and also a literature search by using the variant details as keywords.
Keywords / Main Areas	Clinical Genomics; VUS; Variant Classification; Variant Curation
References	Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PMID: 25741868; PMCID: PMC4544753.

Title (ID.)	Extending SADIP – a resource for unified protein study (ExtSADIP)
Principal Supervisor	Mr Joseph Bonello
Contact Email	joseph.bonello@um.edu.mt
Co-Supervisor (if any)	N/A
Project Description	SADIP stands for Semi-automated data integration system for protein databases. The project tries to offer a unified view of proteins by aggregating data from several databases. It is a particularly valuable resource as it can be a one-stop shop for biologists studying proteins. We want to extend the backend of the existing project, mainly by using Big Data techniques to improve the assimilation of data from various sources and improving the UX of the project as a whole. Technologies used: ■ Back-End (ETL): Kotlin, Apache Spark, Hadoop Distributed File ■ System (HDFS), Apache Parquet, ArcadeDB ■ Back-End (REST API): Kotlin, Spring Boot ■ Front-End (One-stop shop portal): React (JavaScript), HTML, CSS
Outcomes	The intended outcomes include: ■ improving SADIP in terms of the scalability of the backend ■ improve SADIP in terms of the presentation to biologists ■ add new datasources as required (eg AlphaFold)
Keywords / Main Areas	Protein Classification
References	SADIP Thesis https://www.disprot.org/help http://www.moonlightingproteins.org/publications/

Title (ID.)	microRNA target prediction using machine learning (MLmRNA)
Principal Supervisor	Dr Panagiotis Alexiou
Contact Email	panagiotis.alexiou@um.edu.mt
Co-Supervisor (if any)	N/A
Project Description	microRNAs are small regulatory RNAs that are central to post-transcriptional regulation of gene expression. The rules of microRNA targeting are not yet fully understood. Specifically, an important question is how multiple microRNA binding sites on the same target gene, may work together to produce the intended repression. After recent experimental developments, we now have an increasing amount of experimental evidence for microRNA:target site interaction, as well as the effect a microRNA knockout or overexpression has on the transcriptome. Using this experimental data as a training set, we will produce a machine learning model (possibly Recurrent Neural Network) that can predict the cumulative effect of multiple microRNA binding sites on their target genes. We will be building upon already developed interpretable machine learning methods for microRNA target site prediction based on Convolutional Neural Networks (article in preparation). The project will involve collaboration with early stage researchers at the Central European Institute of Technology (CEITEC) in the Czech Republic.
Outcomes	I expect that successful completion of the project will lead to a publication of a research article, and/or presentation in an international conference.
Keywords / Main Areas	MicroRNA; machine learning; neural networks
References	Experimental data: https://doi.org/10.1101/2022.02.13.480296 Computational method: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6067737/

Title (ID.)	Identification of conserved sequences in phage displayed, single chain antibodies targeting SARS-CoV-2 antigens (ConSeqSars)
Principal Supervisor	Prof. David G. Saliba
Contact Email	david.saliba@um.edu.mt
Co-Supervisor (if any)	Dr Jean-Paul Ebejer
Project Description	The aim of this project is to characterise the amino acid sequence of enriched and phage displayed single chain (scFV) an bodies targetting SARS-CoV2 antigens. This project is an offshoot of the ACCELERATE project funded by the IDP-MCST. We have generated polyclonal scFV antibody populations and monoclonal antibodies that bind specifically to either SARS-COV-2 trimer, S1, RBD or ZikaV envelope by phage display biopanning. We have purified the DNA and will be performing long-read NGS using Oxford Nanopore technology of polyclonal populations. Sanger sequencing will also be performed for monoclonal an bodies. A pipeline to analyse the DNA sequences coding for the scFV will have to be set up during this MSc.
Outcomes	**1. Determine Read lengths** Check for Open Reading Frame Determine nucleotide to protein translation; Determine Linker sequence is “conserved” Determine and discard Clones without FR1 for both VH and/or VL; Determine and discard clones without FR4 or myc tag for both VH and/or VL. **2. Check for stop codons** Determine and discard any sequences that contain Stop codons in ScFV nucleo de sequence **3. Determine sequences that are common to non-protein target round** **4. Check for conserved sequences “Dynamic range”** Align and determine common VH (CDR3?) sequences Determine frequency of common scFV sequences (motif analysis) Determine frequency of all scFV sequences Determine phylogenetic relationships between clones
Keywords / Main Areas	NGS; Nanopore; scFV antibodies; Phage library

References	https://doi.org/10.3390/cancers14051325

Title (ID.)	Using Epigenetic Modifiers to Sensitise Acute Myeloid Leukaemia to Retinoic Acid (EpigAemia)
Principal Supervisor	Dr Analisse Cassar
Contact Email	analisse.cassar@um.edu.mt
Co-Supervisor (if any)	Dr Panagiotis Alexiou
Project Description	The projects aims to reapply already existing treatments to acute myeloid leukaemia (AML) for a better prognosis. All-trans retinoic acid (ATRA) differentiation therapy has been the greatest cancer success story to date with a 99% fatal leukaemia to become over 70% curable (1). This state-of-the-art treatment is however applicable to less than 10% of all leukaemia subtypes. It has been confirmed that a combination of epigenetic modifiers as a pre-treatment to ATRA will cause differentiation of three types of AML cell lines. The treatment was tested also on leukaemia patient blood samples with positive effects. The chemicals used are already FDA-approved and have also been tested on healthy lymphocytes to ensure there is no toxicity. With effective results on multiple leukaemia types, this one-year project will focus on identifying the method of action of the treatments and the molecular mechanisms involved using RNAseq and CHIPseq following the best treatments on each type of leukaemia. This will increase the potential of acquiring a patent for the reapplication of these already-approved drugs for the treatment of leukaemia in the aim of creating a new treatment regimen that will provide AML patients with new therapeutic options.
Outcomes	■ Optimised treatments for each AML ■ Cell cycle effects of treatments on each leukaemia ■ Effects of differentiation on each leukaemia type ■ Mechanisms of action for each treatment at an RNA and Chromatin level.
Keywords / Main Areas	Epigenetics; AML; Differentiation; RNAseq; CHIPseq
References	Stahl M, Tallman MS. Acute promyelocytic leukemia (APL): remaining challenges towards a cure for all. Leuk Lymphoma. 2019 Dec;60(13):3107-3115. Doi: 10.1080/10428194.2019.1613540. Cassar A. Investigating the effect of insect conditioned medium and its constituents on the terminal differentiation of leukaemia. 2017 April;

Unpublished Doctor of Philosophy, University of Malta, Msida

Schmid C, Schleuning M, Schwerdtfeger R, Hertenstein B, Mischak-Weissinger E, Bunjes D, et al. Long-term survival in refractory acute myeloid leukemia after sequential treatment with chemotherapy and reduced-intensity conditioning for allogeneic stem cell transplantation. Blood 2006;108(3):1092-1099.

Zhu CY, Chen GF, Zhou W, Hou C, Wang XK, Wang FY, Yang N, Wang L, Fang S, Luo L, Guan LX, Zhang R, Liu YC, Dou LP, Gao CJ. Outcome and Prognostic Factors of High-Risk Acute Myeloid Leukemia After Allogeneic Hematopoietic Stem Cell Transplantation. Ann Transplant. 2019 Jun 7;24:328-340. doi: 10.12659/AOT.915381.

Bolden JE, Peart MJ, Johnstone RW. Anticancer activities of histone deacetylase inhibitors. Nat Rev Drug Discov. 2006 Sep;5(9):769-84. doi: 10.1038/nrd2133.

Sajadian SO, Ehnert S, Vakilian H, et al. Induction of active demethylation and 5hmC formation by 5-azacytidine is TET2 dependent and suggests new treatment strategies against hepatocellular carcinoma. Clin Epigenetics. 2015;7(1):98. Published 2015 Sep 11. doi:10.1186/s13148-015-0133-x

Kuendgen A, Bug G, Ottmann OG, Haase D, Schanz J, Hildebrandt B, Nachtkamp K, Neukirchen J, Dienst A, Haas R, Germing U, Gattermann N. Treatment of poor-risk myelodysplastic syndromes and acute myeloid leukemia with a combination of 5-azacytidine and valproic acid. Clin Epigenetics. 2011 Aug;2(2):389-99. doi: 10.1007/s13148-011-0031-9. Epub 2011 Apr 8. PMID: 22704349; PMCID: PMC3365387.

Kuendgen A, Knipp S, Fox F, Strupp C, Hildebrandt B, Steidl C, Germing U, Haas R, Gattermann N. Results of a phase 2 study of valproic acid alone or in combination with all-trans retinoic acid in 75 patients with myelodysplastic syndrome and relapsed or refractory acute myeloid leukemia. Ann Hematol. 2005 Dec;84 Suppl 1:61-6. doi: 10.1007/s00277-005-0026-8. PMID: 16270213.

Lübbert M, Kuendgen A. Combining DNA methyltransferase and histone deacetylase inhibition to treat acute myeloid leukemia/myelodysplastic syndrome: achievements and challenges. Cancer. 2015 Feb 15;121(4):498-501. doi: 10.1002/cncr.29083. Epub 2014 Oct 21. PMID: 25336205.

Tambaro FP, Dell'aversana C, Carafa V, Nebbioso A, Radic B, Ferrara F, Altucci L. Histone deacetylase inhibitors: clinical implications for hematological malignancies. Clin Epigenetics. 2010 Sep;1(1-2):25-44. doi: 10.1007/s13148-010-0006-2. Epub 2010 Jul 28. PMID: 22704087; PMCID: PMC3365365.

Wagner JM, Hackanson B, Lübbert M, Jung M. Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy. Clin Epigenetics. 2010 Dec;1(3-4):117-136. doi: 10.1007/s13148-010-0012-4. Epub 2010 Nov 9. PMID: 21258646; PMCID: PMC3020651.

Scaglioni PP, Cai LF, Majid SM, Yung TM, Socci ND, Kogan SC, Kopelovich L, Pandolfi PP. Treatment with 5-azacytidine accelerates acute promyelocytic leukemia leukemogenesis in a transgenic mouse model. Genes Cancer. 2011 Feb;2(2):160-5. doi: 10.1177/1947601911410300. PMID: 21779489; PMCID: PMC3111249.

Soriano AO, Yang H, Faderl S, Estrov Z, Giles F, Ravandi F, Cortes J, Wierda WG, Ouzounian S, Quezada A, Pierce S, Estey EH, Issa JP, Kantarjian HM, Garcia-Manero G. Safety and clinical activity of the combination of 5-azacytidine, valproic acid, and all-trans retinoic acid in acute myeloid leukemia and myelodysplastic syndrome. Blood. 2007 Oct 1;110(7):2302-8. doi: 10.1182/blood-2007-03-078576. Epub 2007 Jun 27. PMID: 17596541.

Title (ID.)	Estimating the age of founder variants in the Maltese Population (AgeFVar)
Principal Supervisor	Dr Rosienne Farrugia
Contact Email	rosienne.farrugia@um.edu.mt
Co-Supervisor (if any)	Dr Jean Paul Ebejer
Project Description	Recurrent occurrence of the same rare pathogenic variant in multiple patients is attributed to either a mutation hotspot or to changes in population structure whereby a rare variant introduced into a small population becomes over represented as the population grows, often exponentially, within a relatively short period of time. These can be differentiated by looking at the genetic background on which the mutation occurs. Founder variants typically occur on the same haplotype (the same LD block) in different individuals. The Maltese gene pool is characterised by a number of such founder variants. These variants, a number of them pathogenic, are rare worldwide but found at a higher frequency in the Maltese population. Genetic studies of a number of rare diseases have identified a number of founder variants in the Maltese. Using data from 1000 Whole genome sequences from Maltese individuals and the DMLE+ tool, these variants will be investigated to determine whether these are truly founder variants and how long ago they were introduced into the Maltese population. This might be extended to a genome wide analysis of variants which are overabundant in the Maltese.
Outcomes	The age and founder status of a number of pathogenic variants will be determined.
Keywords / Main Areas	Mutation age; founder variants
References	https://pubmed.ncbi.nlm.nih.gov/31043424/ https://pubmed.ncbi.nlm.nih.gov/11410841/ https://pubmed.ncbi.nlm.nih.gov/12075030/

Title (ID.)	Revisiting Small-Molecule Conformer Generation (CONFGEN)
Principal Supervisor	Dr Jean-Paul Ebejer
Contact Email	jean.p.ebejer@um.edu.mt
Co-Supervisor (if any)	N/A
Project Description	Conformer generation (generating geometries and structures for small molecules) has many important applications in Computer Aided Drug Discovery and Design, such as use in 3D virtual screening methods. In this project we will review and explore the performance of a number of modern freely-available conformer generation methods (*e.g.* ETKDG in RDKit). This work is based on the work of Ebejer et al. (over a decade ago), using a more modern validation dataset and tools and explores the landscape of conformer generation tools. To evaluate the conformer-generation methods we will measure their ability to reproduce experimentally resolved small-molecules in the Protein Data Bank. This work is well-suited to a computational person with a strong interest in small-molecule drug discovery.
Outcomes	The two main outcomes are: 1. Development of a dataset for the evaluation of conformer generation methods 2. Comparison of a number of popular freely-available conformer generation models
Keywords / Main Areas	conformer generation; small-molecule drug discovery; molecular structure
References	Ebejer, J.-P., Morris, G. M., & Deane, C. M. (2012). Freely Available Conformer Generation Methods: How Good Are They? In Journal of Chemical Information and Modeling (Vol. 52, Issue 5, pp. 1146–1158). American Chemical Society (ACS). https://doi.org/10.1021/ci2004658 Hawkins, P. C. D. (2017). Conformation Generation: The State of the Art. In Journal of Chemical Information and Modeling (Vol. 57, Issue 8, pp. 1747–1756). American Chemical Society (ACS). https://doi.org/10.1021/acs.jcim.7b00221 Riniker, S., & Landrum, G. A. (2015). Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. In Journal of Chemical Information and Modeling (Vol. 55, Issue 12, pp. 2562–2574). American Chemical Society (ACS). https://doi.org/10.1021/acs.jcim.5b00654

Title (ID.)	Remapping of HTS data across pseudogene regions (RemapPseudo)
Principal Supervisor	Dr Panagiotis Alexiou
Contact Email	panagiotis.alexiou@um.edu.mt
Co-Supervisor (if any)	Dr Rosienne Farrugia
Project Description	Pseudogenes are non-functional highly homologous copies of functional genes or parts thereof which arise due to gene duplications. Due to the high homology – sometimes as high as 98% - mapping of High throughput sequencing (HTS) data across these regions can be problematic because a sequenced fragment may match almost perfectly to more than one genomic region. One such problematic region is the short arm of chromosome 16 which has multiple pseudogenes. Autosomal dominant polycystic kidney disease (ADPKD) is caused in the majority of cases by variants in *PKD1*, a 47kb gene which has 6 highly homologous partial pseudogenes. Even though homology is very high (97-98%), there are various single nucleotide differences across the exons of *PKD1* that can distinguish it from its pseudogenes. Analysis of HTS data shows that different pipelines and exome data vs genome data results in differences in mapping of reads across this gene. Manual perusal of the sequences of individual reads also highlights instances of mismapping of the reads to the gene when they are pseudogene fragments. Thus, a tool needs to be developed to re-map across pseudogene regions paying particular attention to known variations in DNA sequence that can identify gene from pseudogenes.
Outcomes	A bioinformatics tool that uses prior knowledge of gene and pseudogene sequence to accurately map HTS reads from homologous gene/pseudogene regions to the appropriate place in the genome. This tool would be applied to 20 ADPKD HTS datasets and 1000 WGS datasets from non-ADPKD individuals.
Keywords / Main Areas	Pseudogenes; mapping of individual HTS reads; ADPKD
References	https://pubmed.ncbi.nlm.nih.gov/11414761/ https://pubmed.ncbi.nlm.nih.gov/35707588/ https://pubmed.ncbi.nlm.nih.gov/27429446/

Title (ID.)	Exploring the Mechanisms of Polyphenols in the Mediterranean Diet associated with neurodegenerative diseases through Systematic Bioinformatics Analyses (Polyphenols)
Principal Supervisor	Dr Mario Caruana Grech Perry
Contact Email	mario.caruana@um.edu.mt
Co-Supervisor (if any)	Dr JP Ebejer
Project Description	Alzheimer's disease (AD) and Parkinson's disease (PD) are the most common age-related neurodegenerative disorders and hence pose remarkable socio-economical burdens to both families and state. Dietary polyphenols exert beneficial effects on neurodegeneration in humans. Molecular mechanisms, however, are not completely understood. In addition, research into the chemistry and biology of polyphenol bioactives is prolific but knowledge of their molecular interactions with proteins is limited (Lacroix et al 2018). The aim is to conduct an in-depth integrative bioinformatic analyses to elucidate established proteins in neurodegeneration that interact with or metabolise polyphenols with the objective of mapping underlying neuroprotective effects of specific polyphenols belonging to Mediterranean foods. The project will primarily conduct a systematic literature search to identify studies with polyphenols in the Mediterranean diet that demonstrate improvement of neurodegenerative risk factors through polyphenols-protein interactions. Thereafter a series of bioinformatic analyses will be performed for the identification of molecular mechanisms underlying the beneficial nutraceutical effects of polyphenols and potential targets for future studies.
Outcomes	Bioinformatics analyses to see the effect of a select group of polyphenols predominantly found in Mediterranean foods on intracellular pathways specifically through polyphenols-protein interactions and identify potential targets for translational research.
Keywords / Main Areas	Bioinformatics; Polyphenols; Nutraceutical; Neurodegenerative diseases
References	Ruskovska, T.; Budić-Leto, I.; Corral-Jara, K.F.; Ajdžanović, V.; Arola-Arnal, A.; Bravo, F.I.; Deligiannidou, G.-E.; Havlik, J.; Janeva, M.; Kistanova, E.; Kontogiorgis, C.; Krga, I.; Massaro, M.; Miler, M.; Milosevic, V.; Morand, C.; Scoditti, E.; Suárez, M.; Vauzour, D.; Milenkovic, D. Systematic Bioinformatic Analyses of Nutrigenomic Modifications by Polyphenols Associated with Cardiometabolic Health in Humans—Evidence from Targeted Nutrigenomic Studies. *Nutrients* **2021**, *13*, 2326.

	https://doi.org/10.3390/nu13072326 Di Meo F, Valentino A, Petillo O, Peluso G, Filosa S, Crispi S. Bioactive Polyphenols and Neuromodulation: Molecular Mechanisms in Neurodegeneration. Int J Mol Sci. **2020** Apr 7;21(7):2564. doi: 10.3390/ijms21072564. PMID: 32272735; PMCID: PMC7178158. Lacroix, S., Klicic Badoux, J., Scott-Boyer, MP. *et al.* A computationally driven analysis of the polyphenol-protein interactome. *Sci Rep* 8, 2232 (**2018**). https://doi.org/10.1038/s41598-018-20625-5

Title (ID.)	RBP deep learning model interpretability (RBP-DL)
Principal Supervisor	Dr Panagiotis Alexiou
Contact Email	panagiotis.alexiou@um.edu.mt
Co-Supervisor (if any)	N/A
Project Description	RNA Binding Proteins (RBPs) are a class of proteins that directly bind RNA molecules. There are hundreds of known RBPs in humans currently, with more being constantly identified. Their functions regulate the whole lifecycle of RNA molecules in cells, including processes such as splicing and translation. We have previously produced state of the art machine learning techniques based on convolutional neural networks, that can accurately classify sequences as bound or unbound by several RBPs. These models work on a combination of RNA sequence information, RNA secondary structure information, and evolutionary conservation scores. The goal of this project will be to use machine learning interpretation methods such as Integrated Gradients or SHAP to interpret the learned models, and produce in depth human readable representations from them. The project may be extended, upon student request, to involve training and interpretation of new models on publicly available data from the ENCODE project. Ideal candidate should be passionate about machine learning.
Outcomes	I expect that successful completion of the project will lead to a publication of a research article, and/or presentation in an international conference.
Keywords / Main Areas	RBPs; RNA; machine learning; neural networks; interpretability
References	https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-022-08414-x

Title (ID.)	Discovery of Rare Genetic Variants unique to Maltese ALS patients (VarALS)
Principal Supervisor	Prof. Ruben J Cauchi
Contact Email	ruben.cauchi@um.edu.mt
Co-Supervisor (if any)	Dr. Jean-Paul Ebejer
Project Description	ALS is a progressive neurological disease affecting the motor neurons that control the voluntary muscles. The disease cripples the ability of patients to walk, talk, eat, and, eventually breathe. Genetics plays a strong role in causing ALS. Initial work done by the Principial Supervisor's team revealed that Maltese ALS patients have a different genetic architecture compared to their European counterparts. Nonetheless, the genes that are uniquely responsible for triggering ALS in the Maltese population remain unknown. The Project aims at identifying rare coding and non-coding variants (single nucleotide variants and/or indels) that are uniquely present in Maltese ALS patients but absent in age- and sex-matched healthy volunteers (controls). The student is expected to work with whole genome sequence datasets of patients and controls to identify rare variations, classify these as potentially damaging or tolerated and rank genes according to number of hits or disease-relevant biological pathways. The results of this study will be valuable for genetic counselling, diagnosis and the generation of ALS animal models.
Outcomes	■ Mining of ALS case-control whole genome sequence datasets ■ Identification of rare damaging variants present in cases but absent in controls ■ Ranking of ALS genes according to number of hits and/or disease-relevant biological pathways
Keywords / Main Areas	motor neuron disease; amyotrophic lateral sclerosis (ALS); genetics; genomics; genetic variants
References	Borg R., Farrugia Wismayer M., Bonavia K., Farrugia Wismayer A., Vella M., van Vugt J. J. F. A., Kenna B. J., Kenna K. P., Vassallo N., Veldink J. H., Cauchi R. J., 2021 Genetic analysis of ALS cases in the isolated island population of Malta. Eur J Hum Genet **29**: 604–614. van Rheenen et al., 2021 Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat

	Genet 53: 1636–1648. Goutman S. A., Hardiman O., Al-Chalabi A., Chiò A., Savelieff M. G., Kiernan M. C., Feldman E. L., 2022 Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. Lancet neurology. Smukowski S. N., Maioli H., Latimer C. S., Bird T. D., Jayadev S., Valdmanis P. N., 2022 Progress in Amyotrophic Lateral Sclerosis Gene Discovery: Reflecting on Classic Approaches and Leveraging Emerging Technologies. Neurol Genet **8**: e669. Brown R. H., Al-Chalabi A., 2017 Amyotrophic Lateral Sclerosis. N Engl J Med **377**: 162–172.