Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Roberto Malinverni and Marcus Buschbeck
Email	rmalinverni@carrerasresearch.org
Institution	Josep Carreras Leukaemia Research Institute (IJC)
Website	http://www.carrerasresearch.org/en/Chromatin_Metabolism_and_Cell_Fate
Group	Chromatin, metabolism and cell fate

Project

Computational genomics

Project Title:

Role of macroH2A histone variant in three-dimensional genomics context

Keywords:

Chromatin, macroH2A, HiC, Histone, R, NGS

Summary:

Chromosome conformation capture (C-)techniques allow to assess the nuclear architecture and distribution of chromatin in unprecedented level and have boosted the growth of nuclear organization field during the recent years (1). A large number of different variants of C-techniques including Hi-C and , Hi-ChIP have become routine in basic research, this has led to the creation of a massive amount of data stored in different public databases. Our laboratory has particular interest for years in the study of a particular histone variant called macroH2A (2). Histone variants replace canonical histones in a sub-fraction of the core structural units of chromatin, the nucleosomes. Recently, we demonstrated surprising impact of macroH2A on nuclear organization and heterochromatin architecture (3). The proposal of this master project is to investigate the role of macroH2A and other heterochromatin regulators, through the integration of the data created by our laboratory (Hi-C, HiChIP, ChipSeq, RNAseq) with those present in public databases. Specifically, we will address the following questions: 1. To evaluate association of macroH2A with respect to self-interacting genomic regions such as topological associated domains (TADs) and genome compartments. 2. To modify a framework of our previously created tool (regioneR (4)) to allow its application in a three-dimensional genomics context. 3. To create, pipelines and bioinformatics tools to query and visualize such a complex mass of data. Technically we will mainly use resources in R (Bioconductor) and Python, in Linux environment. High Performance Computer calculation will be carried out at CSUC (www.csuc.cat).

References:

1 - Grob S., Cavalli G. (2018) Technical Review: A Hitchhiker's Guide to Chromosome Conformation Capture. In: Bemer M., Baroux C. (eds) Plant Chromatin Dynamics. Methods in Molecular Biology, vol 1675. Humana Press, New York, NY. 2 - Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. Corujo D, Buschbeck M. Cancers (Basel). 2018 Feb 27;10(3). 3 - MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. Douet J, Corujo D, Malinverni R, Renauld J, Sansoni V, Posavec Marjanović M, Cantariño N, Valero V, Mongelard F, Bouvet P, Imhof A, Thiry M, Buschbeck M. J Cell Sci. 2017 May . 4 - regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. Bioinformatics. 2016 Jan 15.

Expected skills::

Experience in programming languages (preferably R). Basic knowledge of NGS data and Linux operating system. Enthusiasm to answer biological questions.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

This project will be co-supervised by Roberto Malinverni and Marcus Buschbeck

---

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Toni Giorgino
Email	toni.giorgino@cnr.it
Institution	Consiglio Nazionale delle Ricerche
Website	www.giorginolab.it
Group	Istituto di Biofisica

Project

Computational genomics

Project Title:

Machine learning-based assessment of SARS-CoV-2 genome variability

Keywords:

Coronavirus, machine learning, deep neural networks, genome

Summary:

We shall exploit the genetic-epidemiological evidence collected during the present SARS-CoV-2 outbreak to characterise the genomic regions with high mutation potential that could play a role in future outbreaks and in acquisition of drug resistance. Statistical learning and artificial-intelligence methods will be used to produce mutation models; the selected hot-spots will be cross-referenced in order to build early lead strategies of use in future outbreaks. The work is essentially computational. The student may work, either locally or remotely, with the computational group at the Institute of Biophysics of the Italian National Research Council, located at the University of Milan (Italy). Further collaborations are possible.

References:

* Smith M, Smith JC. Repurposing Therapeutics for the Wuhan Coronavirus nCov-2019: Supercomputer-Based Docking to the Viral S Protein and Human ACE2 Interface. 2020 Feb 20 (Chemrxiv) * https://viralzone.expasy.org/8996 * www.giorginolab.it * https://users.unimi.it/biolstru/molbd3-lab.html

Expected skills::

The project is heavily computationally focused. A good grasp of Python and an interest in machine learning are essential.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

Also: structural bioinformatics. Contact: toni.giorgino@cnr.it

---

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Toni Gabaldón
Email	toni.gabaldon.bcn@gmail.com
Institution	Barcelona Supercomputing Centre

Website	http://cgenomics.org
Group	Comparative Genomics

Project

Computational genomics

Project Title:

Evolution of hybrid genomes

Keywords:

Hybridization, genome evolution, phylogenomics, pathogens

Summary:

Evolution of eukaryotic species and their genomes has been traditionally understood as a vertical process in which genetic material is transmitted from parents to offspring along a lineage, and in which genetic exchange is restricted within species boundaries. However, mounting evidence coming from comparative genomic studies indicates that this paradigm is often violated. Horizontal gene transfer and mating between diverged lineages blur species boundaries and complicates the reconstruction of evolutionary histories of species and their genomes. Non-vertical evolution might be more restricted in eukaryotes as compared to prokaryotes, yet it is not negligible and can be common in certain groups. Recognition of such processes brings about the need to incorporate this complexity in our tools and models, as well as to conceptually re-frame eukaryotic diversity and evolution. In this project you will work on several hybrid genomes, including those of some pathogenic species, using comparative genomics and populations genomics tools.

References:

https://www.ncbi.nlm.nih.gov/pubmed/28681409

Expected skills::

Python, Phylogenetics, Variant calling analysis,

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

Alternative projects, within the scope of interests of the group (see publications and webpage) can be discussed.

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Julio Rozas
Email	jrozas@ub.edi
Institution	Universitat de Barcelona
Website	http://www.ub.edu/molevol/EGB/
Group	Evolutionary Genomics & Bioinformatics

Project

Computational genomics

Project Title:

Comparative and evolutionary analysis of repetitive elements in spider genomes

Keywords:

Comparative genomics; Transposable elements; Repetitive elements; phylogenomics; Adaptive genomics; genome annotation

Summary:

Understanding the origin, amplification and functional role of repetitive sequences in eucaryotic genomes is a central question in Evolutionary Biology. Despite that modern high-throughput sequencing (HTS) technologies are currently accessible for many labs, the accurate identification and annotation of gene family is one of the major challenges in the field. This scenario will change in the near future thanks to the irruption of the so called third-generation sequencing technologies (i.e., long-read sequencing). In this sense, our research group is generating new high quality genomic data from a group of Canary Island endemic spiders (Chelicerata) using long-read sequencing technologies but also chromosome-scale assembly techniques, such as Hi-C and Chicago libraries. The objective of this project is to study the abundance, distribution and evolution of repetitive elements in chelicerates and, by extension, in arthropods, including transposable elements (TEs) and other types of repetitive sequences. A large body of evidence suggest that these elements have structural functional significance. TEs, for instance, can generate variability by movement and insertion, are responsible of defining centromeric regions, or can activate gene expression under stress conditions. Our bioinformatic study in a comparative context will enable understanding of the nature and behavior of this important genomic components. The student will participate in the identification, annotation and/or analysis of repetitive elements in complete genomes of several spiders (and chelicerates) species. For that, he/she will use high quality genome sequences (data generated by our group based on third generation sequencing technologies, and sequences already available in databases), bioinformatics tools (software and scripts to manipulate and visualizte sequences and genomic annotations, to identify repetitive elements, to conduct evolutionary genetics analyses). The basic work-flow will consist in the identification and primary annotation of repeats, the determination of families, types and classes, the estimation of gene turnover rates, or the characterization of the distribution of these repetitive sequences across chromosomes or with respect to other genomic elements, such as protein-coding genes . Many of these analyses will be carried out in our high performance computer cluster.

References:

References from our researhch group • Frías-López, C., Sánchez-Herrero, J. F., Guirao-Rico, S., Mora, E., Arnedo, M. A., Sánchez-Gracia, A. and Rozas, J. 2016.DOMINO: Development of informative molecular markers for phylogenetic and genome-wide population genetic studies in non-model organisms. Bioinformatics 32: 3753-3759. doi:10.1093/bioinformatics/btw534. • Rendón-

Anaya, M. et al. 2019.The Avocado Genome Informs Deep Angiosperm Phylogeny, Highlights Introgressive Hybridization, and Reveals Pathogen-Influenced Gene Space Adaptation. Proc. Natl. Acad. Sci. USA. 116: 17081-17089. doi: 10.1101/654285. • Sánchez-Herrero, J. F., Frías-López, C., Escuer, P., Hinojosa-Alvarez, S., Arnedo, M. A., Sánchez-Gracia, A., Rozas, J. 2019.The draft genome sequence of the spider Dysdera silvatica (Araneae, Dysderidae): A valuable resource for functional and evolutionary genomic studies in chelicerates. GigaScience 8: 1-9. doi: 10.1093/gigascience/giz099. • Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J. Sánchez-Gracia, A. 2019.Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation. Mol. Ecol. 28: 4028-4045. doi: 10.1111/mec.15199. • Vizueta, J., Sánchez-Gracia, A., Rozas, J. 2019.BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. bioRxiv XX:. doi: 10.1101/593889. • Vizueta, J., Rozas, J., Sánchez-Gracia, A. 2018.Comparative Genomics Reveals Thousands of Novel Chemosensory Genes and Massive Changes in Chemoreceptor Repertories across Chelicerates Genome Biol. Evol. 10: 1221-1236. doi:10.1093/gbe/evy081. Research Group References: (http://www.ub.edu/molevol/julio/SelPublications.html)

Expected skills::

Basic knowledge on NGS data handling and analysis, especially in genome assembly and annotation, notions of comparative genomics and transcriptomics approaches and phylogenetic methods, and experience with Linux operating systems and some of the high level programing languages commonly used in bioinformatics (Perl, Python, R).

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Anthony Mathelier
Email	anthony.mathelier@ncmm.uio.no
Institution	Centre for Molecular Medicine Norway, University of Oslo
Website	https://mathelierlab.com/
Group	Computational Biology & Gene Regulation

Project

Computational genomics

Project Title:

A pan-cancer computational study of the interplay between transcription factor binding, DNA methylation, and enhancer activity

Keywords:

DNA methylation, Transcription Factor, Enhancer, Quantitative Trait Loci, Machine Learning

Summary:

Methylation of DNA is a prominent DNA modification linked to gene expression alteration in cancers [1,2]. While DNA methyltransferase (DNMT) enzymes de-methylate DNA, Ten-Eleven Translocation (TET) proteins are involved in demethylation. As DNMTs and TETs do not bind DNA in a sequence-specific manner, how these proteins are recruited to their specific sites of action is still an open question. Further, our understanding of the cascading effect of these aberrant DNA methylation patterns on gene expression deregulation is still limited. With a better characterization of the cascading effects of DNA methylation in cancer patients, we could reveal key regulatory networks critical for an improved molecular understanding of the diseases, as we recently showed for breast cancer [3]. In this project, the Master student will perform pan-cancer computational analyses to study the interplay between TF binding, DNA methylation, and enhancer activity. Specifically, the project will aim at (1) unravelling the interplay between DNA methylation and TF-binding in cancer types with limited statistical power and (2) unravelling the interplay between DNA methylation and enhancer activities. 1. We recently computed expression-methylation quantitative trait loci (emQTL) between TF expression and methylation at high-confidence TF-DNA interaction information stored in our UniBind database [4]. emQTL highlighted an interplay between DNA 5mC marks and TF-binding and showed that the binding of key pioneer TFs at their binding sites are likely to trigger local DNA demethylation that could lead to carcinogenesis (unpublished). Unfortunately, the small sample size for some cancer types prohibited the identification of the TFs involved, due to reduced statistical power. The student will use Generative Adversarial Networks (or alike machine-learning approaches) to simulate synthetic data for both methylation and gene expression from available patient data. This approach will alleviate the statistical power bottleneck currently observed. The generated data will be used to perform emQTL analyses and highlight key TFs modulating DNA methylation landscape in these cancer genomes. 2. In the second part of the project, the emQTL framework will be extended to investigate the relationship between DNA methylation and enhancer activity. Specifically, we will use RNA-seq data mapped at enhancers annotated by the FANTOM5 consortium with DNA methylation information from both normal and cancer tissues. The results will be used to investigate how the interplay between DNA methylation, TF binding, and enhancer activity mimics cell fate transition. Indeed, recent reports found that, during cell fate transition, pioneer TFs prime inaccessible enhancers, leading to increased chromatin accessibility and loss of DNA methylation [5]. This project will equip the student with computational biology skills employed in studying gene regulation and cancer genomics. She/he will build computational workflow using Snakemake and scripts in Python, R, and bash. The master student will be introduced to and learn to handle large public cancer genomics data (from ICGC, TCGA, and BASIS) and gene regulation resources (e.g. UniBind, JASPAR).

References:

1. Suzuki T, Maeda S, Furuhata E, Shimizu Y, Nishimura H, Kishima M, et al. A screening system to identify transcription factors that induce binding site-directed DNA demethylation. Epigenetics Chromatin 2017;10:60. 2. Suzuki T, Shimizu Y, Furuhata E, Maeda S, Kishima M, Nishimura H, et al. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. Blood Adv 2017 3. Fleischer T, Tekpli X, Mathelier A, Wang S, Nebdal D, Dhakal HP, et al. DNA methylation at enhancers identifies distinct breast cancer lineages. Nat. Commun. 2017 4. Gheorghe M, Sandve GK, Khan A, Chèneby J, Ballester B, Mathelier A. A map of direct TF–DNA interactions in the human genome. Nucleic Acids Res. 2019 5. Barnett KR, Decato BE, Scott TJ, Hansen TJ, Chen B, Attalla J, et al. ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. Mol. Cell 2020

Expected skills::

Proficiency in Python, R, and/or bash, previous experience in genomics data analysis, team spirit, English proficiency, Exposure to gene regulation and cancer biology will be a plus but not a strict requirement

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Anthony Mathelier
Email	anthony.mathelier@ncmm.uio.no
Institution	Centre for Molecular Medicine Norway, University of Oslo
Website	https://mathelierlab.com/
Group	Computation Biology & Gene Regulation

Project

Computational genomics

Project Title:

Exploring the links between DNA methylation, transcription factor binding, and alternative splicing

Keywords:

Alternative splicing, DNA methylation, transcription factor, cancer genomics, gene regulation

Summary:

RNA splicing is a process involved in mRNA maturation that involves removal of introns from the pre-mRNA. The machinery engaged in this process, the spliceosome, recognizes conserved nucleotide sequences in the introns in order to promote their excision from the pre-mRNA. Alternative splicing is a process that allows the cells to selectively control which parts of a pre-mRNA will be represented in the mature mRNA to be translated into protein. Despite current knowledge, the mechanism of alternative splicing is still not fully understood. The CCCTC-binding factor (CTCF) is a transcription factor (TF) that has recently been shown to be relevant in this process as mutations in its binding site are linked to specific exon inclusion or exclusion [1]. As binding of CTCF to the DNA is altered by DNA methylation [2], the study of the impact of DNA methylation at CTCF binding site on alternative splicing could shed light on aberrant splicing patterns observed in cancers. In this project the Master student will explore the interplay between DNA methylation, transcription factor binding and aberrant alternative splicing in cancers. Specifically, we plan to: (1) detect differentially used exons in cohorts of tumor and normal samples obtained from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC); (2) identify binding sites for CTCF and other TFs in the vicinity of the differentially used exons; and (3) characterize the effects of somatic mutations and DNA methylation at these binding sites on aberrant alternative splicing. Some details are provided below. (1) The student will use the DEXSeq [3] Bioconductor package on the RNA-seq data from normal and cancer samples to detect differentially used exons. This analysis will be performed on cohorts of samples from TCGA and/or ICGC for which RNA-seq, DNA methylation, and somatic mutations are available. (2) In the second step of the project, the student will use our UniBind database of high confident direct TF-DNA interactions [4] and TF binding analyses to highlight binding sites for CTCF and TFs that could be associated with alternative splicing. A strategy similar to what was used by Ruiz-Velasco et al. [1] for CTCF will be implemented. (3) In the final step of the project, the candidate will combine information

from (1) and (2) to overlay somatic mutation and DNA methylation at TF binding sites with the observed alternative splicing events in cancer patients. This project will consolidate the student's knowledge in computational biology for the analysis of genomics data with a focus on gene expression regulation and cancer. Moreover, the student will get familiar with large cancer genomics public data sets available at TCGA and/or ICGC. She/he will learn how to develop computational workflow to analyze large-scale data, such as differential exon usage and differential methylation analyses. The student will also be exposed to different programming languages such as R, Python, and Bash.

References:

1. Ruiz-Velasco M, Kumar M, Lai MC, Bhat P, Solis-Pinson AB, Reyes A, et al. CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Syst. 2017;5: 628–637.e6. 2. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479: 74–79. 3. Reyes A, Anders S, Huber W. Inferring differential exon usage in RNA-Seq data with the DEXSeq package. 2013. Available: https://bioconductor.riken.jp/packages/3.5/bioc/vignettes/DEXSeq/inst/doc/DEXSeq.pdf 4. Gheorghe M, Sandve GK, Khan A, Chèneby J, Ballester B, Mathelier A. A map of direct TF–DNA interactions in the human genome. Nucleic Acids Res. 2018;47: e21–e21.

Expected skills::

Proficiency in Python, R, and/or bash; Previous experience in genomics data analysis; Team spirit; English proficiency

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Renee Beekman in collaboration with Francois Serra and Alfonso Valencia
Email	renee.beekman@crg.eu; francois.serra@bsc.es; alfonso.valencia@bsc.es
Institution	Centre for Genomic Regulation (CRG) in collaboration with the Barcelona Supercomputing Center (BSC)
Website	https://www.crg.eu/en/programmes-groups/beekman-lab; https://www.bsc.es/discover-bsc/organisation/scientific-structure/computational-biology
Group	Single Cell Epigenomics and Cancer Development (CRG) in collaboration with Computational Biology Life Sciences (BSC)

Computational genomics

Project Title:

Unveiling differences between the 3D chromatin structure of homologous chromosomes in the context of translocations in tumour cells

Keywords:

3D Chromatin Structure, Chromosome/Allele-specific Read Mapping, De Novo Genome Assembly, Normal Cells and Tumour Cells, Lymphoma.

Summary:

Each mammalian cell has two copies of each chromosome. For the vast majority of computational analyses, the two copies of each chromosome are combined. However, it is generally known that many biological processes can show differences between the two copies of the chromosomes. A specific gene can for example be expressed from only one of the two chromosomes (allele-specific expression). Or, genetic mutations in tumour samples occur on only one of the two chromosomes (heterozygous mutations). In our group, we aim to distinguish information of the different copies of the chromosomes to better understand the development of cancer. We will do this in the context of lymphomas, which are tumours that originate from normal immune cells. While the chromosomes are considered linear structures, they are actually folded at the three-dimensional (3D) level in a highly organised way (Dekker et al. Nat Rev Genet. 2013). This organisation is needed for the chromosomes to regulate gene expression. Importantly, in lymphoma cells the 3D chromatin structure is altered in comparison to normal cells (Vilarrasa-Blasi & Soler-Vila et al. BioRxiv 2019). In our group, we aim to study the 3D chromatin structure in lymphoma cells in comparison to normal cells. More specifically, we study how genetic translocations (=a piece of one chromosome fuses to another chromosome) in lymphoma cells affect the 3D chromatin structure. In this project, we will use Hi-C data generated to study the 3D chromatin structure in normal cells (Rao et al. Cell 2014) and lymphoma cells (Vilarrasa-Blasi & Soler-Vila et al. BioRxiv 2019 and unpublished data). Hi-C is a molecular technique coupled to next generation sequencing that allows to reconstruct the 3D folding of the genome in the nucleus (Lieberman-Aiden et al. Science 2009). First, we will computationally separate the two homologous copies of chromosome 14. We focus on this chromosome as in lymphomas one of the copies of chromosome 14 is affected by a genetic translocation we aim to study. More specifically, we will use the variation in the genetic code between the two copies of chromosome 14 to distinguish them. To that end, we will use the genomic sequencing data of this chromosome in these samples and perform a de novo assembly to create a reference sequence for the two copies separately. Next, we will use this reference sequence to map the Hi-C reads representing the 3D chromatin structure to one or the other chromosome. Finally, we will reconstruct the 3D-chromatin structure using these separated sets of reads in TADbit (Serra et al. PLoS Comput Biol. 2017). From these reconstructed copies of chromosome 14 we will analyse the differences in the 3D chromatin structure in order to understand the effect of this genetic translocation specific to lymphoma on the 3D chromatin landscape surrounding it. What will you learn: • Computational biology: basics on network analysis; collaborative software development using GIT; to design and use of computational pipelines for high performance computing (in the 30th most powerful supercomputer in the world). • Structural Genomics: to process and analyse data from Chromosome Conformation Capture techniques (mostly Hi-C and Capture-C). • Genomics and Epigenomics: to explore available data at the interface of genomics and epigenomics; to understand the basics of gene regulation mechanisms and to postulate hypotheses about deregulation of these mechanisms in cancer and test them by analysing the data. • Tumour Biology: to understand the genetic and epigenetic mechanisms underlying the development of lymphomas. • Scientific Dissemination: to present in lab meetings and to write a research article resulting from your work.

References:

Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet. 2013 Jun;14(6):390-403. doi: 10.1038/nrg3454. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M. V, Ragoczy, T., Telling, A., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 2009 Jul; 326, 289–93. doi: 10.1126/science.1181369 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665-80. doi: 10.1016/j.cell.2014.11.021. Serra F, Baù D, Goodstadt M, Castillo D, Filion GJ, Marti-Renom MA. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. PLoS Comput Biol. 2017 Jul 19;13(7):e1005665. doi: 10.1371/journal.pcbi.1005665. Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, Russiñol N, Di Stefano M, Chapaprieta V, Clot G, Farabella I, Cuscó P, Agirre X, Prosper F, Beekman R, Beà S, Colomer D, Stunnenberg HG, Gut I, Campo E, Marti-Renom MA, Martin-Subero JI. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. bioRxiv 764910

Expected skills::

A strong background in UNIX command line tools as well as in python or R programming, in combination with creative thinking and enthusiasm to work in a multi-disciplinary team with wet lab and bioinformatic experience.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Our lab in the CRG can be divided into two branches: a wet lab branch and a bioinformatic branch. A key aspect of our group is that these two branches are intermingled, whereby the different team members can interact on a day-to-day basis and during weekly lab meetings. On top of that, most team members will have shared wet lab-bioinformatic projects. Moreover, this project in particular will be conducted in collaboration with the lab of Alfonso Valencia in the Barcelona Supercomputing Center, giving a significant boost in computational power and bioinformatics expertise. We strongly believe that both wet lab and bioinformatic analyses and especially the interaction between these two fields are critical to better understand biological phenomena.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Laura Isús
Email	laura.isus@genomcore.com
Institution	Made of Genes (Genomcore)
Website	https://genomcore.com, https://madeofgenes.com
Group	Bioinformatics Unit

Project

Computational genomics

Project Title:

Design and development of bioinformatic tools for precision medicine

Keywords:

Precision medicine, Computational Genomics, data integration, Report automatization

Summary:

Genomcore/Made of Genes (https://genomcore.com, https://madeofgenes.com) is a company founded in 2015 with the objective to allow the effective implementation of precision medicine in healthcare. We developed a unique B2B technological framework designed to manage large volumes of personal, health-related, highly sensitive biomedical and health data aimed to diagnosis laboratories and healthcare providers. We also feature a packetized B2C/B2B2C personalized healthcare solution that combines genomic and metabolic analysis in a single test. Our innovative solutions are recognized worldwide through different international awards, such as MIT Technology Review Innovators Under 35, Dubai Future Accelerators or Seal of Excellence of the European Commission. If you want to join a unique fast-growing, high-potential, trend-making company, this is your chance. We are looking for a talented and motivated Bioinformatics Student to collaborate with our bioinformatics team in the research, design and development of new bioinformatic tools for precision medicine. The project will entail processing, quality and annotation pipelines for omics datasets. Data visualization and report generation for prevention, diagnostic or treatment recommendations. The project will allow the student to participate and learn from a real setting and a selection of activities aimed to complete the researchers' career development. The position will be located in our Esplugues de Llobregat (Barcelona) offices.

Expected skills::

1) Experience working in Linux environments (Unix tools, Bash scripting, SSH, Unix filesystem…). 2) Experience in scripting language (Python is preferred). 3) Knowledge of general genetics and genetic inheritance. 4) Academic training in both Computer Sciences and Life Sciences (ie, Degree + Master in course). 5) Knowledge of tools for manipulating NGS data (BWA, Samtools, GATK, etc). 6) Experience using public databases (ClinVar, dbSNP, Reactome, OMIM, GO, PharmGKB, etc) will be valued. 7) Fluency in spoken and written English. 8) Fluency in Spanish or Catalan is a plus. Knowledge of other languages is also valued.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

Gross academic aid of 300€ / month

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Ichiro Hiratani
Email	ichiro.hiratani@riken.jp
Institution	RIKEN Center for Biosystems Dynamics Research (RIKEN BDR)
Website	https://www.bdr.riken.jp/en/research/labs/hiratani-i/index.html

Group Laboratory for Developmental Epigenetics

Project

Computational genomics

Project Title:

Integrative nucleome analysis of genome-wide scRepli-seq and Hi-C datasets to explore the 3D genome architecture

Keywords:

3D genome organization, 4D nucleome, Hi-C, scRepli-seq, NGS

Summary:

We welcome students with bioinformatics skills who have a keen interest in 3D genome architecture (4D nucleome) through integrative analysis of genome-wide NGS datasets derived from single-cell Repli-seq (scRepli-seq) and Hi-C experiments. We are looking for curator-type bioinformatics students who are responsible for all the work related to data resources and data integration, described as the "second category" bioinformatician in the following link. https://bitesizebio.com/38236/how-to-become-a-bioinformatician/

Expected skills::

(1) Programming skills for analyzing genomic data (unix/python/R/Perl), (2) Statistics background, (3) Basic molecular biology background

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Rosa Trobajo & David Mann
Email	rosa.trobajo@irta.cat
Institution	IRTA
Website	http://www.irta.cat/ca/grup/aigues-marines-i-continentals/
Group	Aigües Marines i Continentals

Project

Computational genomics

Project Title:

Functional evaluation of HTS reads using protein sequence data

Keywords:

HTS, protein sequence, environmental DNA, microalgae

Summary:

High Throughput Sequencing (HTS) is currently being developed as a substitute for traditional methods in biomonitoring aquatic ecosystems (e.g. for the European Union Water Framework Directive). For example, methods involving counting cells of microscopic algae are being replaced by metabarcoding: a short region of the gene (rbcL) coding for the large subunit of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase, the key enzyme of photosynthesis) is amplified from environmental samples and sequenced by Illumina; the reads are processed through a bioinformatics pipeline, where they are filtered to remove sequences containing errors, and then identified to species by reference to a database of known 'barcodes' from Sanger sequencing. Although these pipelines work adequately, some of the methods used to reject 'faulty' sequences are crude. For example, sequences that occur only rarely in a dataset are often rejected because of the error rate in Illumina sequencing, which leads to incorrect nucleotides occurring ±anywhere along a sequence. A threshold is therefore set, that a particular sequence must be observed twice or more times before it is accepted as real, on the basis that any particular error during sequencing is unlikely to occur many times repeatedly. The proportion of reads rejected on this basis can be of the order of 50% and probably involves many type II errors in the quantification of diversity. The suggested project would involve developing a method that is able to assess HTS reads, even when the sequences have never been encountered before (and are therefore not in the reference database), by taking account of the fact that RuBisCO (like all proteins) has a function and that function can only be performed if the protein folds in the correct way. Hence certain changes in the DNA coding for RuBisCO are evolutionarily 'easy' (because they have no effect on protein function, e.g. many codon 3rd position changes), whereas others are strongly or fully constrained. The project aims to determine the likelihood of different changes at a particular DNA site by evaluating variation among known rbcL gene sequences in the group of organisms being studied and also considering RuBisCO structure (which is well-known), and to use this information to develop an 'intelligent filter' for metabarcoding pipelines. In this, reads would be evaluated on the basis of whether they code for biologically plausible peptides, rather than solely on the basis of their frequency. Though applied to rbcL, the approach developed could be applicable with appropriate modification to any coding sequence used for metabarcoding (e.g. the CO1 gene used to barcode animals).

Expected skills::

Bioinformatics pipeline development, protein structure prediction, programming, sequence alignment

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Comments:

This project would require a combination of skills from different areas of specialization in the syllabus, mainly from computational genomics and structural bioinformatics, and include the need for some programming.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Marta Melé
Email	marta.mele@bsc.es
Institution	Barcelona Supercomputing Center
Website	https://www.bsc.es/discover-bsc/organisation/scientific-structure/transcriptomics-and-functional-genomics-lab-tfgl
Group	Trancritomics and functional Genomics

Project

Computational genomics

Project Title:

Understanding individual variation in splicing in human populations

Keywords:

Transcriptomics, differential gene expression, human populations, splicing, ribosome profiling, posttranscriptional processing, RNA binding proteins.

Summary:

The candidate will join Marta Melé's Transcriptomics and Functional Genomics lab in the Life Sciences Department at the Barcelona Supercomputing Center. The lab is interested in understanding how individual variation in gene expression can explain phenotypic differences between individuals both in the context of health and disease. To address this question, we use large-scale transcriptomic analysis and latest single-cell sequencing technologies combined with methods development to study gene expression, splicing and cell type composition variation across human tissues and phenotypes. In this project, we will perform a large-scale analysis of splicing variation between individuals with different phenotypes and from different ethnic groups. In previous studies, we observed that variation in splicing may play in contrast a comparatively greater role in defining individual phenotypes than variation in gene expression. contributes more to individual variation than to changes in gene expression (Melé et al. Science 2015). Moreover, we observed an enrichment of specific genes showing large splicing variation between individuals that was

especially strong for ribosomal proteins and that will be explored further. Ultimately, in this project we will explore in depth what is the role of splicing in defining why human individuals are different from one another. What you will learn: Development of computational pipelines to analyze and interpret large omics datasets such as RNA-Seq, single-cell RNA-seq, ribosome profiling, and CLiP-seq). Working in a high performance computing (HPC) environment. Effective communication of research findings, scientific writing, critical thinking.

References:

Melé, M. et al. The human transcriptome across tissues and individuals. Science (80-.). 348, 660–665 (2015).

Expected skills::

Availability to start in July 2020 is preferred Strong programming skills in bash, python, R, perl, or similar, Some experience working in HPC clusters Some experience with Next Generation Sequencing data analysis Excellent communication skills in spoken and written English Capacity to contribute to research projects with novel research ideas and analysis Capacity to work as a team in a highly collaborative and diverse environment

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Yasushi Okada
Email	y.okada@riken.jp
Institution	RIKEN, BDR
Website	https://www.bdr.riken.jp/en/research/labs/okada-y/index.html
Group	Laboratory for Cell Polarity Regulation

Project

Computational genomics

Project Title:

Decoding genome by imaging

Keywords:

protein engineering, image processing, machine learning, NGS data analysis

Summary:

We are developing technologies to estimate the epigenetic state of the cell from the super-resolution live cell imaging. The project includes the following four sub-projects, and the intern student can choose one according to his/her interest. 1) Development of the prove for visualization, which includes structure-based designing of the mutant protein probes, imaging of the designed probes, and analysis of the binding sites in the genome by genome-wide sequencing. 2) Development of the program for automation of the microscope system, which includes automatic search for the cell by deep learning. 3) Image processing, which includes denoising, regularization, and quantification, through the combination of the traditional algorithms and machine learning. 4) Development of the computational models to link the image data to the sequence data.

Expected skills::

no

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Manuel Irimia
Email	mirimia@gmail.com
Institution	Centre for Genomic Regulation

Website	https://www.crg.eu/manuel_irimia
Group	Transcriptomics of vertebrate development and evolution

Project

Computational genomics

Project Title:

The role of alternative splicing in the evolution of animal tissue diversity

Keywords:

Alternative splicing, microexons, exon-intron evolution, tissue-specific regulation, parallel evolution.

Summary:

Why evolution of alternative splicing? Alternative splicing (AS) is a molecular process allowing multiple transcripts to arise from the same gene. The power of AS in expanding the functional potential of a gene is well exemplified by the axon guidance receptor Dscam in Drosophila melanogaster. Dscam produces ~38000 distinct transcripts, which uniquely fine-tune the function of the gene in different neurons. AS contributes to functional diversity not only within cell populations but also between cell types/tissues. Our lab is currently investigating the role of AS in the evolution of animal tissue diversity: here we propose a parallel analysis of AS evolution in vertebrates and insects, two monophyletic clades in the bilaterian tree. Set up: we inferred exon orthology groups in a set of 20 bilaterian species (8 vertebrates, 8 insects and two pairs of relative outgroups) and we assembled a comprehensive RNA-seq dataset covering 8 homologous tissues in all species. We used the RNA-seq data to identify AS exons within each species and tissue. Experimental design: the project will be divided into three main parts: 1) We will investigate evolutionary patterns involving the entire tissue AS landscapes. Preliminary results show that neural and muscle AS networks seem to be well conserved in vertebrates but not in insects, suggesting different rewiring rates between the two clades. 2) We will focus on the exons specifically spliced within each tissue. Many tissue-specific exons have acquired tissue-specific regulation millions of years after their birth. An exciting perspective is the identification of a causal relationship between changes in exon regulation and simultaneous phenotypic innovation/adaptations. 3) We will explore the regulatory mechanisms underlying the rise of tissue-specific AS. The master project will be developed as part of this bigger project on exon evolution. The student will become familiar with the principles of alternative splicing and gene regulation, while getting hands-on experience with genome annotations, RNA-seq data analysis, comparative transcriptomics, and network reconstruction.

References:

- Torres-Méndez, A., Bonnal, S., Marquez, Y., Roth, J., Iglesias, M., Permanyer, J., Almudí, I., O'Hanlon, D., Guitart, T., Soller, M., Gingras, A.-C., Gebauer, F., Rentzsch, F., Blencowe, B.J.B., Valcárcel, J., Irimia, M. (2019). A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. Nature Ecol Evol, 3:691-701. - Marletaz, F., Firbas, P., Maeso, I., Tena, J.J., Bogdanovic, O., Perry, M., Wyatt, C.D.R., [+50 authors], Holland, P.W.H., Escriva, H., Gomez-Skarmeta, J.L., Irimia, M. (2018). Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature, 564:64-70. - 6) Burguera, D., Marquez, Y., Racioppi, C., Permanyer, J., Torres-Mendez, T., Esposito, R., Albuixech, B., Fanlo, L., D'Agostino, Y., Gohr, A., Navas-Perez, E., Riesgo, A., Cuomo, C., Benvenuto, G., Christiaen, L.A., Martí, E., D'Aniello, S., Spagnuolo, A., Ristoratore, F., Arnone, M.I., Garcia-Fernàndez, J., Irimia, M. (2017). Evolutionary recruitment of flexible Esrp-dependent splicing programs into diverse embryonic morphogenetic processes. Nat Commun, 8:1799.

Expected skills::

Ideally, experience on RNA-seq analyses and/or comparative genomics. Interest on genome evolution.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Mar Albà
Email	mar.alba@upf.edu
Institution	IMIM-UPF
Website	evolutionarygenomics.imim.es
Group	Evolutionary Genomics, Research Programme on Biomedical Informatics

Project

Computational genomics

Project Title:

Gene expression dysregulation in cancer and neoantigen formation

Keywords:

Transcriptomics; cancer; small ORFs; neoantigens; immunotherapy.

Summary:

In cancer, genomic structural rearrangements and mutations result in the expression of many novel transcripts that are not expressed in normal conditions. Recent studies suggest some of these transcripts translate peptides that can be presented by MHC molecules and be an important source of neoantigens. Such neoantigens are non-self proteins and could thus trigger a potent immune response and be very relevant for immunotherapy approaches to fight against cancer. However, the lack of studies measuring novel transcriptional events in cancer prevents us from fully understanding the contribution of these neoantigens. The aim of the project will be to perform transcriptome assembly directly from RNA-Seq data using large publicly available cancer cell datasets. In the group we have previously employed massive transcriptomics data to identify recently originated transcripts in human and mouse and predict any encoded protein products (Ruiz-Orera et al., 2015; Ruiz-Orera et al., 2018). Here we will use similar techniques to identify novel, non-annotated, transcripts in cancer cell RNA-Seq data and to characterize putative neoantigens.

References:

Ruiz-Orera, J., Hernández-Rodríguez, J., Chiva, C., Sabidó, E., Kondova, I., Bontrop, R., Marqués-Bonet, T., Albà, M.M. (2015). Origins of de novo genes in human and chimpanzee. Plos Genetics, 11(12): e1005721. Ruiz-Orera, J., Grau-Verdaguer, P.,

Villanueva-Cañas, J-L., Messeguer, X., Albà, M.M. (2018). Translation of neutrally evolving peptides provides a basis for de novo gene evolution. Nature Ecology and Evolution, 2:890–896.

Expected skills::

Interest in computational genomics and transcriptomics; knowledge of a programming language; knowledge of R; good command of English.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Miquel Angel Pujana
Email	mapujana@iconcologia.net
Institution	Catalan Institute of Oncology IDIBELL
Website	http://ico.gencat.cat/en/recerca/Programa-ProCURE/index.html
Group	Cancer Resistance & Bioinformatics

Project

Computational genomics

Project Title:

Discovery of “Dr Jekyll & Mr Hyde” genes

Keywords:

Cancer, genetics, outcome, tumor suppressor, oncogene

Summary:

Integration of genomic and clinical information using the Cox proportional hazard model is commonly used to identify biological factors that influence cancer outcome. Typically, this is applied to analyze the connection between gene expression and cancer progression, therapeutic response or patient survival. This approach has generated hundreds of biomarkers, of which several are nowadays applied in the clinic. However, some genes might not show a single facet during the course of the disease: they can act as tumor suppressors or as oncogenes depending on other variables (so called “Dr Jekyll and Mr Hyde” genes). These genes, their features and impact on cancer outcome remain completely unknown. Objectives In this project, we aim to identify this type of genes by pan-cancer interrogation of gene expression and clinical outcomes. This proposal will be integrated into experimental assays performed at the recipient group.

References:

An Integrated TCGA Pan-Cancer Clinical Data Resource to drive high quality survival outcome analytics. Cell. 173, 2: p400-416.e11, 10.1016/j.cell.2018.02.052 (2018). Kourou et al., Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 13, 8–17 (2015).

Expected skills::

Candidate(s) are expected to be proficient in programming in R and to have strong background on statistics.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Oriol Dols Icardo and Jordi Clarimón
Email	odols@santpau.cat and jclarimon@santpau.cat
Institution	Sant Pau Biomedical Research Institute
Website	http://santpaumemoryunit.com/

Group Genetics of Neurodegenerative Diseases Unit

Project

Computational genomics

Project Title:

Deep transcriptome characterization of the frontal cortex of frontotemporal lobar degeneration patients

Keywords:

Neurodegenerative disease; RNA sequencing; Transcriptome; Human brain; RNA alterations

Summary:

Frontotemporal lobar degeneration (FTLD) is a neuropathological term for a group of neurodegenerative dementias, mainly characterized by the aberrant deposition of TDP-43 (FTLD-TDP) or tau (FTLD-tau) proteins in the frontal and temporal lobes. Dysfunction of the RNA metabolism has proven to be one of the major pathological hallmarks of FTLD. In order to investigate RNA alterations in FTLD human brains, we have performed high-throughput RNA sequencing (encompassing total and small RNA) to deeply characterize the transcriptome of the frontal cortex of 12 FTLD-tau, 20 FTLD-TDP and 10 healthy controls. In this project, bioinformatics tools will be applied in order to disentangle gene and isoform differential expression, gene co-expression networks and alternative splicing events associated with FTLD which will be integrated with small RNA sequencing data from the same individuals. Finally, cell-type deconvolution algorithms using human single-nucleus RNA sequencing data will be applied to disentangle cellular heterogeneity in FTLD. Since this approach has not yet been performed in this neurodegenerative disorder, outcomes from this study will have a very high potential to be published in specialized journals.

Expected skills::

Linux/Ubuntu, R and python

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Mar Albà
Email	mar.alba@upf.edu
Institution	IMIM-UPF
Website	evolutionarygenomics.imim.es
Group	Evolutionary Genomics, Research Programme on Biomedical Informatics

Project

Computational genomics

Project Title:

Building transcriptomes with Nanopore sequencing data

Keywords:

Transcriptomics; transcript discovery; long reads; Nanopore; gene expression.

Summary:

Long read sequencing techniques, such as Oxford Nanopore Technologies (ONT), have great potential to sequence complex transcriptomes and to discover new transcripts beyond the annotated ones. While methods that work with Illumina short reads are quite mature, the development of software to work with Nanopore RNA sequencing reads is a very active area of research. We and co-workers have recently developed a method to build transcriptomes from RNA-derived Nanopore reads (cDNA and direct RNA) that does not require a reference genome and that could be used to investigate highly rearranged genomes (such as those in cancer cells) or species that currently lack a sequenced genome (de la Rubia et al., 2020). The aim of the project will be to compare methods based on Nanopore or Illumina reads for building eukaryotic transcriptomes in the absence of a reference genome and to identify novel, non-annotated, transcripts in species that already have a genome and reference annotations. We would like to determine when it is more convenient to use one sequencing technology over the other one, and if the combination of the two tecnologies – using Illumina reads to correct errors in Nanopore reads – is a real advantage. For this we will use already available datasets for yeast, human and mous especies, as well as datasets that are currently being generated in the group.

References:

de la Rubia, I., Indi, J.A., Carbonell, S., Lagarde, J., Albà, M.M., Eyras, E. (2020). Reference-free reconstruction and quantification of transcriptomes from long-read sequencing. bioRxiv, https://doi.org/10.1101/2020.02.08.939942

Expected skills::

Interest in computational genomics and transcriptomics; knowledge of a programming language; knowledge of R; good command of English.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	François Serra
Email	francois.serra@bsc.es
Institution	BSC - Barceona Supercomputing Center
Website	https://www.bsc.es/
Group	Computational Biology

Project

Computational genomics

Project Title:

A dynamic epigenetic network

Keywords:

epigenetics; networks; bioinformatics

Summary:

The laboratory of Alfonso Valencia specializes in many areas of computational biology focusing on the integration of data and development of computational frameworks that could help bridge the gap between fundamental research and personalized medicine. One of the areas of expertise of the lab is epigenomics, producing pioneer studies in the field. The project will be mainly supervised by François Serra an expert in epigenetics and chromatin conformation (coauthor of the works mentioned below on leukemia and cell differentiation), with experience in mentoring master and PhD students. The project we propose here is based on the work of Vera Pancaldi that built an epigenetic network based on the 3D conformation of the chromatin (Pancaldi et al. 2016). We aim to reproduce this work using more genomic interaction data (Davies et al. 2017) and to study it in the dynamic models of leukemia and cell differentiation (Beekman et al. 2018) or cell dedifferentiation (Stadhouders et al. 2018). Concretely the student will work on the development of a computational pipeline to discover interactions between DNA binding proteins and epigenetic marks. Finally, the data will be represented as a network of interactions to be analyzed at different time stages. We expect to be able to understand the functional association between the different actors in the epigenetic landscape and to understand the dynamics behind its remodeling upon disease or in development. What you will learn: - Computational biology: basics on network analysis; collaborative software development using GIT; design and use of computational pipelines for high performance computing (in the 30th most powerful supercomputer in the world). - Epigenomics: explore available data in the interface between genomics and epigenomics, postulate hypotheses about the mechanisms of gene regulation, and analyze the results. - Scientific Dissemination: to present in lab meetings and to write a research article resulting from your work.

References:

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Martens, J.H.A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M., Wingett, S.W., Clot, G., Queirós, A.C., Castellano, G., Blanc, J., Gut, M., Merkel, A., Heath, S., Vlasova, A., Ullrich, S. and Martin-Subero, J.I. 2018. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nature Medicine 24(6), pp. 868–880. Davies, J.O.J., Oudelaar, A.M., Higgs, D.R. and Hughes, J.R. 2017. How best to identify chromosomal interactions: a comparison of approaches. Nature Methods 14(2), pp. 125–134. Pancaldi, V., Carrillo-de-Santa-Pau, E., Javierre, B.M., Juan, D., Fraser, P., Spivakov, M., Valencia, A. and Rico, D. 2016. Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. Genome Biology 17(1), p. 152. Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., Hecht, J., Filion, G.J., Beato, M., Marti-Renom, M.A. and Graf, T. 2018. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nature Genetics 50(2), pp. 238–249.

Expected skills::

1- Critical thinking and creativity 2- Good statistical and programming skills (R/Bioconductor or Python) 3 - Basic knowledge of molecular biology 4- Ability to access and evaluate scientific literature

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

This project is a follow up of the work currently being done by a UPF master student. The new student will benefit from the results and analysis generated until now. It is also wide enough, biologically and methodologically, to leave room for the student to decide in which direction she/he would prefer to go in either way. The project will be supervised by François Serra and co-supervised by Alfonso Valencia.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Ramiro Logares
Email	ramiro.logares@icm.csic.es
Institution	ICM - CSIC
Website	http://www.log-lab.barcelona

Group Ecology of Marine Microbes & log-lab

Project

Computational genomics

Project Title:

Population dynamics and evolution of the ocean microbiome

Keywords:

microbiome, ocean, metagenomics, evolution, ecology

Summary:

The global ocean and the tiny organisms it contains are crucial for global ecosystem function. Microbial phytoplankton in the ocean fix as much carbon from the atmosphere as land plants, and other heterotrophic microbes guarantee that most of the fixed carbon is circulated through food webs. The genomic machinery that marine microbes use for performing a myriad of metabolic processes remained unknown until ca. 15 years ago, when large-scale DNA sequencing projects became feasible. With the advent of high-throughput DNA sequencing, we started unveiling the ocean microbiome at unprecedented levels of detail. During the last 5 years, very large genomic datasets have been extracted from the global ocean microbiome. In particular, the global expeditions TARA-Oceans (https://oceans.taraexpeditions.org) and Malaspina (http://www.expedicionmalaspina.es) have produced a goldmine of genomic data that we are continuously explored. This data is the best representation we have of the diversity and function of marine microbes, and considers mostly metagenomes and metatranscriptomes (ca. 30 Terabytes of compressed DNA data). My group at the ICM-CSIC (log-lab http://www.log-lab.barcelona at the EMM https://emm.icm.csic.es) is involved in both global marine expeditions. The proposed project aims at interrogating these datasets in order to 1) determine the population variation of selected microbes (using mutations; a.k.a. SNPs or Single Nucleotide Polymorphisms) in the global ocean and 2) find out whether some of the previous variation is due to evolutionary processes that occurred relatively recently in geological time. For investigating the above, we will build metagenome-assembled genomes (MAGs) and then map metagenomic or metatranscriptomic reads from the global ocean to a number of selected MAGs. Afterwards, we will perform a SNP calling analysis, aiming to determine fine-grained genomic variation. The analysis of these SNPs is what will indicate how much variation is present in the selected microbial populations and whether part of this variation has emerged through adaptive evolution. Most analyses for this work will be performed at our marine bioinformatics platform Marbits https://marbits.icm.csic.es

References:

Sunagawa, S., et al., Structure and function of the global ocean microbiome. Science, 2015. 348(6237): p. 1261359. Carradec, Q., et al., A global ocean atlas of eukaryotic genes. Nat Commun, 2018. 9(1): p. 373. Logares, R., et al., Disentangling the mechanisms shaping the surface ocean microbiota. 2020. Microbiome. In press. https://www.researchsquare.com/article/rs-7862/v2 Falkowski, P., The power of plankton. Nature, 2012. 483(7387): p. S17-20. Alberti, A., et al., Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. Sci Data, 2017. 4: p. 170093. de Vargas, C., et al., Eukaryotic plankton diversity in the sunlit ocean. Science, 2015. 348(6237): p. 1261605.

Expected skills::

Proficiency with bash and R. Familiar with python.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

motivation, interest to learn new bioinformatics techniques and to work in clusters

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Shigehiro Kuraku
Email	shigehiro.kuraku@riken.jp
Institution	RIKEN BDR
Website	https://www.bdr.riken.jp/en/research/labs/kuraku-s/
Group	Laboratory for Phyloinformatics

Project

Computational genomics

Project Title:

Elucidating the rules of genomic scaling: how does the size of DNA regions influence their physiological output?

Keywords:

genomic scaling, rate of living, c-value paradox, longevity

Summary:

Genome sizes exhibit a remarkable variation even among vertebrate animals. This project is assumed to be conducted only with computational solutions and aims at understanding the effect of variable physical spacing between exons and genes in animal genomes, by investigating which portion of the genomes are susceptible to the variation, with an emphasis on genes responsible for physiological controls.

References:

Hara et al. Nat Ecol Evol, 2018 2:1761- (https://www.nature.com/articles/s41559-018-0673-5) and Kowalczyk et al., eLife 2020 9:e51089 (https://elifesciences.org/articles/51089)

Expected skills::

Basic skills of programming, basic knowledge of molecular biology

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Tomas Marques-Bonet
Email	tomas.marques@upf.edu
Institution	UPF
Website	http://biologiaevolutiva.org/tmarques
Group	Comparative Genomics

Project

Computational genomics

Project Title:

Population genomics and conservation for whole genomes of 200 species of primates

Keywords:

Whole genomes DNA sequencing, Population genetics, variant calling, admixture

Summary:

Genomic diversity is at the core of many evolutionary inferences. The finer study of primates, our closest relatives, is relevant for

several reasons. They are the only living organisms with whom we share a higher proportion of genetic material as we have a shared evolutionary history over time. Thus, studying the genetics of the primates is a necessary endeavour to define the similarities among primates, the uniqueness of humans, and to strengthen the foundations of primate management and conservation. The latter of which should be an international effort, as these species should be considered a treasure of humanity. In the recent years, we have shown that it is possible to study full genome information from apes (Prado-Martinez et al. Nature 2013, Xue et al. Science 2015; deManuel et al. Science 2016; Nater et al. Current Biology 2017). Considering the population decline that all primates are experiencing, it is time-sensitive to act rapidly and generate the global dataset of variation for all primates. We have generated high quality full genome information for a large panel of primates all over the world. By using samples from the wild, we will further elucidate the role of demography, admixture and selection on genome diversity. In so doing, fundamental insights will be gained into the study of primates with multiple ramifications to biology.

References:

Prado-Martinez et al. Nature 2013, Xue et al. Science 2015; deManuel et al. Science 2016; Nater et al. Current Biology 2017

Expected skills::

Programming, population genetics.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Sarah Djebali
Email	sarah.djebali@inserm.fr
Institution	IRSD, INSERM U1220
Website	www.en.irsd.fr
Group	Genetic and regulation of iron metabolism

Project

Computational genomics

Project Title:

Bioinformatics methods for the identification of enhancer/gene relationships in vertebrate genomes

Keywords:

enhancer/gene regulatory relationships; high-through functional sequencing data; chromatin structure; program evaluation; machine learning

Summary:

For many complex genetic diseases, the majority of the identified variants are located outside protein-coding genes [1], making it difficult to understand their function. And when variants are located far away from any gene, they are usually assumed to act on the nearest gene, which can often prove totally wrong [2]. The regulatory element that can explain this long distance action of the variant on the gene is the enhancer. Enhancers are genomic regions on which transcription factors bind, and which activate the expression of one or several genes by being brought close to (in 3D) the uptream regulatory elements (promoters) of those genes. Enhancers can therefore be far away from the genes they activate on the 1D genome, but being close to them in the 3D space of the nucleus. Today the best approaches to identify enhancer/gene relationships in the genomes are genetic screening [3] and targeted chromatin structure (3D), such as polymerase II ChIA-PET [4] or promoter capture HiC [5]. The problem is that the first one can only targets a handful of genes and the second one is very difficult and costly to generate. For this reason and because many international consortia such as ENCODE, FANTOM or Epigenome Roadmap have recently produced and made publicly available large quantities of functional 1D data (such as RNA-seq, ATAC-seq, histone marks or methylation data), the favoured approach is the integration of high-throuput functional 1D data. Although many programs exist to identify enhancer/gene relationships from functional 1D data [6,7,8], there is no consensus about what the best approach is. Here we would like to fill in this gap by assessing the different existing methods on reference sets and proposing a new method that uses a minimal amount of different data. The student will therefore have to: - Make a complete state-of-the-art of the existing 1D methods - Plan the evaluation * Define reference sets * Define criteria to include programs in the evaluation * Define the input data to use for each program to evaluate - Make the programs to evaluate work on small and real evaluation datasets - Determine the best approach and propose a new one that uses as few different input data as possible

References:

[1] Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. 2009 Jun 9;106(23):9362-7. [2] Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, Nguyen ML, Rubin AJ, Granja JM, Kazane KR, Wei Y. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nature genetics. 2017 Nov;49(11):1602. [3] Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, Nguyen TH. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nature Genetics. 2019 Dec;51(12):1664-9. [4] Zhang J, Poh HM, Peh SQ, Sia YY, Li G, Mulawadi FH, Goh Y, Fullwood MJ, Sung WK, Ruan X, Ruan Y. ChIA-PET analysis of transcriptional chromatin interactions. Methods. 2012 Nov 1;58(3):289-99. [5] Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nature genetics. 2015 Jun;47(6):598. [6] He B, Chen C, Teng L, Tan K. Global view of enhancer–promoter interactome in human cells. Proceedings of the National Academy of Sciences. 2014 May 27;111(21):E2191-9. [7] Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, Mok MT, Cheng C, Fan X, Gerstein M, Cheng AS. Reconstruction of enhancer–target networks in 935 samples of human primary cells, tissues and cell lines. Nature genetics. 2017 Oct;49(10):1428. [8] Li W, Wong WH, Jiang R. DeepTACT: predicting 3D chromatin contacts via bootstrapping deep learning. Nucleic acids research. 2019 Jun 4;47(10):e60-.

Expected skills::

Linux command line; Programming skills (bash, awk, python, ...); having already manipulated high-throughput (functional) sequencing data; know the basics of statistics and the R language; know how to run jobs on a cluster; understand written English well

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

PhD funding is not available yet but several options can be envisionned. This question also depends on the success of applications for funds that will be done during the fall.

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Nuria Lopez-Bigas
Email	nuria.lopez@irbbarcelona.org
Institution	IRB Barcelona
Website	bbglab.irbbarcelona.org
Group	Biomedical Genomics

Project

Computational genomics

Project Title:

Understanding cancer biology

Keywords:

Cancer drivers, selective advantage, mutational processes, tumorigenesis

Summary:

A tumor has between hundreds and thousands of mutations and only a few are directly involved in tumorigenesis, frequently called driver mutations. These mutations affect genes which when mutated confer the cell with a growth advantage with respect to its neighbors. Our lab has developed methods to identify these driver genes, and has analyzed tens of thousands of tumors, producing a catalog of the genes underlying tumorigenesis in the most frequent cancer types. Currently, we are interested in cataloguing the downstream effect that mutations affecting these driver genes have in different tumor types. While many mutations in driver genes are capable of driving tumorigenesis, some are not, and the range of driver mutations of a cancer gene varies between tumor types. Understanding the functional effect of driver mutations thus constitutes a key goal of cancer genomics research.

References:

Tamborero et al, 2018. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Medicine. 10:25 Pich et al, 2018. Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove

around Nucleosomes. Cell doi:10.1016/j.cell.2018.10.004 Sabarinathan et al., 2016. Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature 532, 264-267 Mularoni et al, 2016. OncodriveFML: A general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biology. 17: 128 Rubio-Perez et al, 2015. In silico prescription of anti-cancer drugs to cohorts of 28 tumor types reveals novel targeting opportunities. Cancer Cell. 27(3):382-396 Gonzalez-Perez et al, 2013. IntOGen-mutations identifies cancer drivers across tumor types. Nature Methods. doi:10.1038/nmeth.2642

Expected skills::

Basic programming, data analysis and statistics skills. Willing to learn

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Chaysavanh Manichanh
Email	cmanicha@gmail.com
Institution	Vall d'Hebron Research Institute
Website	https://sites.google.com/site/manichanhlab/ and http://www.vhir.org/portal1/
Group	Microbiome Lab

Project

Computational genomics

Project Title:

Development of bioinformatics and statistical tools to integrate meta-omics data to decipher the human microbiome

Keywords:

Human Microbiome; Metagenomics; Metatranscriptomics; Metabolomics; Composition and functions

Summary:

Meta-omics approaches have been intensively used over the last 20 years to study the composition and functions of the human microbiome (the other Human Genome) in health and disease conditions. The aim of the present work is to develop and/or implement bioinformatics tools to analyze and integrate meta-omics data. • You will work in the dry-lab conducting bioinformatics and biostatistical research. You will be integrated in a young and collaborative environment: medical doctors, nutritionist, molecular biologists, bioinformaticians, statistician. • You will learn from your colleagues, and take responsibility, in writing your conclusions into academic papers, which eventually will be published in High Impact Journals. We want to help you build solid foundations on the research method, so you will be assisted by more experienced colleagues.

References:

https://sites.google.com/site/manichanhlab/our-publications

Expected skills::

Fluent in English (most of our team are foreigners, thus English is our language); Theoretical and practical knowledge of classical statistical inference and Machine Learning; Strong coding experience

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

We are looking for a motivated student who is seeking to pursue his/her career in research. The candidate will be remunerated 1000 euros/month (gross salary) during his master internship and will be offered the possibility to apply for a PhD fellowship (INPhINIT “la Caixa”, FPU, AGAUR, VHIR...).

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor Rory Johnson

Email	rory.johnson@ucd.ie
Institution	University College Dublin
Website	https://www.gold-lab.org/
Group	Laboratory for Genomics of Long Noncoding RNAs and Disease (GOLD Lab)

Project

Computational genomics

Project Title:

CRISPR, Cancer, Noncoding RNAs

Keywords:

CRISPR, cancer, lncRNA

Summary:

One of the biggest biological surprises of the last decade has been the discovery of a completely new class of genes in the human genome – long non-coding RNAs (lncRNAs). These RNA transcripts are not translated into protein, but instead seem to function as regulatory molecules that control the expression of other genes. As part of the international ENCODE consortium, our group has helped catalogue >10.000 of these genes, 99% of which remain completely uncharacterised. LncRNAs represent an extremely promising source of new drug targets. The objective of our lab is to develop a new generation of anti-cancer therapies based on designed lncRNA inhibitors. We identify lncRNA targets via in-house developed, interdisciplinary strategies combining bioinformatics with CRISPR-Cas9 genome-engineering tools. We can offer a variety of tailor-made projects to interested students. These may be, for example, integrative data analysis to make testable predictions about lncRNA functionality, or creation of pipelines for identification of cancer-causing lncRNAs from our CRISPR screens. Previous MSc students have gone on to publish first-author papers based on their MSc thesis, and have successful scientific careers with us and other groups (eg from UPF: Carlevaro-Fita, Pulido-Quetglas, Mas-Ponte, Lanzos – see Pubmed). Students in our lab get exposed to latest bioinformatic and experimental practices on a daily basis. They get closely mentored and have numerous opportunities to present their work internally. Our lab is involved in several international collaborations and consortia, including the International Cancer Genome Consortium (https://www.nature.com/collections/afdejfafdb) and we were recently awarded a prestigious Future Research Leaders grant from the President of Ireland (https://www.sfi.ie/research-news/news/president-higgins-honours/). If you are motivated to work at the forefront in computational cancer genomics and genome-engineering in a fun, supportive and motivated team, then contact us for more information! See also: gold-lab.org https://twitter.com/GOLDLab_Bern https://people.ucd.ie/rory.johnson

References:

Selected recent papers: Rheinbay E... PCAWG Consortium (including Johnson R, Carlevaro-Fita J, Lanzos A) Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020 Feb;578(7793):102-111. Bergadà-Pijuan J, Pulido-Quetglas C, Vancura A, Johnson R#. CASPR, an analysis pipeline for single and paired guide RNA CRISPR screens, reveals optimal target selection for long noncoding RNAs. Bioinformatics. 2019 (In Press) Carlevaro-Fita J, Polidori T, Das M, Navarro C, Zoller TI, Johnson R#. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. Genome Research 2019 Feb;29(2):208-222 Joana Carlevaro-Fita, Rory Johnson#. Global Positioning System: Understanding long noncoding RNAs through subcellular localisation. Molecular Cell 2019 Mar 7;73(5):869-883 Roberta Esposito, Núria Bosch, Andrés Lanzós, Taisia Polidori, Carlos Pulido-Quetglas, Rory Johnson#. Hacking the cancer genome: Profiling therapeutically-actionable long noncoding RNAs using CRISPR-Cas9 screening. Cancer Cell 2019 Apr 15;35(4):545-557. Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R#, Johnson R#. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nature Genetics 2017 Dec;49(12):1731-1740 Uszczynska-Ratajczak B, Lagarde J, Frankish A, Guigó R, Johnson R#. Towards a complete map of the human long non-coding RNA transcriptome. Nature Reviews Genetics 2018 Sep;19(9):535-548.

Expected skills::

Unix, R, python

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Miquel Angel Pujana
Email	mapujana@iconcologia.net
Institution	Catalan Institute of Oncology IDIBELL
Website	http://ico.gencat.cat/en/recerca/Programa-ProCURE/index.html
Group	Cancer Resistance Research & Bioinformatics

Project

Computational genomics

Project Title:

Discovering unexpected cancer-protective effects of common medications

Keywords:

Cancer, therapy, common medication, genetics, GWAS, epidemiology

Summary:

Development of a new cancer-target drug costs hundreds of millions of EUR and on average 10 years of experimental work before approval. However, overall success rate is less than 10%. Thus, drug repurposing is received much attention and new indications of existing drugs are accounting for 20% of new products. Systematic analyses of thousands of developed/approved drug or compounds have found many with previously unrecognized anti-cancer activity. While these evidence mainly derive from in vitro cellular assays, large-scale studies of population-based health care records integrated into genetic information are currently possible. Preliminary data from our group has discovered that certain drugs used for non-cancer common conditions have large protective effects regarding cancer progression and metastasis. In this project, we aim to estimate the beneficial effects of common medications on cancer patient survival by integrating and modeling epidemiological and health care data from two European

populations. The effects will be further deciphered at the germline genetic level by meta-analyses of GWASs. This proposal is integrated into experimental assays also performed at the recipient group.

References:

• Bycroft et al., The UK Biobank resource with deep phenotyping and genomic data, Nature 562, 203–209(2018). • Bolivar et al., SIDIAP Database: Electronic Clinical Records in Primary Care as a Source of Information for Epidemiologic Research, Med Clin. 138(14):617-21 (2012). • Pantziarka et al., Hard Drug Repurposing for Precision Oncology: The Missing Link? Front Pharmacol. 9: 637 (2018). • Corsello et al., Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer doi:10.1038/s43018-019-0018-6 (2020).

Expected skills::

Candidate(s) are expected to be proficient in programming in R and to have strong background on statistics.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Lluís Ribas de Pouplana
Email	lluis.ribas@irbbarcelona.org
Institution	IRB Barcelona
Website	www.irbbarcelona.org
Group	Gene Translation Laboratory

Project

Computational genomics

Project Title:

Identification of domain-specific adaptations for protein synthesis

Keywords:

Protein synthesis, proteome diversity, transfer RNA, modified bases, evolution

Summary:

We have developed tools to detect and analyze protein sequences that are impossible to synthesize for some organisms. Those species that can make this proteins do so thanks to special adaptations ('upgrades') of the protein synthesis machinery. Now we want to further develop and use these tools to map the global proteome landscape and identify all possible protein sequences possibly unique to some group of species thanks to the existance of 'upgrades'. We offer one or two payed (modestly) positions to carry out these analyses. We look for candidates interested in evolution, biochemistry, and confident in the use of R.

References:

References: 1. The mitochondrial tRNA conundrum. (2020) Ribas de Pouplana L. Nat Rev Mol Cell Biol. doi: 10.1038/s41580-020-0220-5. 2. Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. (2019) Torres AG, Reina O, Stephan-Otto Attolini C, Ribas de Pouplana L. Proc Natl Acad Sci U S A. 116(17):8451-8456. 3. The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes. (2019) Rafels-Ybern À, Torres AG, Camacho N, Herencia-Ropero A, Roura Frigolé H, Wulff TF, Raboteg M, Bordons A, Grau-Bove X, Ruiz-Trillo I, Ribas de Pouplana L. Mol Biol Evol. 36(4):650-662. 4. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. (2018) Rafels-Ybern À, Torres AG, Grau-Bove X, Ruiz-Trillo I, Ribas de Pouplana L. RNA Biol. 2018;15(4-5):500-507.

Expected skills::

R. Desirable but not essential: Python, and experience in phylogenetic analysis.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Mario Cáceres
Email	mcaceres@icrea.cat
Institution	Institut de Biotecnologia i de Biomedicina (IBB), Unversitat Autònoma de Barcelona (UAB)
Website	https://invfest.uab.cat/
Group	Comparative and Functional Genomics Group

Project

Computational genomics

Project Title:

Functional and evolutionary impact of polymorphic inversions in the human genome

Keywords:

Expected skills depend on the actual line of research chosen, but should include scripting/programming skills (python, bash, R and/or perl) and experience in genomic variants and functional analysis. Knowledge of MySQL and PHP would also be helpful for working with the InvFEST database.se.

Summary:

The master student will integrate in a young, interdisciplinary and highly-dynamic group. In particular, the proposed tasks span a diverse range of themes focused in the functional and evolutionary impact of inversions, which are a little studied class of genomic variants, and the project could vary according to the interest and background of the candidate. 1. Bioinformatic analysis of the functional consequences of inversions and their association with phenotypic traits and disease susceptibility through imputation of inversion genotypes in large-scale datasets, in which the effect of these changes has been typically missed. 2. Development of new functionalities and visualization tools for our human polymorphic inversion data base InvFEST (http://invfestdb.uab.cat/), the world reference of human inversions. 3. Comparative study of known human inversion regions in other mammal species genomes to determine if there are inversion recurrence hotspots conserved over long evolutionary distances that might indicate a potential functional role.

References:

M. Puig et al. Determining the impact of uncharacterized inversions in the human genome by droplet digital PCR. Genome Research (in press) (2020). C. Giner-Delgado et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nature Communications 10: 4222 (2019). D. Vicente-Salvador et al. Detailed analysis of inversions predicted between two human genomes: errors, real polymorphisms, and their origin and population distribution. Human Molecular Genetics 26:567-581 (2017). M. Puig et al. Functional impact and evolution of a novel human polymorphic inversion that disrupts a gene and creates a fusion transcript. PLoS Genetics 11(10): e1005495. doi:10.1371/journal.pgen.1005495 (2015). A. Martínez-Fundichely et al. InvFEST, a database integrating information of polymorphic inversions in the human genome. Nucleic Acids Research 42 (D1): D1027-D1032 (2014).

Expected skills::

Expected skills depend on the actual line of research chosen, but should include perl, python and bash programming and experience in working with DNA sequence data and functional analysis. Knowledge of MySQL and PHP would also be helpful for working with the InvFEST database.

Possibility of funding::

Yes

Possible continuity with PhD: :

Yes

Comments:

Depending on the degree of experience of the candidate and the task performed it is possible to obtain financial support for the master practice, Also, at the end of the master there is the possibility to apply for a PhD fellowship.

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Albert Jordan
Email	ajvbmc@ibmb.csic.es
Institution	Institute of Molecular Biology of Barcelona (IBMB-CSIC)
Website	http://www.ibmb.csic.es/groups/chromatin-regulation-of-human-and-viral-gene-expression
Group	Chromatin regulation of human and viral gene expression

Project

Computational genomics

Project Title:

Occupancy of histone H1 variants genome-wide and consequences of altering H1 levels on human chromatin organization.

Keywords:

Chromatin, histones, genomics, 3D nuclear structure, ChIP-seq

Summary:

We focus our research on the control of gene expression in human cells by chromatin organization, components and modifications. We investigate the role and specificity of histone H1 variants in chromatin organization and gene expression control. By RNA interference of the different human H1 variants we have found that they have different involvement in cellular processes such as cell cycle progression and gene expression. We have also described a differential role of H1 variants in pluripotency and differentiation. Currently, we are investigating the occupancy of H1 variants genome-wide by ChIP-seq (NGS) and the consequences of altering H1 levels on chromatin organization (ATAC-seq, DNA methylation, hiC, etc), with an extensive use of Genomics and

Bioinformatics. Additionally, we are performing proteomics of H1 variant specific protein complexes in chromatin and nucleoplasm.

References:

▲ Izquierdo-Bouldstridge A*, Bustillos A*, Bonet-Costa C, Aribau P, Garcia D, Dabad M, Esteve-Codina A, Pascual L, Peiro S, Esteller M, Murtha M, Millán-Ariño Ll, Jordan A (2017) Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. Nucleic Acids Research 45(20): 11622-42. ▲ Millán-Ariño Ll, Izquierdo-Bouldstridge A, Jordan A (2016) Specificities and genomic distribution of somatic mammalian histone H1 subtypes. BBA Gene Regulatory Mechanisms 1859(3): 510-19. ▲ Mayor R*, Izquierdo-Bouldstridge A*, Millán-Ariño Ll, Bustillos A, Sampaio C, Luque N, Jordan A (2015) Genome distribution of replication-independent histone H1 variants shows H1.0 associated with nucleolar domains and H1X associated with RNA polymerase II-enriched regions. Journal of Biological Chemistry 290(12):7474-91. ▲ Millán-Ariño Ll, Islam A, Izquierdo-Bouldstridge A, Mayor R, Terme JM, Luque N, Sancho M, López-Bigas N, Jordan A (2014) Mapping of six somatic linker histone H1 variants in human breast cancer cells uncovers specific features of H1.2. Nucleic Acids Research. doi: 10.1093/nar/gku079 ▲ Terme JM*, Sesé B*, Millán-Ariño L, Mayor R, Izpisua-Belmonte JC, Barrero MJ, Jordan A (2011) Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. Journal of Biological Chemistry 286(41):35347-57 ▲ Sancho M, Diani E, Beato M, Jordan A (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLOS Genetics- Oct;4(10):e1000227.

Expected skills::

Strong motivation for research. Background or interest in Biologgy/Biomedicine and Epigenetics. The student will work in analyzing high-throughput genomic data such as ChIP-seq, RNA-seq, ATAC-seq and hi-C. To do so, experience in handling aligners, peak calling softwares, differential gene expression analysis and statistics tests will be an advantage. In addition, programming skills in R, Python and/or Perl are also necessary.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Robert Castelo
Email	robert.castelo@upf.edu
Institution	Universitat Pompeu Fabra
Website	https://functionalgenomics.upf.edu

Group Functional Genomics

Project

Computational genomics

Project Title:

Functional genomics

Keywords:

genetics, genomics, statistics, bioconductor

Summary:

The research in the functional genomics group is geared towards the development of computational methods and pipelines to address questions of biological and clinical relevance. Depending on the profile of the candidate, different types of master projects are possible, ranging from software engineering and development in R/Bioconductor, development of new methods for the analysis of high-throughput genomics data, to the analysis of specific datasets to answer particular biological and clinical questions. Some of our contributions in all these aspects can be found in the list of references.

References:

1. Costa et al. Genome-wide postnatal changes in immunity following fetal inflammatory response. medRxiv, 19000109, 2020. 2. Roverato and Castelo. Path weights in concentration graphs. Biometrika, in press (arXiv:1907.05781) 3. Puigdevall et al. Genetic linkage analysis of a large family identifies FIGN as a candidate modulator of reduced penetrance in heritable pulmonary arterial hypertension. Journal of Medical Genetics, 56:481-490, 2019. 4. Puigdevall and Castelo. GenomicScores: seamless access to genomewide position-specific scores from R and Bioconductor. Bioinformatics, 18:3208-3210, 2018. 5. Roverato and Castelo. The networked partial correlation and its application to the analysis of genetic interactions. Journal of the Royal Statistical Society Series C -Applied Statistics, 66:647-665, 2017. 6. Costa and Castelo. Umbilical cord gene expression reveals the molecular architecture of the fetal inflammatory response in extremely preterm newborns. Pediatric Research, 79:473-481, 2016. 7. Baumstark et al. The propagation of perturbations in rewired bacterial gene networks. Nature Communications, 6:10105, 2015. 8. Tur et al. Mapping eQTL networks with mixed graphical Markov models. Genetics, 198(4):1377-1383, 2014. 9. Hänzelmann et al. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics, 14:7, 2013.

Expected skills::

Programming, scripting, minimum understanding of statistics.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Josep F Abril
Email	jabril@ub.edu
Institution	Department of Genetics, Microbiology & Statistics
Website	https://compgen.bio.ub.edu/
Group	Computational Genomics Lab @ UB

Project

Computational genomics

Project Title:

Refactoring Viral Metagenomic Pipelines

Keywords:

metagenomics, sequence analysis, kmer frequencies, taxonomy annotation pipelines

Summary:

In collaboration with the VirCont research lab, we have already developed a number of analysis procedures for the characterization of viral species found from high-throughput sequencing experiments of complex samples, as well as the diversity parameters from environmental samples. We want to integrate those into a semi-automatic/fully-automatic pipeline to perform the whole process, from data-gathering to the generation of summary reports. We will try to extend the current analyses with k-mer based approaches, as well as more efficient ways to assign species by fast homology-based approaches.

References:

Natalia Timoneda PhD Thesis: https://compgen.bio.ub.edu/dl2650

Expected skills::

Student should master Unix/bash, python/perl/C, R, SQL, web apps (HTML/CSS/shiny/django)..

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Comments:

If applicant is interested in doing a PhD, there is the possibility to apply for a Generalitat FI or Ministerio FPU PhD grants on the next announcement.

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Arnau Sebe-Pedros
Email	arnau.sebe@crg.es
Institution	CRG
Website	https://www.crg.eu/en/programmes-groups/sebe-pedros-lab
Group	Single-cell genomics and evolution

Project

Computational genomics

Project Title:

Evolutionary modeling of cell type gene regulatory networks using single cell genomics data

Keywords:

evolution; gene regulation; cell types; transcription factors; chromatin accessibility

Summary:

Our group studies genome regulation from an evolutionary systems perspective. In particular, we are interested in deciphering the evolutionary dynamics of animal cell type programs and in reconstructing the emergence of genome regulatory mechanisms linked to cell type differentiation (from transcription factor binding through chromatin states to the physical architecture of the genome). To this end, we apply advanced single-cell genomics and chromatin experimental methods to molecularly dissect cell types and epigenomic landscapes in phylogenetically diverse organisms. We also develop computational tools to integrate these diverse data sources into models of cell type gene regulatory networks and we use phylogenetic methods to comparatively analyse these models.

Our recent work has provided the first whole-organism cell type atlases in different species and mapped key regulatory genome features underlying these cellular programs. By sampling additional species and chromatin features at single-cell resolution, we now aim at dissecting the evolution of cell types and their underlying gene regulatory networks. We are seeking highly motivated master students to join our team and work on inferring and comparing cell type gene regulatory networks (GRNs) across species. Methodologically, this project will involve the integrative computational analysis high-throughput single-cell genomics and chromatin data in different systems.

References:

https://www.ncbi.nlm.nih.gov/pubmed/29856957 https://www.ncbi.nlm.nih.gov/pubmed/29942020
https://www.ncbi.nlm.nih.gov/pubmed/27114036

Expected skills::

R and Python programming; experience working on a computing cluster; good understanding of functional genomics methods and experience analyzing genomic data.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Biola M. Javierre
Email	bmjavierre@carrerasresearch.org
Institution	Josep Carreras Leukaemia Research Institute (IJC)
Website	http:www.carrerasresearch.org/
Group	3D Chromatin Organization

Project

Computational genomics

Project Title:

Dissecting The Role Of Spatial-Temporal Genome Architecture In Pediatric Acute Lymphoblastic Leukemia

Keywords:

leukemia, ¨omics¨ data, 3D genome architecture

Summary:

Most of mutations and epimutations associated with complex diseases lie in non-coding regions, frequently at regulatory regions, and potentially exert their functions by altering the regulation of the target genes. The vast majority of regulatory elements that regulate each gene in each cell type are uncharted, constituting a major missing link in understanding genome control. We previously developed a new method called Promoter Capture Hi-C (PCHi-C), which allows the pioneer genome-wide systematic identification of the long-range regulatory elements that control more than 20000 genes. Using this method, we connected for the first-time non-coding autoimmune disease variants to putative target promoters prioritizing thousands of disease-candidate genes and implicating disease pathways, quarters of which not previously implicated (Cell 2016). Based on preliminary data recently generated, we hypothese that the novel description of the regulatory elements that control each gene along human B lymphpoiesis could allow to understand the contributions of mutations and epimutations in B cell cancer development and to discover new genes potentially implicated in malignant transformation. First, we are developing a novel experimental and computational methodology to genome-wide detect distal interacting regions of the genome for all genes in rare cell types with an improved resolution. Second, using this new methodology and other omics such as CHiP-seq, RAN-seq and ATAC-seq, we will unravel the dynamic rewiring of promoter interactomes along B cell differentiation. Third, we will link non-coding mutations and epimutations to their putative target genes, describing potential novel genes and gene pathways associated with B cell pediatric acute lymphoblastic leukemia. In summary, this interdisciplinary project will provide unprecedented insights into our understanding of how cells decide their identity with an impact on regenerative medicine, autoimmunity, immunodeficiency and B cell malignancies. SPECIFIC AIMS. - Mapping, filtering, interaction peak calling and analysis of the new method inspired on PCHiC data (HICUP and CHiCAGO pipelines). - Mapping, filtering, calling and analysis of CHIP-seq, ATAC-seq and RNA-seq - Analysis of GWAS data, WGS and WBGS data. - Integration of non-coding mutations and epimutations (Differentially Methylated Regions) with the previously omics data to define new genes and gene pathways associated with pediatric acute lymphoblastic leukemia.

References:

-Javierre, B.M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167, 1369-1384.e19 (2016). This study focused on 17 human primary hematopoietic cell types, demonstrates that promoter interactions are highly cell-type- and lineage-specific and that they allow the association of non-coding mutations with potential target genes. - Cairns, J. Freire-Pritchett, P., Wingett, S.W., Várnai, C., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.M. et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 17, 127 (2016). This paper describes the algorithms for detecting significant interactions from capture Hi-C. - Pancaldi, V., Carrillo-de-Santa-Pau, E., Javierre, B.M. et al. Integrating epigenomic data and 3D genomic structure with a new measure of chromatin assortativity. Genome Biol. 17, 152 (2016) This manuscript summarizes the computational method used to calculate the enrichment of specific epigenomic features in the chromatin fragments constituting the nodes of the network. - Azagra A, Marina-Zarate E, Ramiro AR, Javierre BM #, Parra M # (#Corresponding author). From Loops to Looks: Transcription Factors and Chromatin Organization Shaping Terminal B Cell Differentiation. Trends Immunol (2020) This review summarizes the role of genome architecture in B cell differentiation and biology - Watt S, Vasquez L, Walter K, Mann AL, Kundu K, Chen L, Yan Y, Ecker S, Burden F, Farrow S, Farr B, Iotchkova V, Elding H, Mead D, Tardaguila M, Ponstingl H, Richardson D, Datta A, Flicek P, Clarke L, Downes K, Pastinen T, Fraser, P, Frontini M, Javierre BM #, Spivakov M#, Soranzo N# (#Corresponding author). Variation in PU.1 binding and chromatin looping at neutrophil enhancers influences autoimmune disease susceptibility. Nat Commun. (Under review) bioRxiv 620260; doi: https://doi.org/10.1101/620260 This manuscript describes the interplay between SNPs, transcription factor binding, gene expression, histone modifications, 3D chromatin organization and disease.

Expected skills::

High level of motivation and interest, Proficiency in at least one scripting or programming language, Proficiency in scripting environments for statistics and data analysis, Competitive CV, High level of collaborative and communicative skills, Good level of English speaking and writing skills.

Possibility of funding::

No

Possible continuity with PhD: :

Yes

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Andreu Paytuvi
Email	apaytuvi@sequentiabiotech.com
Institution	Sequentia Biotech
Website	http://www.sequentiabiotech.com
Group	Sequentia Biotech

Project

Computational genomics

Project Title:

Analysing human microbiomes: towards personalized medicine

Keywords:

microbiome nutrition health-care

Summary:

The study of the microbiomes present in the human body is of fundamental importance as it is highly relevant for clinical applications. For instance, dysbiosis in distinct communities has been related to some diseases. Traditionally, studying these populations required the isolation and culture of each individual microorganism, which is a significant limitation considering that small portion prokaryotes are culturable. However, using sequencing technologies allows the study of these populations in a high-throughput manner. These technologies have been essential for the development of metagenomics, which is defined as the culture-independent genomic analysis of all the microorganisms in an environmental niche. Human microbiomes are taxonomically different whether they come from the gut, skin, vagina or from the mouth. For instance, the genus Bacteroides is very abundant in the gut while it is Lactobacillus in the vagina. Changes in the normal microbiota composition (dysbiosis) have been linked to some diseases such as diabetes (gut), obesity (gut), autism (gut), fertility (vagina), acne (skin), Parkinson (gut), among others. About 16,000 samples from the American Gut Project (AGP) have been already analysed with Gaia to obtain the taxonomic profile of the samples. Metadata for these 16,000 samples is available. With this taxonomic matrix and the available metadata, the student will develop methods, especially related to machine-learning, that will help doctors to diagnose. Therefore, the final aim of the project is the development of models to classify a sample (e.g. from a patient) to specific groups (e.g. potentially diabetic, Parkinson-like profile, etc.).

Expected skills::

The student considering this thesis proposal must have a strong Bash (command-line) and Python/R knowledge.

Possibility of funding::

No

Possible continuity with PhD: :

No

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Josep Vilardell
Email	josep.vilardell@ibmb.csic.es
Institution	institute of Molecular Biology of Barcelona
Website	www.ibmb.csic.es/vilardell
Group	Mechanisms of pre-mRNA splicing

Project

Computational genomics

Project Title:

Impact of the spliceosome on protein synthesis

Keywords:

splicing, spliceosome, ribosome, mRNA. differential expression

Summary:

The information content of genomes can be greatly expanded by pre-mRNA splicing. Virtually all human pre-mRNAs need to be spliced to become mRNAs. Furthermore, most pre-mRNAs can be spliced into different mRNAs by alternative splicing. Therefore, it is hardly surprising that perturbations in splicing are linked to disease. However, we know little on how the splicing of particular RNAs may be affected, and even less on how a number of splicing changes are coordinated during development or disease. To start addressing this question, we are analyzing WGS and RNASeq data from a number of cancer datasets. Although we are interested in all events of regulated splicing, we pay special attention to those related to the biosynthesis and function of the ribosome. A cycling cell depends on a suitable set of ribosomes to provide the necessary amount of structural and functional proteins before mitosis; paradoxically, making this machinery requires most of the cell's energy (as an illustrative example, a growing HeLa cell is making 1.6 x105 ribosomal proteins per minute). Thus, we expect that fast-growing cell, subjected to a strong selection (such as a tumor cell), will tweak this process to get any advantage. However, the analysis of the transcriptome of ribosomal proteins presents specific challenges because (a), it includes the mRNAs that are most abundant in the cell, but the amounts of each one are variable (while the ribosome has one copy of each protein); (b), the corresponding pre-mRNAs undergo little alternative splicing; and (c), the majority of human pseudogenes come from them, which introduces ambiguity when mapping reads to the genome. Our initial results suggest that processing of this set of transcripts is altered in cancer in unexpected ways, and we plan on strengthening our conclusions by expanding our analyses. In this context there are many opportunities for those with a strong motivation to document genomic strategies that control the tran-scriptome of specific gene families, like those related to the ribosome or the spliceo-some. The tasks involve quality analysis of raw RNASeq data, mapping using standard tools (for example, Hisat, STAR, and those related to direct sequencing of RNA), statistical analysis (Ballgown, Salmon, Vast-tools, DexSeq, or others), and modeling. Subject to progress, we would explore the use of transcriptomics data as a disease prognosis tool; namely, is a distinct distribution of transcripts indicative of a particular disease out-come?

References:

* Hussain, S. (2018) “Native RNA- Sequencing Throws its Hat into the Transcriptomics Ring” TiBS 1434. https://doi.org/10.1016/j.tibs.2018.02.007 * Guimaraes, JC. and Zavolan, M. (2016) “Patterns of ribosomal protein expression specify normal and malignant human cells” Genome Biol. 17:236-248 * Gupta, V. and J. R. Warner (2014). "Ribosome-omics of the human ribosome." RNA 20: 1004-1013. * Bitton, D. A., et al. (2014). "LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq." Genome Res 24(7): 1169-1179. * Acuna, L. I. and A. R. Kornblihtt (2014). "Long range chromatin organization: a new layer in splic-ing regulation?" Transcription 5. * Kawashima T et al (2014) Widespread use of non-productive alternative splice sites in Saccharo-myces cerevisiae. PLoS Genet. 2014 Apr 10;10(4):e1004249. * Zhang, J. and J. L. Manley (2013). "Misregulation of pre-mRNA alternative splicing in cancer." Cancer Discov 3(11): 1228-1237. * Fu, R. H., et al. (2013). "Aberrant alternative splicing events in Parkinson's disease." Cell Trans-plant 22(4): 653-661. * Plass, M., et al. (2012). "RNA secondary structure mediates alternative 3'ss selection in Saccha-romyces cerevisiae." RNA 18(6): 1103-1115

Expected skills::

knowledge of R is highly desirable

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Comments:

We are a wet lab but with knowledge of Bioinformatics and several questions to be approached using Bioinformatics but for which we have many molecular data. This is therefore an excellent setting for any knowledgeable, independent, ambitious, and highly motivated Bioinformatics student.

Master project 2020-2021

Personal Information

Supervisor	David Comas
Email	david.comas@upf.edu
Institution	Universitat Pompeu Fabra (UPF)
Website	http://www.biologiaevolutiva.org/dcomas/
Group	Human Genome Diversity Group

Project

Computational genomics

Project Title:

Human genome diversity: demography and adaptation

Keywords:

Human genome, genome diversity, demography, adaptation

Summary:

Our research is focused on the understanding of the current genomic diversity in human populations in order to establish the mechanisms, causes and consequences of this genetic variation. We are mainly focused on trying to disentangle two types of processes: - Demographic processes. Population history, such as migrations, expansions, bottlenecks and admixtures have modelled the extant genome diversity of humans. Using several genetic markers, such as mitochondrial or Y-chromosome lineages as well as high-throughput SNP coverage of several human populations, we have addressed some demographic questions, from regional aspects such as the genetic impact of the Bantu expansion in Central Africa to more global issues such as the colonization of whole continents. - Selective processes. The human genome as also been modelled by selective processes as a result of adaptations during the species history. We have analyzed parts of our genome in order to detect genetic signals yield by selective processes, such as adaptation to different environments and its relationship with human diseases. Disentangle both types of processes is not an easy task and our research deals with the analysis of the diversity of the human genome at a population level in order to detect demographic and selective processes.

Expected skills::

Basic bioinformatic skills in genome data analysis

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Tanya Vavouri
Email	tvavouri@carrerasresearch.org
Institution	Josep Carreras Leukaemia Research Institute (IJC)
Website	http://www.carrerasresearch.org/en/regulatory-genomics
Group	Regulatory Genomics

Project

Computational genomics

Project Title:

Evolution and biogenesis of mammalian PIWI-interacting RNAs (piRNAs).

Keywords:

genomic repeats, small non-coding RNAs, gene regulation

Summary:

Transposons and other repeats substantially contribute to genetic diversity in a species, to spontaneous mutations and regulatory innovations. PIWI-interacting RNAs (piRNAs) bound to PIWI proteins repress transposon activity in the germline. Repression of transposons is essential for normal progression of mammalian spermatogenesis. Transposons are highly enriched among piRNA producing loci and are transcriptionally and post-transcriptionally repressed by piRNAs. Nearly half of all piRNA-producing loci are protein-coding genes but, to date, it remains unknown why/how certain protein-coding genes are targeted for piRNA production during gametogenesis. The dynamic landscape of mammalian transposon insertions in genes and the strong association between piRNAs and transposons raise the question whether transposon insertions in genes have triggered piRNA production from these genes. The goal of this project is to use bioinformatics tools and available data (both from the lab and from other publications) to understand the effect of transposon insertions on gene function in the mammalian male germline. The specific objectives are to understand the extent and genetic causes of inter-individual variation in piRNA expression in mouse and to gain mechanistic insight into piRNA production from protein-coding genes.

References:

Ozata, D.M., Gainetdinov, I., Zoch, A. et al. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet 20, 89–108 (2019). https://doi.org/10.1038/s41576-018-0073-3

Expected skills::

R, scripting in bash, perl/python

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Eulàlia de Nadal
Email	eulalia.nadal@irbbarcelona.org
Institution	IRB Barcelona
Website	https://www.irbbarcelona.org/en/research/cell-signaling
Group	Cell Signaling

Project

Computational genomics

Project Title:

Decoding transcriptional heterogeneity one cell at a time (Deletome-seq)

Keywords:

single-cell RNA-seq, transcriptional heterogeneity, stress-responses, SAPK

Summary:

Single cell RNA-seq (scRNA-seq) has become the method of choice to dissect complex samples. These studies have provided striking insights such as the identification of novel cell types, but they also unveiled an unexpectedly high degree of transcriptional heterogeneity. The molecular mechanisms underlying this variability are not understood. Yet, cell-to-cell heterogeneity provides a mechanism to alter cell fate and cell identity. Currently, it remains a challenge to understand which mechanisms regulate transcriptional heterogeneity and their consequences. Here we propose to combine single cell transcriptomics with functional genome-wide genetic screen to identify the principles underlying transcriptional heterogeneity.

References:

- Nadal-Ribelles M&, Islam S&, Wei W&, Latorre P&, Nguyen M, de Nadal E, Posas F, Steinmetz LM. Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. Nat Microbiol. 4:683-692 (2019). - Nadal-Ribelles M, Islam S, Wei W, Latorre P, Nguyen M, de Nadal E*, Posas F*, Steinmetz LM*. Yeast Single-cell RNA-seq, Cell by Cell and Step by Step. Bio-Protocol Bio-protocol 9: e3359 (2019). - de Nadal E*, Posas F*. Osmostress-induced gene expression - a model to understand how stress-activated protein kinases (SAPKs) regulate transcription. FEBS J. 282: 3275-85 (2015). - de Nadal E, Ammerer G, Posas F. Controlling gene expression in response to stress. Nat Rev Genet. 12: 833-45. (2011).

Expected skills::

Biology, biochemistry or related fields

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Anna Bigas
Email	yguillen@imim.es
Institution	IMIM
Website	https://www.imim.es/programesrecerca/cancer/en_annabigas.html
Group	Cancer and Stem Cells

Project

Computational genomics

Project Title:

Comparative genomics and transcriptomics in stem cells and cancer

Keywords:

Transcriptomics, Genomics, leukemia, public repositories, cancer genetics

Summary:

Comparative genomics has become an essential tool for understanding genetic changes among different organisms, tissues and diseases. Alongside the latest development of powerful sequencing techniques and complex computational algorithms, researchers have been able to identify genetic drivers of cancer, as well as to provide insights into the biological pathways involved in carcinogenesis. Comparative genomics is thus considered a key element in cancer sciences, and bioinformatics is nowadays implemented in every multidisciplinary research team. The Stem Cells and Cancer group led by Anna Bigas is focused on the study of the molecular mechanisms involved in hematopoietic stem cells generation and hematologic malignancies, specially T-cell acute lymphoblastic leukemia (T-ALL). Moreover, it is implicated in the exploration of biological processes underlying colorectal cancer as it works in close cooperation with Colorectal Cancer group led by Lluis Espinosa. The expertise of Yolanda Guillén, a bioinformatics-trained biotechnologist, is crucial to perform the computational part of these projects and to understand the biological relevance of the results. We are glad to host an enthusiastic and motivated student willing to participate in: • The implementation of different bioinformatics pipelines in ongoing projects. We do use multiple approaches to analyze transcriptional (RNA-Seq and microarrays), genomics (ChIP-Seq) and epigenomics (ATAC-Seq) data. The student will learn how to prepare, run and interpret the results, from the raw sequencing data to the final output. Importantly, we do have access to a supervised computational cluster, which will make the student possible to understand how to work in such computational environment. • The exploration of transcriptional changes in T-ALL. Our main objective is to collect transcriptional data, mainly RNA-Seq, from public resources in order to screen for genetic expression changes in T-ALL. We are not only interested in identifying genes differentially expressed in T-ALL compared to normal cells, but to detect isoform switching patterns in cancer.

Expected skills::

Bash and R basic programming

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Master project 2020-2021

Personal Information

Supervisor	Rosa Fernández
Email	rosa.fernandez@ibe.upf-csic.es
Institution	Institute of Evolutionary Biology
Website	rmfernandezgarcia0.wixsite.com/metazomics
Group	Metazoa Phylogenomics Lab

Project

Computational genomics

Project Title:

Understanding how animals protect their DNA against UV light damage through the lens of comparative genomics

Keywords:

Comparative genomics; Phylogenomics; UV light-induced DNA damage repair; Nonmodel Organisms; Animals

Summary:

Almost 500 millions of years ago, several animal lineages conquered terrestrial environments from marine ones. One of the main challenges they needed to overcome was the protection of their DNA against UV light-induced damage, a threat that did not exist under water. Interestingly, we know almost nothing about how non-vertebrate animals repair their DNA after UV light-induced DNA damage. How do they do it? Do the different animals that conquered land (including nematodes, arthropods, earthworms or planarians, among other creatures) use the same mechanisms or different ones? This project aims at shedding light into the genomic underpinnings of UV light-induced DNA damage repair in non-model organisms through a comparative genomics spyglass.

Expected skills::

Python and/or perl programming. Knowledge on phylogenetics and comparative genomics desirable but not essential.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Ilario De Toma
Email	ilario.detoma@gmail.com
Institution	CRG
Website	
Group	Mara Dierssen

Project

Computational genomics

Project Title:

EGCG induced change in the phospho-proteome of DYRK1A overexpressing cells

Keywords:

Down syndrome; Proteomics; Time-course; EGCG; Hippocampus

Summary:

DYRK1A is a gene triplicated in Down syndrome that regulates the phosphorylation of several targets. Mice overexpressing DYRK1A show cognitive alterations. Interestingly, EGCG, the main polyphenol extracted from green tea, it is a DYRK1A inhibitor and ameliorates the cognitive impairment in DYRK1A transgenic mice and other DS mouse models. We performed an iTRAQ experiment on hippocampal primary neuronal cultures, labeling 5 different time points upon EGCG treatment, 5 uM (0, 5', 15', 30', 120') both in transegenic and wild type cells. This will shed lights in the acute phase action of EGCG actions, with the future goal to improve its efficacy for treatment purposes.

Expected skills::

Using R and RStudio

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Davide Piscia
Email	davide.piscia@cnag.crg.eu
Institution	CNAG-CRG
Website	www.cnag.cat
Group	Data Analysis Team, Bioinformatics Unit

Project

Computational genomics

Project Title:

Deep learning models applied to rare diseases: where do we stand?

Keywords:

deep learning, rare diseases, variant effect prediction, non-coding regions

Summary:

Deep learning models have been used extensively in image recognition and natural language processing, but in the last years they have been also applied to genomics. In the last years some interesting initiatives were started such as kipoi (https://kipoi.org) and selene (https://selene.flatironinstitute.org) whose aim is to facilitate the use of deep learning in biological contexts. One of the objectives of this project is to evaluate and summarize the state of the art of genomics deep learning, especially for variant effects prediction in non-coding regions . Once the most promising models have been selected, the candidate will have to apply it to some of the rare disease whole genome datasets hosted at CNAG-CRG. In this task she/he will have to be able to run the models in the CNAG-CRG HPC cluster (GPUs enabled) and do a first assessment of the model predictions. The long-term goal of this work is to integrate deep learning as a functionality in the RD-Connect GPAP platform (https://platform.rd-connect.eu).

Expected skills::

The candidate is expected to have good computational skills, especially in python.

Possibility of funding::

No

Possible continuity with PhD: :

To be discussed