

Bioinformatics Student Exchange Program

CSIRO – Germany – China



1 BSEP 2017

Australia was featured in a recent [Nature article](#) stating that “Scientists from across the world are attracted to the country, which competes internationally by focusing on its strengths”. [The Commonwealth Scientific and Industrial Research Organisation \(CSIRO\)](#) is one of the largest and most diverse scientific organisations in the world. By igniting the creative spirit of our people, we deliver great science and innovative solutions that benefit industry, society and the environment.

In order to give students from Germany and China the opportunity to contribute to world-class research and gain experience in an international research environment, the eHealth program is running the Bioinformatics Student Exchange Program (BSEP) with German and Chinese Universities. The program is aimed at Master and Honours students and invites them to join CSIRO to conduct original research. This is an exciting opportunity to forge new collaboration with CSIRO as the hub for bioinformatics research in Australia.

Master and Honours students in Bioinformatics will have the opportunity to join CSIRO for 23 weeks (5 months) and undertake a research project that contributes towards their Thesis. The project will be proposed by CSIRO researchers who also agree to co-supervise the student and assist in writing the thesis.

| University | Contact Person |
|---|---|
| Freie Universität Berlin | Prof. Dr. Annalisa Marsico RNA Bioinformatik Phone: +49 30 8413 1843 Fax: +49 30 8413 1960 Email: Annalisa.Marsico@fu-berlin.de |
| Eberhard Karls University Tübingen | Dr. Julian Heinrich Applied Bioinformatics Group Email: heinrich@informatik.uni-tuebingen.de |
| Justus-Liebig-University Giessen | Prof. Dr. Alexander Goesmann Bioinformatik und Systembiologie Tel. +49 (0)641 99-35801 Email: Gwyneth.schulz@computational.bio.uni-giessen.de |
| CSIRO | Dr. Denis Bauer Transformational Bioinformatics, eHealth, CSIRO Phone: +61 2 9325 3174 Email: denis.bauer@csiro.au |

1.1 Key dates

| Date | |
|--------------------------------|---|
| June | CSIRO calls for project proposals |
| 31 st July | Program Booklet sent to the Universities |
| Early August to early November | Deadline for PROMOS or equivalent funding application |
| Dec | Thesis committee assesses suitability of projects and identifies appropriate co-supervisor amongst the faculty. |
| Jan | Students choose proposals and CSIRO starts recruitment process (interview, visa) |
| May | Students commence research in Australia |
| Oct | Students return home |
| Nov | Students finalise reports and write master thesis with input from CSIRO researchers |

1.2 Funding

Students are encouraged to apply for funding. Unless stated otherwise, the projects will not provide funding.

1.3 Germany

PROMOS

German funding through **PROMOS** (Deadline Early October to early November), which will cover

- from 300 to 500 EUR per month and / or
- Traveling costs up to 1950 EUR

Note, PROMOS is not explicitly paying a health insurance, this hence needs to be covered by the student.

DAAD

The DAAD offers **FIT- Internationale Forschungsaufenthalte in der Informationstechnologie für Masterstudierende**, which can be applied for at any stage, with notification of success within 3 Months (recommended application date no later than October)

- 875 EUR per months
- contributions to travel costs
- contributions to insurances

There are also other funding sources available such as <http://www.ranke-heinemann.de>.

1.4 How to apply and other resources

Please choose the project you are interested in and get in touch with your contact person listed above. Your first step will be to organize funding by applying for PROMOS or equivalent sources (DAAD). After a successful interview in January, CSIRO will issue a contract with a visa sponsorship number. It is crucial to apply for the Australian Visa quickly as it can take up to 3 months to be approved.

VISA:

<https://www.border.gov.au/Trav/Visa-1/402->

Information about the VISA subclass 402 Trainee and Research.

Address where to send the application:

<https://www.border.gov.au/Lega/Lega/Help/Location/australia>

Tasmania-Hobart office

Health insurance:

<http://www.health.gov.au/internet/main/Publishing.nsf/Content/Overseas+Student+Health+Cover+FAQ-1#insurersofferohc>

e.g. the BUPA caters for VISA subclass 402.

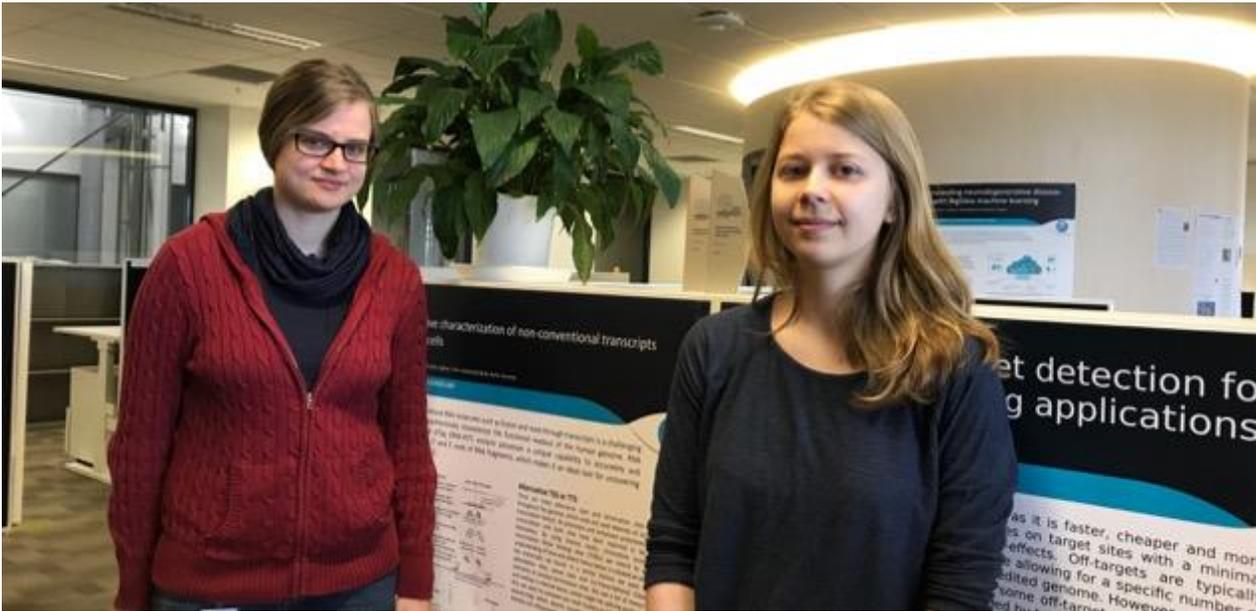
German information on going to Australia:

<http://www.reisebire.de/>

Official government website with information about studying and living in Australia

www.studyinaustralia.gov.au

1.5 Experience Report from 2016



We are Denise and Sara, two students from FU Berlin currently undertaking our Master thesis at CSIRO in the group Transformational Bioinformatics of Dr. Denis Bauer. We like to share our experience and advice with you in order to make your application as easy as possible and to prepare you for your possible exciting exchange project.

CSIRO is one of the biggest government-funded research organisations and our group is working in many cutting-edge fields such as genome engineering, cloud-based tool development and functional genomics. As the team is not too big a productive and pleasant atmosphere is created where people support each other and share thoughts about different strategies and approaches to solve problems. From the very first moment it was a very supportive way of working that we experienced. You will get insights into what it means to collaborate with industry and organize bigger research teams in CSIRO-internal meetings but you also get to attend external conferences where you can present your project to a broader range of scientists, collect more ideas and new ways of thinking that might support you while working on your thesis and even open up some opportunities after graduation. Depending on your project, you could also come along to meet with Google and/or AWS in their Sydney headquarters. All in all, you will be introduced to exciting research working with people who are enthusiastic and very open to guide and assist the development of your own exciting project.

In addition, we want to share some useful advice that should help you simplify the application and organization process. First of all, you will need to look for funding. One of us got funding by PROMOS and one by the DAAD. The advantage of the DAAD is that it will provide you with more money, on the other side it is more difficult to obtain. For both funding sources, you need a motivation letter, your transcript of records and a recommendation of a professor. For the DAAD you would need to provide a research proposal containing a brief plan of attack of your thesis as well.

As soon as you get the confirmation of your funding, which can take a few months you should apply for a visa. This process might take 2 to 3 months and it can also happen that you get asked for more information to provide which extends the process again. Also, you should book a variable flight or book the flight after you received the visa. Fortunately, CSIRO is flexible with the start of your project. It happened to Denise that her visa arrived later than expected but it was no problem at all to start a week later.

If you managed to get your funding and visa as expected, you should think about your accommodation. As we both stay and work in Sydney, we are sharing some ideas concerning this city in the following. Sydney consists of many districts so there are plenty of options. The inner city is close to all cafes, bars and restaurants but on the other hand more expensive and half an hour train ride away from the CSIRO department at North Ryde. Living next to our workplace is convenient and cheaper but on the other side also a little outside of the inner-city life. We tried both and it is more or less depending on your wishes and needs what you'd prefer. You might want to consider if your place has a proper heating system as most places in Sydney are badly isolated and heaters are rare. Winter in Sydney might be warm and sunny during the day but also quite cold at night.

Sydney itself is a multicultural city that offers all kinds of different activities which makes it an amazing place to live. The inner city and surrounding districts offer all types of shopping, awesome coffee, concerts, bars and amazing restaurants (check out Newtown). In the east famous beaches such as Bondi and Manly are awaiting which even in winter are amazing places to be at and release the stress with beautiful walks along the coast. All around Sydney and even in the surroundings of Sydney there are national parks, the closest one to CSIRO is just 15 minutes away. Two hours from Sydney are the beautiful Blue Mountains where you can spend a day or a weekend hiking, exploring canyons, caves, forest and wildlife. Besides Sydney offers a good traffic system with trains, busses and ferries, so that it is convenient to get around the city even at night.

For any place and city we would recommend you the meetup app to get to know new people and feel at home in a new environment. It offers different group activities including all sorts of sports, movies, discussions, meeting other new people for drinks or just going for a walk.

We would warmly recommend taking the opportunity to join a research team at CSIRO to work on an exciting project with an amazing team, get insights into new research areas and benefit from the work abroad. If you have further questions concerning the organization, feel free to ask us at any time.

Cheers,
Denise and Sara
(denise.thiel@csiro.au, sara.hetzel@csiro.au)

1.6 Projects

| | | |
|----------|---|----------|
| 1 | BSEP 2017 | 2 |
| | <i>BSEP01: CRISPR-Cas9 applications in human health</i> | 8 |
| | <i>BSEP02: Comparative analysis of CD4+ and CD8+ TCR repertoire by machine learning</i> | 9 |
| | <i>BSEP03: Linkage-directed prioritisation for disease variant detection</i> | 10 |
| | <i>BSEP04: Distribution of transposable elements distributed in the Tiwi genomes</i> | 11 |
| | <i>BSEP05: Extending the CursedForest (a distributed random forest) framework for genomics</i> | 12 |
| | <i>BSEP06: Targeted eQTL Analysis of SNPs and expression markers for Alzheimer's disease.</i> | 13 |
| | <i>BSEP07: Deep learning approaches to computational biology problems</i> | 14 |
| | <i>BSEP08: Genetically invariant genes as drug targets in pathogenic nematodes</i> | 15 |
| | <i>BSEP09: Detecting true genetic variation in polyploid crop genomes</i> | 16 |
| | <i>BSEP10: Web-based framework for genetic risk prediction</i> | 17 |
| | <i>BSEP11: An integrated database of on-target sequences of the CRISPR-Cas system for bread wheat</i> | 18 |
| | <i>BSEP12: Automated grading of diabetic retinopathy relying on deep learning</i> | 19 |

BSEP01: CRISPR-Cas9 applications in human health

The development of the CRISPR-Cas9 system has revolutionized genome engineering, making it possible to directly target almost anywhere in the genome for editing. The reliable application of the CRISPR-Cas9 technology requires the identification of the optimal target site, as activity can vary substantially between sites. This is particularly important if the technology is hoped to be applied in the human health space.

The goal of this project will be to use BigData and Machine Learning approaches to understand what factors contribute to CRISPR-Cas9 activity and how these can be leveraged to improve predictions of target activity.

The student will build upon the team's extensive CRISPR software portfolio, contributing to the next version of GT-Scan with the option of writing a first-author paper. Furthermore, the student will have the opportunity to develop cloud-based services, likely in close collaboration with AWS Solution Architects in Sydney.

The student will perform

- Research and review of potential applications of CRISPR-Cas9 in the area of human health
- Analysis of CRISPR-Cas9 activity *in vitro*
- Development of predictive models

Relevant field/s of study

- Machine Learning
- Genome Engineering
- BigData

Supervisor

Laurence Wilson, Denis Bauer (CSIRO, Health&Biosecurity)

email Laurence.wilson@csiro.au

Location: Sydney NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP02: Comparative analysis of CD4+ and CD8+ TCR repertoire by machine learning

The T-cell Receptor (TCR) repertoire is diverse and serves as a reservoir for recognizing all potential antigens. It is understood that the amino acid composition of TCR is random. Amongst the two major T-cell types, CD4⁺ and CD8⁺, the β chain of TCR is extremely diverse and distinct. The richness of TCR β in CD4⁺ is approximately 5 times greater than CD8⁺, and the TCR β amino acid compositions from these two T-cell types are almost mutually exclusive, which is much lower than random expectation (Li et al., 2016. *J Leukoc Biol*). Building on these findings, this project aims to

- Analyse and compare the TCR repertoires of CD4⁺ and CD8⁺ cells from a large Tuberculosis cohort;
- Understand the molecular mechanism for the exclusiveness of TCR β amino acid composition between CD4⁺ and CD8⁺ cells;
- Implement classifier of CD4⁺ and CD8⁺ cells based on TCR repertoire information using machine learning algorithms.

This work will potentially lead to a first-author publication in a high impact journal.

The student will perform

- Analysing TCR repertoire sequencing data from CD4⁺ and CD8⁺ T-cells
- Developing Spark-based machine learning algorithms for classifying CD4⁺ and CD8⁺ T-cells based on TCR repertoire information

Relevant field/s of study

- Bioinformatics
- Machine learning
- Immunology

Supervisor

Oscar Luo (CSIRO, Health&Biosecurity) Guobing Chen (Guangzhou Medical University)

email: Oscar.Luo@CSIRO.au

Location: Sydney NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP03: Linkage-directed prioritisation for disease variant detection

Even healthy individuals have hundreds of variants that disrupt the normal function of the resulting protein. Hence identifying causative variants by filtering for deleterious variants may sometimes not be sufficiently powerful. Inherited high impact deleterious variants amongst family members is observed clinically as a familial syndrome. Recently developed tools such as pVAAST (Nat Biotech, 2014) make use of shared chromosomal segments between related individuals to identify genetic variants that directly influence disease risk. These approaches extend the variant prioritization and case-control association features with linkage analysis methods specifically designed for sequence data. These models are broadly similar to traditional linkage analysis but capable of modelling de novo mutations and handling incomplete penetrance or locus heterogeneity.

This project proposes to apply pVAAST to existing whole genome and whole exome data of two disease cohorts, one from familial colorectal cancer and one from familial motor neuron disease patients. We also wish to compare with other approaches, including non-parametric linkage in Merlin, the Generalized Family-Based Association Test and methods for identifying shared chromosomal segments in distantly related individuals. This will require parsing VCF files into standard linkage/association file formats used by PLINK and Merlin. We propose to extend SeqAn to output these formats.

The student will perform

- Perform a literature review to identify all linkage-aware association predictors
- Apply pVAAST and related tools to two disease cohorts
- Develop a SeqAn-based parser for linkage file formats
- Compare pVAAST results to other linkage and family-based association test outputs
- Develop approaches for identifying distantly related individuals

Relevant field/s of study

- Knowledge of genetics
- Bioinformatics

Supervisor

Jason Ross (CSIRO, Health&Biosecurity)

phone +61 2 9490 5015 or email Jason.Ross@CSIRO.au

Location: Sydney NSW

BSEP04: Distribution of transposable elements distributed in the Tiwi genomes

A large proportion of the human genome (approximately 45%) comprises of transposon and transposon-like repetitive DNA elements, although only a very small number remain active. They comprise a large number of families based on sequence comparison. Depending on their genomic location and activity, they can influence gene expression, individual phenotypic diversity and disease.

The McMorran group has generated 120 complete human genome sequences from the Tiwi people, generated as part of a study investigating causes of renal disease. This represents the first resource of its kind.

This project aims to conduct a bioinformatics characterisation of transposable elements in whole genome sequences of the Tiwi people. Elements will be identified, characterised and mapped for each individual. An assessment of common and unique elements that may influence gene expression and protein function will also be conducted.

The student will perform

- Whole genome sequence analysis identifying transposable elements and distinguish them against other repetitive elements using third party tools
- Utilize annotation information (ENCODE, UCSC) to characterise functional classes
- Compare findings between 120 individuals to identify structural groups

Relevant field/s of study

- Proficiency with SeqAn for motif search
- C++ advanced
- statistical concepts

Supervisor

Brendan McMorran (ANU), Denis Bauer (CSIRO, Health&Biosecurity)

brendan.mcmorran@anu.edu.au

Location: Sydney NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP05: Extending the CursedForest (a distributed random forest) framework for genomics

CursedForest is a tailored Hadoop/Spark-based implementation of random forests specifically designed to cater for ``big" (many samples) and ``wide" (many features) datasets, which was recently featured in the Databricks Engineering Blog [1].

The current implementation has been successfully used in detecting SNPs associated with a phenotype in application with up to 80M variables. Making CursedForest applicable to more genomic research areas, such as transcription analysis, this project aims to add some of these features:

- 1) handling of categorical variables (by subset selection);
- 2) a regression penalty term (currently uses Gini index);
- 3) proximity matrix;
- 4) conditional inference trees.

[1] <https://databricks.com/blog/2017/07/26/breaking-the-curse-of-dimensionality-in-genomics-using-wide-random-forests.html>

The student will perform

- design of implementation
- coding in Scala and perhaps other languages
- testing of implementation, documentation

Relevant field/s of study

- Computer science
- machine learning
- statistics

Supervisor

Denis Bauer (CSIRO, Health&Biosecurity), Robert Dunne, Piotr Szul (CSIRO, D61)

Email: Denis.Bauer@CSIRO.au

Location: Sydney

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP06: Targeted eQTL Analysis of SNPs and expression markers for Alzheimer's disease.

Using a targeted approach to molecular pathways and eQTL analyses, the current project will investigate approximately 2000 SNPs and genes to ascertain their relationship with Alzheimer's disease. Implementing a novel genomic design, the candidate will use complex statistical methods to define the interrelationships between SNPs, gene expression and disease phenotype. The outcome would be to build a Shiny App that would take data from a set of SNPs, data from a set of gene expression, and data regarding disease phenotype, and perform analyses to define optimal sets of markers associated with outcome. The app will be adjustable to vary numbers of markers identified, and present graphics to represent network based associations. It will be applicable to anyone who wants to analyse their favourite set of SNPs, and customise the parameters for network analyses.

The student will perform

- The student will perform statistical analyses, Shiny App design, and preparation of a manuscript to showcase the app.

Relevant field/s of study

- Bioinformatics
- Biostatistics

Supervisor

James Doecke (CSIRO, Health&Biosecurity)

+617 3253 3697 or james.doecke@csiro.au

Location: Herston Brisbane

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP07: Deep learning approaches to computational biology problems

Deep learning networks have achieved some significant success in application areas like image and speech recognition. They are now being more widely applied to genomics problems. See Deep learning for computational biology [1] for a review of recent applications of deep learning in regulatory genomics.

This project would involve the use of the Tensorflow [2] framework to design and implement an appropriate deep learning network for the application problem.

As first use case, Tensorflow will be used to predict ethnicity from WGS data (the 1000 Genomes project data) and compared against other machine learning methods, such as random forests.

The student will have the opportunity to develop cloud-based services, likely in close collaboration with Google Solution Architects in Sydney.

[1] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4965871/>

[2] <https://www.tensorflow.org/>

The student will perform

- data preparation
- network design and testing

Relevant field/s of study

- bioinformatics
- machine learning
- statistics

Supervisor

Denis Bauer (CSIRO, Health&Biosecurity), Robert Dunne (CSIRO, D61)

Email: Denis.Bauer@CSIRO.au

Location: Sydney, NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP08: Genetically invariant genes as drug targets in pathogenic nematodes

We will aim to identify genes which might encode useful drug targets for the control of pathogenic nematode parasites, which are important in both human and veterinary medicine. A variety of criteria have been used in the past to define gene products as “druggable”, but the extent of genetic variation has not been explored as a criterion before. The group of nematode parasites, Strongylid, or clade V nematodes, contains many parasitic species of importance with most having very high levels of genetic variation.

Using 31 NGS genomic libraries for the nematode *Haemonchus contortus*, you will identify exons from well covered “druggable” gene sequences. Subsequently, you will identify the synonymous and non-synonymous variants (SNP and INDEL) within these exons and genes and further classify them into bins with differing levels of non-synonymous variation. To ascertain the function of these snps/indel, you will perform functional annotation, gene-ontology (GO) enrichment comparison between bins and map the metabolic pathways of key druggable gene families using KEGG maps. For some subsets of gene families of interest, you will produce phylogenetic gene trees (using maximum likelihood and Bayesian MCMC algorithm) to elucidate the evolutionary relationships between the *H. contortus* druggable genes variation you identified and ortholog/paralog genes within related strongylid species.

At the end of this project, we aim to identify genes with low levels of genetic variation, especially those present in multiple species, and those in functional categories defined as potentially “druggable”. These will be used in subsequent work to develop new controls for parasitic nematode diseases in humans and animals.

The student will perform

- align *H. contortus* genome sequence data and identify exons with sufficient coverage of sequence
- map the synonymous and non-synonymous variants
- divide the genes into bins with differing levels of non-synonymous variation
- map the membership of these bins to functional categories
- choose genes from “druggable” categories for further analysis
- Using publically available data for other species, identify genes in these same groups within other Strongylid species
- Using available public data, determine if the genes in other species have similar or differing levels of non-synonymous genetic variation
- Study gene families to determine if some gene family members are more evolutionarily constrained than others
- Produce a list of the “best” candidate genes based on this analysis pathway in *H. contortus*
- A scientific publication may be prepared as a result

Relevant field/s of study

- Genetic variant mapping
- phylogenetic tree building
- functional annotation of genes
- developing bioinformatics pipelines
- vaccine and drug discovery from genetic data

Supervisor

Peter Hunt, Marielle Babineau (CSIRO)
phone +61 2 6776 1321 or email peter.hunt@csiro.au
Armidale NSW

Funding

CSIRO offers a top-up stipend for this project: \$2500 one-off for air travel.

BSEP09: Detecting true genetic variation in polyploid crop genomes

Polyploidy – where more than two complete genomes are found in a cell – is a fascinating biological phenomenon that may help organisms adapt rapidly to change. However, despite being common in commercially relevant crop species such as wheat, our understanding of the where, how, and why of polyploidy is incomplete.

Understanding of polyploidy is limited in part by the unique bioinformatics challenges it poses. For instance, software packages for finding genetic variants commonly use models that assume the organism is diploid or have limited support for high ploidy levels. Other software without these restrictions are extremely limited in their ability to distinguish between true variants versus sequence errors, relying on arbitrary ‘frequency cutoffs’.

Here, you will develop a command-line variant calling software package that can accurately identify true genetic variants in polyploid data, by developing statistical tests based on unique features of sequencing errors, e.g. position in a read, strand bias, quality scores etc. without any assumptions about genome copy number. You will also have the opportunity to apply this software to real data to address important biological questions. In addition to an in-demand software package, this project is likely to result in at least one highly cited scientific paper.

The student will perform

- Investigation of the properties of errors vs variants, e.g. choosing the most appropriate statistical distribution to model the quality score of an error or true variant
- Design of statistical tests to distinguish errors from true variants
- Developing code to implement these tests, e.g. using the samtools C API
- Apply the resulting software to:
 - simulated data,
 - real sequence reads with known variants,
 - experimental crop species data

Relevant field/s of study

- genomics
- computer programming
- statistics

Supervisor

Kerensa McElroy (CSIRO), Alex Whan (CSIRO)

Funding

CSIRO offers a top-up stipend for this project; exact contact supervisor for details.

BSEP10: Web-based framework for genetic risk prediction

The continuously decreasing cost of sequencing (NovaSeq US\$100) and the landmark decision by the FDA to approve genetic testing as done by 23andMe, a commercial company, has paved the way for a second wave of direct-to-consumer genetic products. CSIRO hence aims to expand its portfolio in this space. Building on the team's LifeDNA framework, which predicts obesity risk from genomic profiles, the student will utilize Genome England's PanelApp [1] to build a generic testing framework that takes expert approved genetic loci associated with a specific trait and joins them in a genome wide risk score.

[1] <https://panelapp.extge.co.uk/>

The student will perform

- Build framework to read in a JSON object that holds the formula for combining loci
- Develops the API to interact with the PanelApp framework
- Design and implement the web service to display the result

Relevant field/s of study

- Python
- Web services, API
- Genomic/Genetic

Supervisor

Denis Bauer (CSIRO, Health&Biosecurity)

email Denis.Bauer@CSIRO.au

Location: Sydney, NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP11: An integrated database of on-target sequences of the CRISPR-Cas system for bread wheat

CRISPR-Cas9 and CRISPR-Cpf1 are endonucleases for the CRISPR system in genome engineering applications and have been applied to engineer targeted modifications in polyploid genomes. Simultaneous editing of multiple homoeoalleles has been reported in hexaploid bread wheat.

Here, we would evaluate the genes associated with important agronomic traits, such as yield include flowering time, height, disease resistance, and identify putative on-target sequences for CRISPR-Cas9 and CRISPR-Cpf1 enzymes in the promoter of these genes, and predict the on-site activity using appropriate prediction models.

This integrated database aims to accelerate the research of genome editing in hexaploid bread wheat and would be freely accessible online.

The student will perform

- Identifying genes associated with specific important agronomic traits
- Identifying putative CRISPR-Cas9 and CRISPR-Cpf1 on-target sequences in the promoter of above genes
- Predicting on-target activities using appropriate prediction models.
- Developing web portal for related data visualisation using appropriate framework (such as using GBrowse2 framework)

Relevant field/s of study

- Bioinformatics
- Genomics

Supervisor

Kaitao Lai (CSIRO, Health&Biosecurity)

phone +61 (0)2 93 253 068 or email Kaitao.Lai@csiro.au

Location: Sydney, NSW

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

BSEP12: Automated grading of diabetic retinopathy relying on deep learning

Diabetic retinopathy (DR) is a serious eye disease that occurs due to diabetes mellitus and it has become one of the most common causes of blindness in the present world. Based on latest reports by 2030 there is an epidemic rise of 4.4% in the global prevalence of diabetes. Early detection of diabetic retinopathy can enable timely treatment by minimizing further deterioration. Therefore regular screening of diabetic patients' retina is very essential and automated or computer-assisted analysis of diabetic patients' retina can help eye care specialist to screen larger populations of patients. The aim of this project is to develop automated methods to perform computer-assisted grading of DR using color fundus images of the eye.

Deep learning, as a field of machine learning, has dramatically pushed the performance of many computer-assisted analysis; however, many important research questions are still open. How should one interpret its decision making process? How domain specific knowledge can be incorporated into the system for further improvement in the performance?

The goal of this project will be to critically analyse those research questions and to augment the deep learning system for the diagnosis of diabetic retinopathy. More specifically, the project will focus on understanding deep networks and its inner workings, and then incorporating domain (e.g. diabetic retinopathy) specific information in the decision making process to ensure higher grading accuracy.

The student will perform

- Literature review (will be guided)
- Algorithm development (using Python/ C++/ Matlab)
- Writing reports/papers

Relevant field/s of study

- Computer Science
- Applied Mathematics

Supervisor

Dr Sajib Kumar Saha, Dr Di Xiao, Prof Yogesan Kanagasingam

phone on (061) 893336116 or email Sajib.Saha@csiro.au

Location: Floreat, Perth WA

Funding

CSIRO offers a top-up stipend for this project: \$300/week contribution towards living expenses.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

FOR FURTHER INFORMATION

Health and Biosecurity
Denis Bauer
t +61 2 9325 3174
e Denis.Bauer@csiro.au
w <https://aehrc.com>

AT CSIRO WE SHAPE THE FUTURE

We do this by using science to solve real issues. Our research makes a difference to industry, people and the planet.

As Australia's national science agency we've been pushing the edge of what's possible for over 85 years. Today we have more than 5,000 talented people working out of 50-plus centres in Australia and internationally. Our people work closely with industry and communities to leave a lasting legacy. Collectively, our innovation and excellence places us in the top ten applied research agencies in the world.

WE ASK, WE SEEK AND WE SOLVE