

2027 UNDERGRADUATE STUDENT PROJECTS

COMPUTATIONALLY ENABLED DRUG DISCOVERY GROUP

Project 1

Project Title:	Engineering RNA splicing to uncover new cancer vulnerabilities
Supervisor:	Pragya Gupta PGupta@ccia.org.au
Suitable for:	Undergraduate students 8 weeks Biological Sciences
Available:	Term 3 2027; Summer 2026-2027; Term 2 2027

Project outline:

Background and aims

Cancer cells can change the way they process RNA to help them grow, survive and resist treatment. This project will investigate pre-mRNA splicing, a process in which non-coding regions called introns are removed and coding regions called exons are joined to produce mature messenger RNA.

Some genes contain poison exons. When these exons are included in an RNA transcript, they introduce an early stop signal and can reduce the amount of functional RNA—and therefore protein—produced by the gene.

In this project, you will explore whether poison-exon inclusion can be controlled in selected cancer-related genes. This could help us develop new tools for studying gene regulation and identify potential weaknesses that could be targeted in aggressive cancers.

Approach

Our team is developing modified versions of U1 small nuclear RNA (U1 snRNA), a molecule that helps cells identify where RNA should be cut and joined during pre-mRNA splicing. These engineered U1 snRNAs are designed to change how specific poison exons are processed in cancer-related genes.

As part of this project, you will test whether the engineered U1 snRNAs produce clear and specific changes in their target RNA transcripts.

You will also help establish SHAPE-MaP, a sequencing-based technique used to study RNA structure. This method will allow us to examine whether engineering U1 snRNA changes how the molecule folds or affects the accessibility of important regions.

Expected outcomes and skills acquired

- By completing this project, you will gain practical experience in:
- *RNA biology and pre-mRNA splicing.
 - *Molecular and cell biology techniques.
 - *Designing and testing engineered RNA molecules.
 - *RNA structure analysis and next-generation sequencing.
 - *Interpreting experimental data and evaluating whether a molecular tool works as intended.

2027 UNDERGRADUATE STUDENT PROJECTS

THINK (THERapeutic INnovations for Kids)

Project 2

Project Title: Scaling cellular measurement for AI-guided drug discovery
Supervisor: AJ Sethi | ASethi@ccia.org.au
Suitable for: Undergraduate students | 8 weeks | Biological Sciences; Biotechnology
Available: Summer 2026-2027

Project outline:

Background and aims

Three children in Australia die from cancer every week, and bringing a new medicine to them takes 17 years on average. Much of that time goes into generating the data that tells us how cancer cells respond to candidate drugs. In the Computational Drug Discovery Biology group at Children's Cancer Institute, we build and orchestrate large scale sequencing-based drug screens to accelerate the process of drug discovery. These data steer drug discovery campaigns for childhood cancers and train the next generation of predictive models at the intersection of biology and AI.

Most of our sequencing-based screens require the extraction of RNA – which encodes cellular phenotypes – from 384-well plates. Conventional column-based extraction is manual, slow and costly, and it cannot scale to our target sample throughput. We are replacing it with magnetic-bead extraction performed directly in 384-well plates by liquid-handling instruments, and this project will own that step.

This project aims to:

1. Quantify RNA yield, purity, integrity and well-to-well cross-contamination across conditions, benchmarked against a conventional column kit.
2. Assess between-well contamination using RT-qPCR of well-specific spike in transcripts.

Approach

The student will join CEDD scientists and own this one problem end to end.

Phase 1 — foundations (two weeks). Induction, training in RNA extraction and RNA-safe handling, and the current manual and automated extraction workflows. Introduction to bulk liquid handling robots. Introduction to RT-qPCR.

Phase 2 — optimisation (four weeks). Starting from the group's existing magnetic-bead extraction workflow, the student will evaluate a small, predefined set of conditions, focusing on two selected parameters, against a conventional column-based method. Each condition is scored for yield (fluorometric quantification), purity (absorbance ratios), integrity (automated capillary electrophoresis) and cross-contamination (empty and alternating-well controls), with RT-qPCR of spike-in transcripts as an orthogonal check.

Phase 3 — validation and write-up (two weeks). The best-performing protocol is locked in and a full 384-well plate is taken end to end through RNA extraction and RT-qPCR. The student analyses the QC metrics (library yield, ribosomal fraction, documents the protocol, and presents the results to the group.

Expected outcomes and skills acquired

Outcomes. An introduction to high-throughput RNA-based workflows, and familiarisation with key protocol and chemistry involved therein. Substantial contributions will be recognised in any resulting methods publication. Students who enjoy the work will find a natural path into an Honours project with the group.

Skills developed:

- RNA-safe laboratory practice
- Design of experiments and quantitative method optimisation
- Nucleic-acid QC instrumentation: fluorometry, spectrophotometry, capillary electrophoresis and qPCR
- RT-qPCR experimental workflow and analysis
- Electronic lab notebooks and reproducible protocol documentation

2027 UNDERGRADUATE STUDENT PROJECTS

COMPUTATIONAL BIOLOGY

Project 3

Project Title: Rapid genomic analysis tool using nanopore based long-read DNA sequencing
Supervisor: Piyushkumar Mundra | PMundra@ccia.org.au; Chenkai Ma | cma@ccia.org.au

Suitable for: Undergraduate students | 8 weeks | Other (Bioinformatics); Mathematics and Computing; Bioengineering

Available: Term 1 2027

Project outline:

Background and aims

Current approach requires ~14 days to generate data, convert to sequencing files, processing and analysing data to call cancer associated variants. Rapid DNA sequencing using Oxford Nanopore Technology (ONT) have created opportunities to obtain clinically relevant genomic information much faster than traditional sequencing workflows. While rapid approaches have already shown considerable potential for cancer classification, there is an opportunity to extend this concept beyond determining tumour type and develop a broader real-time cancer genomics toolkit. Such a toolkit could progressively identify important genomic alterations as sequencing data are generated, potentially providing information relevant to diagnosis, tumour biology and treatment selection.

This project will use existing sequencing data from 14 matched tumour–normal pairs. Using the timestamps associated with sequencing reads in the BAM files, the project will reconstruct different stages of a sequencing run and determine how quickly reliable genomic results could have been obtained. The main aim is to develop and evaluate a computational framework for time-resolved cancer genome analysis, initially focusing on how sequencing depth and elapsed sequencing time influence the detection of clinically and biologically relevant genomic alterations.

Approach

The student will work with existing, de-identified tumour and matched-normal sequencing datasets that have already undergone comprehensive analysis. The completed analyses will provide a reference against which rapid or early-stage results can be compared. Using timestamps associated with sequencing reads, the student will computationally recreate sequencing data available at successive points during each run, for example by generating subsets representing progressively increasing sequencing time and genome coverage. At each stage, established bioinformatics tools will be used to investigate cancer-associated structural variants and potentially single nucleotide variants, small insertions and deletions. Further, ONT sequencing provides DNA methylation information. The student will also develop pipeline to provide methylation calls.

The student will compare results obtained at different stages with the final analysis to determine when individual alterations can first be detected reliably. The work will also involve developing scripts and workflows, primarily using Python, R and standard genomic analysis tools, to automate data processing, summarise results and visualise how genomic information accumulates during sequencing. The resulting framework can then be tested across the 14 tumour–normal pairs to assess its reproducibility across different samples.

Expected outcomes and skills acquired

The major outputs of the project will be:

- 1) A prototype computational toolkit for time-resolved cancer genomic analysis, together with an evaluation of how rapidly different types of genomic alterations can be detected during sequencing. The project is expected to generate preliminary data and summary metrics showing the relationship between sequencing time, sequencing depth and the accuracy or completeness of cancer genomic results.
- 2) It should also identify which classes of genomic alterations can be detected confidently at relatively low sequencing depth and which require more extensive sequencing.
- 3) Depending on progress, the toolkit could provide the foundation for subsequent development of a more comprehensive real-time cancer genome interpretation pipeline and future validation in prospective sequencing studies.

2027 UNDERGRADUATE STUDENT PROJECTS

Through the project, the student will gain practical experience in cancer genomics, next-generation and long-read sequencing, tumour–normal analysis, genomic data formats such as BAM and VCF, Linux-based computing, scripting in Python and/or R, reproducible bioinformatics workflow development, data visualisation, and interpretation of genomic results. Importantly, the student will also develop experience in evaluating the accuracy, limitations and potential clinical utility of computational genomics methods.

2027 UNDERGRADUATE STUDENT PROJECTS

COMPUTATIONAL BIOLOGY

Project 4

Project Title:	Knowledge Extraction to Support AI Reasoning in Childhood Cancer Drug Repurposing
Supervisor:	Ah-Jung Jeon AJeon@ccia.org.au
Suitable for:	Undergraduate students 8 weeks Biological Sciences; Other (Artificial Intelligence)
Available:	Term 1 2027; Summer 2026-2027

Project outline:

Background and aims

Background

Medical research produces far more publications than researchers can read, and important knowledge is scattered across thousands of papers. AI models can reason over this knowledge to suggest new research ideas, but they work best with well-organised, reliable information rather than raw text.

The Health Data and AI team at Children's Cancer Institute has built a pipeline that extracts assertions from publications. An assertion is a short, structured statement of a finding, linked to its supporting evidence. The pipeline organises these assertions into a knowledge graph, a network connecting drugs, genes and cancer types, which an AI agent can then draw on as background knowledge.

A key use of this pipeline is drug repurposing, which means finding new uses for existing approved drugs. This is valuable in childhood cancer, where few drugs are developed specifically for children. Two things currently limit the pipeline. Processing full papers is costly, because AI models charge by the amount of text they read. And AI-extracted assertions can contain errors, so they need to be checked.

Aims

This project aims to make the pipeline a reliable, cost-effective source of knowledge for AI reasoning in childhood cancer drug repurposing, by:

- Extending the assertion format to capture drug repurposing information, such as the drug's target, how it works, and the type of evidence.
- Testing whether reading only selected sections of open-access papers reduces cost without losing accuracy, compared with reading the full text.
- Extend the existing web interface so that paper co-authors can review extracted assertions, and assertions are marked as "validated" or "draft".

Approach

The student will work with the Health Data and AI team's existing pipeline and web interface, using a focused set of open-access papers on one childhood cancer type drawn from text and data mining (TDM) output. All AI processing will run within a fixed budget.

Weeks 1–2: Orientation and design. The student will learn how the pipeline works, review background reading on drug repurposing and knowledge extraction, and select the paper set. With supervisors, they will define the additional information each assertion should capture for drug repurposing, such as the drug's target, how it works, the cancer type and the type of evidence.

Weeks 3–5: Cost-reduction experiments. The student will first process a small set of papers in full to create a baseline. They will then test cheaper approaches that read only selected sections, such as the abstract, results and discussion. Each approach will be compared with the baseline on cost, number of assertions extracted and accuracy.

Weeks 4–6: Author review. The student will extend the existing web interface so that co-authors can review assertions from their own papers and mark them as validated or draft. They will pilot it with Children's Cancer Institute researchers. The reviewed assertions will also serve as the reference for checking accuracy in the cost experiments.

2027 UNDERGRADUATE STUDENT PROJECTS

Weeks 7–8: Analysis and reporting. The student will analyse the results, recommend the most cost-effective extraction approach, and present their findings and a demonstration to the team.

Expected outcomes and skills acquired

Expected outcomes

By the end of the project, the student is expected to deliver:

- An extended assertion format that captures the information needed for drug repurposing.
- A comparison of extraction approaches on cost and accuracy, with a recommended approach for future use.
- An extended web interface that supports co-author review, together with an initial set of validated assertions.
- A short written report and a presentation with a demonstration to the team.

Skills acquired

On the technical side, the student will gain hands-on experience with large language models (LLMs) and their programming interfaces. This includes structured data extraction from text, knowledge graphs, web interface development, and designing and running experiments that measure both cost and accuracy. They will also develop domain knowledge in drug repurposing, childhood cancer research and reading scientific literature. Professionally, they will build experience in working within a multidisciplinary research team, documenting their work, presenting results, and applying responsible AI practice, including how to check AI-generated outputs before they are relied on.