Thesis with Ritz Fall 2024 Syllabus

Bio470: Thesis with Anna Ritz

Last Updated September 2024 for students starting thesis in Fall 2024 or Spring 2025. Note that the details in this syllabus pertain only to students mentored by Anna Ritz; other faculty might have different guidance and requirements for thesis.

The CompBio lab uses computational tools to analyze patterns in large biological datasets; this can take thesis projects in a multitude of directions. While a programming background is useful for some projects, other topics require no previous computer science experience (though you might learn some along the way). My main research uses molecular interaction networks, represented as graphs, to study intracellular signaling pathways and the ways in which they are disrupted in disease. Thesis projects have studied the outputs of these tools to prioritize genes that may be associated with diseases, and have developed new ways of investigating these networks.

- Anna Ritz's Website
- CompBio Research Blog

Instructor Details

Instructor: Anna Ritz (she/her Biology 200B)

General Office Hours: Wednesday 1-2:30 and Friday 11-noon

Thesis & Group Meetings: TBD

Departmental Deadlines and Instructions are on the Moodle

Thesis Description & Goals

The College describes the Reed thesis as "your opportunity to explore a problem or answer a question that holds particular significance to you." This is manifested as a year-long, two-unit course where you pursue a biological question relevant to your major.

The nature of the thesis project varies from student to student, as this is the most individualized course offered at Reed. In some cases, you will learn how to plan,

execute, and interpret scientific experiments. In other cases, you will learn how to synthesize existing research to investigate and contextualize a new question. In all cases, you will learn how to **obtain and deepen your understanding of biological phenomena or processes** and how to **communicate the scientific knowledge** you gain, orally and in writing.

The department has put together learning outcomes for thesis (also found on Moodle). Upon completing a senior thesis in Biology as student will have demonstrated the ability to:

- Choose and define important and contemporary topics of inquiry from the major field
- Independently investigate that topic with the support of an advisor
- Develop new knowledge and/or ideas
- Apply a wide variety of skills learned in their coursework including active involvement in the development of hypotheses and in experimental design, methodological refinement, analysis and interpretation of data, discussion and criticism of research findings, and formal presentation of their work
- Apply ethical standards to research
- Write a coherent document that is substantially longer than a traditional term paper or project
- Defend their work orally to scientific and non-scientific audiences

Thesis Evaluation

There are three parts to the thesis evaluation. We will work on all of these parts consistently throughout the year.

1. Engagement and progress with the project chosen/developed. We will have regular check-in meetings throughout the year.
2. The thesis document: a comprehensive summary of the thesis work, including a substantial background section with citations. You will write pieces of the thesis draft throughout the year.
3. An oral examination with a board of four faculty readers in accordance with the guidelines of your major, which occurs after the written thesis is submitted. You will practice communicating with others about your project throughout the year.

The department has also written a statement about the thesis evaluation - this can be found on Moodle.

A Note on Co-Advised Theses and Thesis Scope

In some cases, your thesis may be advised by multiple faculty. Any agreements made by all co-advisors, even if they deviate from this syllabus, are allowed. For example, attending all meetings for all advisors may be untenable, or advisors may have different expectations in terms of internal deadlines. Communication

plans may differ when collaborating with multiple advisors. When in doubt, defer to your collective advisors' advice over this document.

Make sure that your thesis satisfies the expectations of your major. BMB/Neuroscience/ES-Bio majors may want to confirm with a member of the relevant committee that their thesis falls within the expectations for the major. I encourage you to check in with the appropriate committee at least once before the thesis is due to make sure that your thesis is in scope.

Thesis Logistics & Support

Meetings

Constant communication is crucial for both student and advisors to be on the same page. Therefore, we have organized weekly meetings to stay up to date on progress.

Weekly Group Meetings. We will have weekly group meetings in the computational biology lab (B203). Group meetings include thesis students and other researchers in the CompBio lab, and are an opportunity to provide updates and solicit advice about your project at different stages throughout the year.

Weekly Individual Meetings. We will also have time set aside for individual meetings with each thesis student (ideally with any co-advisors). These individual meetings are *your time* to get guidance on any aspect of the thesis project. Before the meeting, spend some time organizing your thoughts about (a) what you have accomplished since the last meeting, (b) any obstacles that you have come across, and (c) what your next steps could be. Individual meetings may need to be postponed due to other conflicts, and can also be replaced by written communication about the three items above if you have a good sense about where to go next.

Communication & Technology

Official communications about thesis deadlines and oral examinations will come from the Office of the Registrar and the Biology Department emails. These will also be posted on the 2024-25 Thesis Moodle page.

All thesis students will join the CompBio Slack channel, and we will have informal communication through Slack. The **required logistics for a thesis project** are that: 1. You use the Reed Thesis template. 2. You decide on a way to back up your work (for your thesis document and any datasets/analyses). 3. You keep track of your daily and weekly work towards your thesis, in any format you choose. 4. You communicate consistently with me and other co-advisors throughout the entire process.

When considering other technology you will use, my rule of thumb is to follow a path of least resistance. This may mean writing programs in a language you

are familiar with (e.g. R, Python,...), generating figures with software that you are familiar with (e.g., R, Python, JMP, Excel,...), and using a text editor that you are familiar with (e.g. Word, LaTeX,...). This is a great time to learn something new, and if you explicitly want to learn a new technology or programming language, we will work that into the goals of your thesis. Keep in mind, though, that if you deviate far from your advisors' expertise, we have limited means to help you.

I strongly suggest using **GitHub** to maintain any code you will write as part of your thesis. You can make a private repository and add me as a collaborator - this will help with code reviews and helping with bugs.

Use of generative AI

Generative AI is technology that is trained to generate text, images, or code from natural language prompts. Two examples of generative AI that you might have seen before are ChatGPT (which returns text based on prompts) and GitHub Copilot (which returns code based on prompts).

Everyone must write a thesis document. This document represents *your* work, *your* results, and *your* conclusions. Further, it can be written in *your* style, which might be as formal as a scientific paper. When you write papers, you are claiming the words and the ideas within them as your own. In order to demonstrate the learning objective that you have writen a coherent document, it needs to be in your words. As a result, you may not use generative AI to compose all or part of your thesis document. Generative AI is acceptable to use to help organize your thoughts in the planning/outlining phase (you will need to acknowledge this in the thesis). You should also be using spellcheck software, though this doesn't need to be ChatGPT.

If you use generative AI for *any* part of your thesis (even for planning/organization), this needs to be clearly acknowledged in the thesis document.

Using others' code and tools

If you use any code you found online in your own programs (such as from StackOverflow, or other sites), you must acknowledge this in your code - include the author, a brief description, and the URL.

If you use tools from other groups, you will also acknowledge these in your thesis. Typically, tools have an accompanying paper that you should cite.

If you have any questions, ask me.

Thesis Support

As your thesis advisor, one of my jobs is also to put you in touch with others who can support you through the thesis process. One of the most important

groups are your peers doing thesis alongside you! The other computational biology thesis students (and other thesis students in general) can be a wonderful resource.

Postbac Mentors

This year, we are lucky to have two post-bac researchers working in the compbio lab: Oliver Anderson and Altaf Barelvi. They will attend our group meetings to learn about your thesis project, and they will be available to meet with you during the year for help.

Thesis Funding and Borrowing Equipment

Computational thesis projects rarely need a budget for supplies. However, several possible sources of funding exist for thesis project supplies, the main mechanism being the Biology Undergraduate Research Project (BURP) grants to support thesis work offered through the Biology Department. BURP grants may fund licenses for software (e.g. Matlab) or access to databases if they are not publicly available.

There is also a Student Technology Equipment Program (STEP) program where seniors doing thesis can apply to borrow equipment from Computer User Services.

Thesis Formatting

The Help Desk will hold drop-in thesis formatting hours, resources for templates, etc. - check out their page.

Academic Accommodations

Thesis is a course, and like all Reed courses I have a commitment to accommodating any documented needs you may have. Please discuss any documented accommodations with me, especially those that affect one or more of the three-part evaluation for thesis. Please let me know if there are useful supplements to thesis meetings (videos, large-font texts, captioning, etc.). Disability Accommodation Notification Letters can be obtained from Disability & Accessibility Resources (DAR). All discussions will remain confidential.

Disruptions to the Academic Year

Many students will have personal circumstances that may affect their performance this year. In fact, it is likely that we *all* have been and will be deeply affected by events locally, nationally, and globally. 1. You are encouraged to seek guidance from me and your co-advisors on any anticipated or realized issues. We will work with you to make adjustments to milestones as needed. 2. Mental and physical health is more important than your thesis project, though it may not always feel that way. If you miss multiple meetings, I will reach out to make sure you are okay.

Diversity and Inclusion

Historically, biological discoveries and advances in computer science have been dominated by privileged voices, namely those of white men. Computational biology, while a more recent field, also lacks diversity along many important axes (including race, gender, nationality, class, sexuality, religion, ability, etc.). As we proceed with thesis, I acknowledge my own bias that stems from systemic privilege in academia, and will aim to make the resources I offer more inclusive. I welcome any other suggestions to make thesis a more inclusive learning process for Reed students.

Land Acknowledgement

As we come together to learn at Reed College, we acknowledge that the territory on which Reed stands is that of Indigenous and Native peoples. The Portland Metro area rests on traditional village sites of the Multnomah, Wasco, Cowlitz, Kathlamet, Clackamas, Bands of Chinook, Tualatin, Kalapuya, Molalla, and many other tribes who made their homes along the Columbia River (Portland Indian Leaders Roundtable, 2018). Native lands have been taken through hundreds of breaches of treaty laws, and in 1953 more then sixty tribes in Oregon were terminated by the federal government which removed millions of acres of land from Indigenous stewardship (The Native American Community in Multnomah County: An Unsettling Profile). This statement is one small step in acknowledging the history that brought us to reside on this land. Portland also has a vibrant and diverse Indigenous community and has the ninth largest urban Native American population in the US. We honor the present Indigenous communities in the area. Visit native-land.ca to explore an interactive map of Indigenous lands.

Milestones

Time management is crucial for a successful thesis. You will work closely with me and your co-advisor(s) to establish a set of internal deadlines in order to make steady progress on your thesis. We will start from this working set of milestones, which are set by quarter.

	Reading Goals	Writing Goals	Experiment Goals
Fall Q1	Clarify Research Question Collect & Read Main Papers	Summarize Related Work Outline Methods	Identify Data Summarize Data

	Reading Goals	Writing Goals	Experiment Goals
Fall Q2	Identify Supporting Papers	Draft Background Chapter Outline Methods Chapter	Reproduce Paper Results or Prototype New Ideas
Spring Q1	Read Theses & Papers for Formatting Examples	Write Methods Chapter Outline Main Results	Implement New Ideas & Analyze Results
Spring Q2	Find Papers that Support/Refute Findings	Write Results & Discussion Revisit All Chapters	Visualize Results

Deadlines (Fall 2024 - Spring 2025)

You will submit thesis proposals in the middle of Week 2 and be assigned thesis advisors by the beginning of Week 3; refer to the Moodle page for more information (link TBD).

We will start from these working set of deadlines for students starting thesis in the Fall of 2024. Deadlines may shift and they may be tailored to each project. We will also have smaller milestones for group meeting activities.

Deadline	Milestone	Description
Fall Week 4 (9/27)	Research Question	A document describing (a) a brief introduction to the topic, (b) the research question, and (c) why the question is important to study. Include at least two citations.

Deadline	Milestone	Description
Fall Week 7 (10/18) *Fall Break*	Annotated Bibliography	A document with a list of at least ten papers that are relevant to your project. For each paper, write one paragraph that (a) summarizes the main contributions of the paper and (b) explains why the paper is relevant to your question.
Fall Week 11 (11/22)	Methods Outline & Updated Bibliography	An updated annotated bibliography with at least 15 papers and a document with a brief description of your methods. The methods outline can take many forms - it can be a list of steps you plan to do, pseudocode for existing or new methods, or instructions for running a tool.
Fall Week 14 (12/11) *Winter Break*	Reflection	A document that describes (a) your current progress in thesis, (b) your planned progress for spring semester, and (c) your immediate next steps. Be sure to address your progress in **writing**, your progress in **reading**, and your progress in **experiments/analysis**. This is an opportunity to re-evaluate the thesis question or goal, and adjust if necessary.

Deadline	Milestone	Description
Spring Week 4 (2/21)	Background & Methods Chapters in Thesis Template	**Move all of your existing drafts to one of the thesis templates.** Your thesis should have complete Background & Methods chapter drafts and ideally an outline of your results chapter.
Spring Week 8 (3/21) *Spring Break*	Results & Discussion Outlines/Drafts	Outline your results in the thesis template, and add any results you currently have. Outline the main points you want to discuss in your Discussion chapter.
Spring Week 11 (4/18)	Full Draft	Your thesis draft should have full drafts of all chapters. **This is the deadline to guarantee Anna's feedback - any writing after this date might not get edited.**
Spring Week 13 (5/2)	Thesis Due	Submit your approved thesis to the registrar before 3pm.

Feedback

After submitting each milestone, you will receive feedback from me within one week in the form of edits directly on your document and, in some cases, a summary of your progress so far.

You can submit drafts for me to review at any time! I will aim to get you edits within two weeks if you submit additional writing. Please send me reminders if you don't hear from me in these off-deadline editing requests.

As we get further into spring semester, I will focus on giving you feedback for parts that I have not read yet (rather than feedback on parts I've already edited). Note that your thesis must be approved by me before being submitted to the registrar, and I will require reviewing certain parts of the document (the Discussion, Results, etc.) before giving that approval.

Previous Theses

These theses are all available in the Thesis Tower in the library; a subset is also available in the Electronic Thesis Archive and in my office.

Student	Co-Advisor	Title
Aliya Ghassaei '24		Leveraging Comparative Graph Theory to Model Gene Co-expression Data
Ainsley Norman '24	Jeremy Coate	Seurat & Suffering: Identification of Endoreduplication Using Computational Methods in Soybean Nodules
Caroline Spiggle '24		Navigating Neurodevelopmental Biology via Computational Genomics: An Accessible Perspective on ADHD & Autism
Zaida Vinaimont '24		Gene Regulatory Network Inference for Subpopulations of Quiescent Müller Glia
Tommy Yoon '24		A Higher-Order Network Analysis of Molecular Dynamics Simulations
Frank Zhuang '23	Kara Cerveny	A RARE Catch: The Potential of Computational Algorithms in Aiding Discovery of Novel Genetic Regulation Interactions
Ananke Krishnan '22	Julia Michaels	Impacts of Backyard Habitat Creation for Birds at Single-Yard and City-Wide Scales

Student	Co-Advisor	Title
Ingrid Zoll '22	Aaron Ramirez	Tree Health from Space: Modeling Urban Tree Health using Multispectral Satellite Imagery in Portland, OR.
Delaney Brubaker '21	Julia Michaels	Can we build a framework for the future of sustainable agriculture knowledge?
Tayla Isensee '21	Erik Zornik	Sex-biased conserved gene calling expansion with orthogroups.
Jiarong Li '21	Jim Fix	Complex network comparison using graphlets.
Hannah Meier '21	Sam Fey	Phytoplankton on the Go! Understanding the effect of thermal acclimation on the movement strategies of the green algae *Chlamydomonas reinhardtii.*
Aryeh Stahl '21	David Ramirez	A generalized approach to graph de-anonymization.
Gabe Preising '20	Suzy Renn	A computational network approach to examine biological crosstalk during mouthbrooding in the cichlid Astatotilapia burtoni
Sol Taylor-Brill '20	Jay Mellies	Computationally analyzing the effect of cannabidiol on Enterohemorrhagic E. coli using transcriptomic data.
Madeline Doak '20	Erik Zornik	Sex-biased gene expression in the brain conserved among taxa

Student	Co-Advisor	Title
Alexander King '19		Multi-omic analysis of genetically induced conditions
Miriam Bern '19	Jeremy Coate	A comparison of two methods for identifying alternative splicing in diploid and tetraploid Arabidopsis thaliana strains
Amy Rose Lazarte '19	Sam Fey	Just keep swimming: resolving the patterns and dynamics of body temperature in freshwater phytoplankton
Ananthan Nambiar '19	Mark Hopkins	Computing the language of life
Petra Wijngaard '18	Derek Applewhite	How Cytoscouts got Shot out of Dr. Ritz's new interactome
Moira Differding '17		Your microbes are super models: using microbiome data to reproducibly predict Crohn's Disease
Olive Franzese '17	Adam Groce	Examining the practicality of shortest hyperpaths for signaling pathway analysis: the Cheating Hyperpath Algorithm as an alternative approach
Olivier Ezell '16		Pathway models of signal dysregulation in schizophrenia
Cameron Shard Milne '16	Sarah Schaack	Identification of transposable element subfamilies using a phylogenetic network approach
Barney Potter '16	Jim Fix	Prize-collecting Steiner trees in directed signaling hypergraphs

Thesis Project Ideas (2024-25)

This is a list of example thesis topics – new ideas are welcome. If you want to get a broad sense of computational network biology, please read Current and Future Directions in Network Biology.

You don't need to have computer science experience to do a computational thesis! Some thesis students use this as an opportunity to learn new skills; others have taken deep dives into interpreting the results of current tools. I have also mentored expository/library theses, which synthesize big complex ideas, and meta-analyses, which collect data from multiple papers and statistically analyze the aggregate findings.

PPI Networks and Signaling Pathways

Protein-protein interaction networks have been extremely useful for identifying and analyzing signaling pathways from experimental data. We model protein protein interactions as a graph, where nodes denote proteins and edges denote physical protein interactions. We have software to connect proteins of interest within this network in many different ways, but each method requires different interpretations of the connections. Many of these projects are inspired by the Signaling Pathway Analysis Streamliner (SPRAS) Project, in collaboration with Tony Gitter at UW-Madison.

1. **Integrate disease data to identify signaling pathway disruption.** We have existing algorithms to reconstruct signaling pathways. How do we use cancer data (or other disease data) to identify dysregulated signaling?

2. **Benchmark landmark papers with new methods/data.** A number of important papers in network biology have helped shape the field; however, we now have more complete datasets and different tools to apply to the problems. Do the findings of these landmark papers hold up in light of new information and methodologies?
3. **Investigate pathways involved in cell-cell fusion.** Cells fuse together as part of fundamental processes, yet we still don't know all of the proteins that regulate this process. This project will apply existing graph algorithms to study proteins that may regulate cell-cell fusion. (Collaboration with Derek Applewhite)
4. **Investigate regulators of manganese transport.** Maintaining a proper balance of metal ions is an extremely important cellular process. We have done some work to identify how manganese is regulated within bacterial cells through a combination of transcription factor binding motifs, network analysis, and transcriptomic analysis. Tommy Yoon '24 most recently worked on this project in the summer of 2023 and wrote about it here. Can we integrate these datasets to better understand potential proteins and mechanisms of regulation in *Bacillus subtillis*? (Collaboration with Shivani Ahuja).

Adding Context to PPI Networks

We have recently done work to help researchers explore molecular networks, and provided additional information to place their questions in the context of physical and regulatory interactions. These projects are inspired by ProteinWeaver, a network visualization tool developed by post-bacs in the group.

1. **Predict protein function from the Gene Ontology.** The Gene Ontology is a categorization of all protein function we know of; however, it is not complete. ProteinWeaver has connected the Gene Ontology to physical and regulatory interactions among proteins, and we have an initial method to predict a protein's function from this information. Can we add additional context (such as protein domain information, sequence similarity, predictd 3D structure) to improve these predictions?
2. **Investigate the biological meaning behind network motifs.** Motifs are small subgraphs that can be found within larger networks. We have found that signaling pathways and disease modules are enriched for certain subgraphs. More recently, we've categorized motifs that are comprised of both physical and regulatory interactions within ProteinWeaver. Do these enriched patterns have any biological interpretation?
3. **Evaluate physical, regulatory, and combined interactions for protein context across taxa.** ProteinWeaver's network can be used as a PPI network, a gene regulatory network, or a combination of physical and regulatory interactions. How do these three networks affect the interpretation for a protein's relevance, and how does this change for different species?

Graphs and Graph Generalizations

Some of these topics may be good for someone who has taken multiple CS courses.

1. **Define a mathematical objective function for signaling pathway reconstruction.** Many algorithms for signaling pathway reconstruction do not directly optimize a function of the graph; and those that do are missing a parameter to increase the size of the subnetwork. We have developed a first attempt to define an objective function for signaling pathway reconstruction (preprint), but can we improve it?
2. **Random walks on directed hypergraphs.** We have done work to develop shortest path and connectivity algorithms on directed hypergraphs (where edges can capture many-to-many relationships). Random walks are useful for identifying node relevance within a graph structure; can we apply them to directed hypergraphs (where an edge captures many-to-many relationships)?

3. **Applications for directed hypergraphs.** Directed hypergraphs have been useful for signaling pathway representations, and they hold promise

for other biological networks. This would be an excellent survey of potential use cases for hypergraphs and other graph generalizations for computational biology.

Other Types of Biological Networks

I have mentored theses on networks relevant for ecology, animal behavior, and neuroscience; I am open to talking about potential projects in these realms. We have also developed Graphery, an interactive biological network tutorial with real-world examples across biological scales. Having an idea of the available data for thesis is especially important for these projects, which are outside my main area of expertise.