

---

# University of Wisconsin-Madison

## Spring 2024

ONCOLOGY 778 Bioinformatics for Biologists

Tuesday/Thursday 1:00 - 2:15 PM 6571 WIMR II

---

**Credits:** 3

**Instructional Mode:** In person

**Instructors:** Eric Johannsen, M.D., WIMR 6531, ejohannsen@medicine.wisc.edu, Phone: 262-9952  
Mitch Hayes, WIMR 6559, mhayes1@oncology.wisc.edu, Phone: 262-6697

**Office Hours:** Mondays 1p-2p (after Problem Set distribution) or by appointment.

**Description:** Familiarizes students with bioinformatics theory and principles. Provides students with real-world experience that they can apply to their own work. Provides a foundation of knowledge that students can use to critically evaluate existing bioinformatics tools that can be used in their work, and in the absence of an appropriate tool, identify the analyses that demand the development of novel tools.

**Format:** This course will meet on TU/TH for 75 minutes from 1:00-2:15 PM. Class sessions involve traditional lecture interspersed with class discussion. Students will be assessed on the following: class participation (20 points), homework (140 points), and their final project (40 points).

**Lecture Topics:** Introduction to Bioinformatics, Statistics, Algorithms; Nucleic Acid Alignment; NextGen Sequencing Analysis; Proteomic Analysis; and Motif Search.

**Textbook (Suggested; available online through UW Libraries):\*** An Introduction to Bioinformatics Algorithms (N. C. Jones and P.A. Pevzner), ANE Books, Cambridge, MA (2004).

**Requisites:** Graduate/Professional Standing and consent of instructor.

---

\*This text provides additional context for the algorithmic underpinnings of the bioinformatics tools used in this course. Available online: <http://go.wisc.edu/8gy576>

---

## Course Outline:

<b>Lecture 1 - Introduction</b> .....	1/23
Objectives and Overview	
<b>Practical 1 - Preliminaries</b> .....	1/25
Computer setup, introduction to data structures and flow control	
<b>Part I - Bioinformatics of DNA/RNA</b>	
<b>Lecture 2 - Global Alignment</b> .....	1/30
Scoring schemes, mismatch penalties, dynamic programming	
Needleman-Wunch algorithm	
<b>Practical 2 - Global Alignment Exercises</b> .....	2/01
Solving global alignment problems by hand and with computational tools	
<b>Lecture 3 - Local Alignment</b> .....	2/06
Smith-Waterman Algorithm, Basic Local Alignment Search Tool (BLAST), MegaBLAST	
<b>Practical 3 - Contig Assembly</b> .....	2/08
Sanger Sequencing, BLAST, pLannotate	
<b>Lecture 4 - Introduction to Next Generation Sequencing (NGS)</b> .....	2/13
Next-Gen sequencing platforms, Quality scores, data formats (e.g., FASTQ)	
<b>Practical 4 - NGS Basics</b> .....	2/15
NGS Workflows; File Formats; SRA; Galaxy; FASTQC	
<b>Lecture 5 - Short Read Alignment DNA</b> .....	2/20
Suffix Trees and other indices, Aligners (e.g., STAR), Mismatch handling, SAM format	
<b>Practical 5A - DNA Short Read Alignment (I)</b> .....	2/22
Indices, Unaligned Reads, Alignment, UCSC Genome Browser	
<b>Practical 5B - DNA Short Read Alignment (II)</b> .....	2/27
SNPs/Variants; Alignment-Free Methods	
<b>Lecture 6 - Short Read Alignment RNA (transcriptomes)</b> .....	2/29
Strandedness and Libraries, Transcriptomes (GTF Format), Two-pass methods, CIGAR strings	
<b>Practical 6 - RNA Short Read Alignment</b> .....	3/5
Indices, handling unmapped reads, CIGAR strings	
<b>Lecture 7 - Differential Expression Analysis</b> .....	3/7
DESeq2 and GSEA	
<b>Practical 7A - Assigning Reads to Genes</b> .....	3/12
Processing STAR output, featureCounts, DESeq2 (Galaxy)	
<b>Practical 7B - Differential Expression Analysis</b> .....	3/14
DESeq2 (Local), GSEA	

<b>Lecture 8 - ChIP-Seq Analysis</b> .....	3/19
Peak Calling, Poisson distribution, Expectancy Maximization (EM), Motif searches	
<b>Practical 8 - ChIP-Seq Analysis Exercises</b> .....	3/21
MACS; Visualization of ChIP-Seq tracks on UCSC Genome Browser	
<b>Spring Break - 03/25-03/29</b>	
<b>Part II - Bioinformatics of Proteins</b>	
<b>Lecture 9 - Protein Alignment</b> .....	4/2
Matrices (BLOSUM, PAM), BLASTP (+TBLASTN, BLASTX, & TBLASTX), synonymous/non-synonymous mutations	
<b>Practical 9 - Protein Alignment Exercises</b> .....	4/4
BLOSUM, BLASTP	
<b>Lecture 10 - Multiple Sequence Alignment</b> .....	4/9
Distance Trees, Progressive Alignments (CLUSTAL), MSA applications: PsiBLAST	
<b>Practical 10 - Multiple Alignment Exercises</b> .....	4/11
Multiple alignment of fluorescent proteins, Plasmid annotation	
<b>Lecture 11 - Protein Structure</b> .....	4/16
Primary vs. Secondary vs. Tertiary, Secondary Prediction (HMM), Tertiary Prediction (Threading)	
<b>Practical 11 - Web-based tools for protein structure analysis</b> .....	4/18
TM prediction, domain identification, and threading	
<b>Part III - Final Projects</b>	
<b>Student Project Presentations</b> [4/23, 4/25, 4/30, 5/2]	

### Learning Outcomes:

1. Identify the appropriate analysis tools for common bioinformatics problems, format data, execute the analysis, and adjust necessary parameters.
2. Interpret results of common bioinformatic analyses including their own results as well as the published works of others.
3. Demonstrate competence in bioinformatic analysis by conducting an original analysis of primary data from their own lab or public databases.
4. Understand the assumptions and principles of bioinformatic pipelines and their implications for validity and statistical significance.

**Grade Policy:** The final grade will be determined by the percentage earned of total points available for the course.

The final grade scale is:

Graduate Students

A 186-200 points

AB 176-185 points

- 
- B 166-175 points
  - BC 156-165 points
  - C 146-155 points
  - D 130-145 points
  - F 0-129 points

**Grading:** Component breakdowns are as follows:

- 10% Participation
- 70% Problem Sets
- 20% Final Project

**Problem Sets:** Problem Sets will be distributed to students throughout the semester. These are intended to provide hands-on experience in the evaluation and application of bioinformatics algorithms and tools. Some problems will require the use of bioinformatics software (both locally installed on students' computers and cloud-based tools) to solve real-world problems.

Problem Set 1 Distributed .....	Feb 6
Due .....	Feb 13
Problem Set 2 Distributed .....	Feb 20
Due .....	Feb 27
Problem Set 3 Distributed .....	March 5
Due .....	March 12
Problem Set 4 Distributed .....	March 19
Due .....	April 2
Problem Set 5 Distributed .....	April 9
Due .....	April 16

**Final Project:** Problem Sets will be distributed to students throughout the semester. These are intended to provide hands-on experience in the evaluation and application of bioinformatics algorithms and tools. Some problems will require the use of bioinformatics software (both locally installed on students' computers and cloud-based tools) to solve real-world problems.

**Guidelines for Evaluating Participation:** During the final four class sessions, each student will give a 20 minute presentation on an original bioinformatic project building on one or more concepts covered in the course, ideally relating to their thesis work. An example might include analysis of RNA-seq data from their lab or a public database. Students are expected to focus primarily on the bioinformatics principles and pitfalls encountered in this analysis rather than the conclusions of the research per se.

1. Outstanding Contributor: Contributions in class reflect exceptional preparation. Ideas offered are always substantive; provide one or more major insights as well as direction for the class. Challenges are well substantiated and persuasively presented. If this person were not a member of the class, the quality of discussion would be diminished markedly. (Outstanding contributors will receive full credit = 20 points.)

- 
2. **Good Contributor:** Contributions in class reflect thorough preparation. Ideas offered are usually substantive; provide good insights and sometimes direction for the class. Challenges are well substantiated and often persuasive. If this person were not a member of the class, the quality of discussion would be diminished. (Good contributors will receive 15 out of 20 points.)
  3. **Adequate Contributor:** Contributions in class reflect satisfactory preparation. Ideas offered are sometimes substantive, provide generally useful insights but seldom offer a new direction for the discussion. Challenges are sometimes presented, fairly well substantiated, and are sometimes persuasive. If this person were not a member of the class, the quality of discussion would be diminished somewhat. (Adequate contributors will receive 10-14.5 out of 20 points.)
  4. **Non-Participant:** This person says little or nothing in class. Hence, there is not an adequate basis for evaluation. If this person were not a member of the class, the quality of discussion would not be changed. (Non-participants will receive 5.5- 9.5 out of 20 points.)
  5. **Unsatisfactory Contributor:** Contributions in class reflect inadequate preparation. Ideas offered are seldom substantive; provide few if any insights and never a constructive direction for the class. Integrative comments and effective challenges are absent. If this person were not a member of the class, valuable air-time would be saved. (Unsatisfactory contributors will receive 0-5.0 out of 20 points.)

**Academic Integrity:** By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion. <https://conduct.students.wisc.edu/syllabus-statement/>

**Accommodations for Students with Disabilities:** The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. <https://mcburney.wisc.edu/instructor/>

**Diversity and Inclusion:** Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

---

The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world. <https://diversity.wisc.edu/>

**Artificial Intelligence:** AI tools (ChatGPT, Bard, etc.) may not be used for Problem Sets or Projects without prior approval of the instructors.

**Important Dates:**

First Session ..... Jan 23  
Last Session ..... May 2