IMPORTANT: This syllabus form should be submitted to OAA (gsbs_academic_affairs@uth.tmc.edu) a week before the start of each semester.

NOTE to STUDENTS: If you need any accommodations related to attending/enrolling in this course, please contact one of the Graduate School's 504 Coordinators, Cheryl Spitzenberger or Natalie Sirisaengtaksin. We ask that you notify GSBS in advance (preferably at least 3 days before the start of the semester) so we can make appropriate arrangements.

Term and Year: **Spring 2025**

Course Number and Course Title:
GS01 1143: Introduction to Bioinformatics

Credit Hours: **3**

Meeting Location: **UT MD Anderson Building**

Building/Room#: **GSBS Computer Lab**
BSRB S3.8112

Program Required Course: Yes

Approval Code: Yes

(If yes, the Course Director or the Course Designee will provide the approval code.)

Audit Permitted: **Yes**

Classes Begin: **January. 14, 2025**

Classes End: **May 2, 2025**

Final Exam Week: **May 5-9, 2025**

Class Meeting Schedule

Day	Time
Tuesday	**1:00-4:00 pm**

Course Director

Name and Degree: **Traver Hart, PhD**

Title: **Associate Professor**

Department: **System Biology**

Institution: **MDACC**

Email Address: GHart1@mdanderson.org

Contact Number: **713-794-4946**

Course Co-Director:

Name and Degree: **Suprateek Kundu, PhD**

Title: **Associate Professor**

Department: **Biostatistics**

Institution: **MDACC**

Email Address: SKundu2@mdanderson.org

Instructor/s

1. **Traver Hart, PhD**
   Institution: MDACC
   Email Address: GHart1@mdanderson.org

2. **Suprateek Kundu, PhD**
   Institution: MDCC
   Email Address : SKundu2@mdanderson.org

3. **Ken Chen, PhD**
   Institution: MDACC
   Email Address : KChen3@mdanderson.org

4. **Nick Navin**
   Institution: MDACC
   Email Address : NNavin@mdanderson.org

Contact Number: 877-555-1235

NOTE: Office hours are available by request. Please email me to arrange a time to meet.

Teaching Assistant:

Yitao Tang
Yitao.Tang@uth.tmc.edu

5. **Lulu Shang, PhD**

   Institution: MDACC

   Email Address : LShang@mdanderson.org

6. **Yiwen Chen PhD**

   Institution: MDACC

   Email Address: YChen26@mdanderson.org

7. **Han Liang, PhD**

   Institution: MDACC

   Email Address: HLiang1@mdanderson.org

8. **Rehan Akbani, PhD**

   Institution: MDACC

   Email Address: RAkbani@mdanderson.org

9. **Kunal Rai, PhD**

   Institution: MDACC

   Email Address: KRai@mdanderson.org

10. **Ye Zheng**

    Institution: MDACC

    Email Address: YZheng8@mdanderson.org

10. **Tsung-Hung Yao**

    Institution: MDACC

    Email Address: TYao@mdanderson.org

Course Description:
This course is intended to be an introduction to concepts and methods in bioinformatics with a focus on analyzing data merging from high throughput experimental pipelines such as next-gen sequencing. Students will be exposed to algorithms and software tools involved in various aspects of data processing and biological interpretation. Though some prior programming experience is highly recommended, it is not a requirement.

Textbook/Supplemental Reading Materials

Optional: Bioinformatics & Functional Genomics 2nd edition (Jonathan Pevsner)

<u>Course Objective/s</u>:

Upon successful completion of this course, students will work with multiple types of bioinformatics datasets, with incorporation of computer programming, statistics and data visualization.

***Specific Learning Objectives*:**

1. A broad exposure to topics and methods in bioinformatics, computational biology and genomics.
2. Learn how to explore and download data from biological databases.
3. Learn the fundamental concepts in statistics.

Student Responsibilities and Expectations:

1. Attend all scheduled classes, labs, and workshop sessions (~2 hours).

2. Active participation in class discussions and activities is essential. Engage with the material through thoughtful questions and contributions to discussions.

3. Submit all assignments and projects by the specified deadlines. Late submissions may result in a grade penalty unless a valid reason is provided.

4. Ensure that all work submitted is your own. Plagiarism or cheating of any kind is unacceptable and will be dealt with severely.

5. Understand the most critical elements from each week needed for the assignments and homeworks.

Students will convene once a week for a two-hour session dedicated to hands-on bioinformatics exercises or in-depth analysis of workshops. The timing of these sessions may vary based on the availability of computational resources and data sets. These sessions are designed to provide real-world experience in handling and analyzing biological data, using various bioinformatics tools and software. When a specific project or case study is undertaken, students may be required to complete additional analysis or write-ups outside of the scheduled hours.

In workshop sessions, students will integrate their hands-on experiences with broader bioinformatics concepts. This integration is crucial for understanding the application of bioinformatics in various biological and clinical contexts. The debriefing sessions will also include discussions on how the week’s activities relate to other areas in the field that students may not have directly explored.

The course emphasizes the elucidation and reinforcement of key concepts and skills acquired during practical exercises. This approach ensures a deeper understanding and retention of the material.

Required Material:
(Optional) Bioinformatics & Functional Genomics 2nd edition (Jonathan Pevsner). Lecture material includes lecture slides and workshop session slides.

Grading System: Letter Grade

Student Assessment and Grading Criteria : *(May include the following:)*

Percentage	**Description**
Homework (90 %)	3 homework assignments (30% each)
Participation and/or Attendance (10%)	

CLASS SCHEDULE: Spring 2025

Week	Date	Chapter	Lecturer	Topics
1	Jan. 14	Overview	Dr. Kundu	• Overview of Genomics and Next-Generation Sequencing Technologies • Human Genomics • Cancer Genomics • Computer set up
2	Jan. 21	Statistics 1	Dr. Shang	• Discrete and Continuous probability distributions • Normal distribution • Mean and Confidence interval • Introduction to R
3	Jan. 28	Statistics 2	Dr. Tsung-Hung Yao (Postdoc)	• Statistical Inference: testing • Multiple testing • Feature Selection
4	Feb 4th	Statistics 3	Dr. Shang	• Principal Component Analysis • Clustering (Hierarchical, k-means)
5	Feb. 11th	UNIX	Dr. Hart	• File system operations • Connecting to Servers • Manipulating Files • Text Editors
6	Feb. 18th	Programming	Dr. Ye Zheng	• String processing • Writing programs and functions • Variables and Data Structures • Logical Operators • File Input-Output
~~7~~	Feb. 25th	NO Class	No Class	----No Class
8	Mar 4th	Cancer Genomics	Dr. K. Chen	• SNV calling • Functional Annotation • Mutational Signature
9	Mar 11th	Spring Break		•

10	Mar. 18th	Functional Cancer Genomics	Dr. Hart	• Genome-scale experiments • false discovery rates • Enrichment tests
11	Mar 25th	Functional Genomics	Dr. Hart	• Enrichment tests, continued
	April 1st	Biological Databases	Dr. Akbani	• cBioportal, Database Resources • COSMIC, Onco-mine and UCSC and TCGA • Direct Database Connections
11	April 8th	Phylogenetics	Dr. Liang	• Overview of phylogenetic tree construction • Various methods (distance, neighbor joining, MP, ML) • Bootstrap analysis
12	Apr. 15th	Single-cell genomics	Dr. Navin	• Seurat • Single cell RNA methods • Single cell DNA methods • Clustering and high-dimensional analysis
14	Apr. 22nd	RNA-seq	Dr. Yiwen Chen	• Pre-processing of RNAseq vs DNAseq • gene expression levels and RPKM identification of fusion transcripts
15	Apr. 29th	Epigenetics	Dr. Rai	• Basic epigenomic methods • ChIP-seq analysis

TH-SK/jal