

**IMPORTANT:** This syllabus form should be submitted to OAA ([gsbs\\_academic\\_affairs@uth.tmc.edu](mailto: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.

<p>Term and Year: <b>Spring 2024</b></p> <p>Course Number and Course Title: <b>GS01 1143: Introduction to Bioinformatics</b></p> <p>Credit Hours: <b>3</b></p> <p>Meeting Location: <b>UT MD Anderson Building</b></p> <p>Building/Room#: <b>GSBS Computer Lab BSRB S3.8112</b></p>	<p><b>Program Required Course: Yes</b></p> <p><b>Approval Code: Yes</b> (If yes, the Course Director or the Course Designee will provide the approval code.)</p> <p>Audit Permitted: <b>Yes</b></p> <p>Classes Begin: <b>Jan 9, 2024</b></p> <p>Classes End: <b>April 30, 2024</b></p> <p>Final Exam Week: <b>N/A</b></p>
---	---

**Class Meeting Schedule**

Day	Time
Tuesday	1:00-3:00 pm

<p><b>Course Director</b></p> <p>Name and Degree: <b>Ken Chen, PhD</b></p> <p>Title: Associate Professor</p> <p>Department: Bioinformatics and Computational Biology</p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:KChen3@mdanderson.org">KChen3@mdanderson.org</a></p> <p>Contact Number: 713-794-1348</p> <p><b>Course Co-Director:</b></p> <p>Name and Degree: <b>Nick Navin, PhD</b></p> <p>Title: Associate Professor</p> <p>Department: Genetics</p>	<p><b>Instructor/s</b></p> <p>1. <b>Ken Chen, PhD</b> Institution: MDACC Email Address : <a href="mailto:KChen3@mdanderson.org">KChen3@mdanderson.org</a></p> <p>2. <b>Nick Navin, PhD</b> Institution: MDCC Email Address : <a href="mailto:NNavin@mdanderson.org">NNavin@mdanderson.org</a></p> <p>3. <b>Lulu Shang, PhD</b> Institution: MDACC Email Address : <a href="mailto:lshang@mdanderson.org">lshang@mdanderson.org</a></p>
--	--

<p>Institution: MDACC</p> <p>Email Address: <a href="mailto:NNavin@mdanderson.org">NNavin@mdanderson.org</a></p> <p>Contact Number: 713-563-1287</p> <p><b>NOTE:</b> Office hours are available by request. Please email me to arrange a time to meet.</p> <p><b>Teaching Assistant:</b></p> <p>    <b>Wenhao Li</b> (<a href="mailto:wli91@mdanderson.org">wli91@mdanderson.org</a>)</p> <p>    <b>Haijing Jin</b>(<a href="mailto:hjin3@mdanderson.org">hjin3@mdanderson.org</a>)</p>	<p><b>4. Yiwen Chen PhD</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:YChen26@mdanderson.org">YChen26@mdanderson.org</a></p> <p><b>5. Han Liang, PhD</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:HLiang1@mdanderson.org">HLiang1@mdanderson.org</a></p> <p><b>6. Traver Hart, PhD</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:GHart1@mdanderson.org">GHart1@mdanderson.org</a></p> <p><b>7. Rehan Akbani, PhD</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:RAkbani@mdanderson.org">RAkbani@mdanderson.org</a></p> <p><b>8. Kunal Rai, PhD</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:KRai@mdanderson.org">KRai@mdanderson.org</a></p> <p><b>9. Wenhao Li</b></p> <p>Institution: MDACC</p> <p>Email Address: <a href="mailto:wli91@mdanderson.org">wli91@mdanderson.org</a></p>
---	---

**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 2<sup>nd</sup> edition (Jonathan Pevsner)

**Course Objective/s:**

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:)*

<b>Percentage</b>	<b>Description</b>
Homework (100 %)	3 homework assignments

**CLASS SCHEDULE - Spring 2024****GS01 1143: Introduction to Bioinformatic**

<b>Week</b>	<b>Date</b>	<b>Chapter</b>	<b>Lecturer</b>	<b>Topics</b>
1	Jan. 9	Overview	Dr. Navin	<ul style="list-style-type: none"><li>• Overview of Genomics and Next-Generation Sequencing Technologies</li><li>• Human Genomics</li><li>• Cancer Genomics</li><li>• Computer set up</li></ul>
2	Jan. 16	Statistics 1	Dr. Shang	<ul style="list-style-type: none"><li>• Discrete and Continuous probability distributions</li><li>• Normal distribution</li><li>• Mean and Confidence interval</li><li>• Introduction to R</li></ul>
3	Jan. 23	Statistics 2	Mr. Li	<ul style="list-style-type: none"><li>• Statistical Inference: testing</li><li>• Multiple testing</li><li>• Feature Selection</li></ul>
4	Jan. 30	Statistics 3	Dr. Shang	<ul style="list-style-type: none"><li>• Principal Component Analysis</li><li>• Clustering (Hierarchical, k-means)</li></ul>
5	Feb. 6	UNIX	Dr. Navin	<ul style="list-style-type: none"><li>• File system operations</li><li>• Connecting to Servers</li><li>• Manipulating Files</li><li>• Text Editors</li></ul>
6	Feb. 13	Programming	Dr. Navin	<ul style="list-style-type: none"><li>• String processing</li><li>• Writing programs and functions</li><li>• Variables and Data Structures</li><li>• Logical Operators</li><li>• File Input-Output</li></ul>
7	Feb. 20	Sequence Analysis	Dr. K. Chen	<ul style="list-style-type: none"><li>• Human genome Assembly</li><li>• Sequence Analysis algorithms</li><li>• Genome Browsers</li></ul>
8	Feb. 27	Cancer Genomics	Dr. K. Chen	<ul style="list-style-type: none"><li>• SNV calling</li><li>• Functional Annotation</li><li>• Mutational Signature</li></ul>
9	Mar. 5	Functional Cancer Genomics	Dr. K. Chen	<ul style="list-style-type: none"><li>• CNA Calling</li><li>• SV calling</li><li>• Medical relevance</li></ul>
10	Mar. 12	RNA-seq	Dr. Yiwen Chen	<ul style="list-style-type: none"><li>• Pre-processing of RNAseq vs DNAseq</li><li>• Gene expression levels and RPKM</li></ul>

				<ul style="list-style-type: none"> <li>• Identification of fusion transcripts</li> </ul>
11	Mar. 19	Spring Break		
12	Mar. 26	Phylogenetics	Dr. Liang	<ul style="list-style-type: none"> <li>• Overview of phylogenetic tree construction</li> <li>• Various methods (distance, neighbor joining, MP, ML)</li> <li>• Bootstrap analysis</li> </ul>
13	Apr. 2	Single-cell genomics	Dr. Navin	<ul style="list-style-type: none"> <li>• Seurat</li> <li>• Single cell RNA methods</li> <li>• Single cell DNA methods</li> <li>• Clustering and high-dimensional analysis</li> </ul>
14	Apr. 9	Functional genomics	Dr. Traver Hart	<ul style="list-style-type: none"> <li>• Enrichment tests,</li> <li>• Gene signatures,</li> <li>• GSEA</li> </ul>
15	Apr. 16	Biological Databases	Dr. Akbani	<ul style="list-style-type: none"> <li>• cBioportal, Database Resources</li> <li>• COSMIC, Onco-mine and UCSC and TCGA Direct Database Connections</li> </ul>
16	Apr. 23	Epigenetics	Dr. Rai	<ul style="list-style-type: none"> <li>• Basic epigenomic methods</li> <li>• ChIP-seq analysis</li> </ul>
17	Apr. 30	Final Project?		