جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Course Syllabus: Computational BioSci & Machine Learning - B 390M

Division	Biological and Environmental Sciences & Engineering Division
Course Number	B 390M
Course Title	Computational BioSci & Machine Learning
Academic Semester	Spring
Academic Year	2018/2019
Semester Start Date	01/27/2019
Semester End Date	05/23/2019
Class Schedule (Days & Time)	02:30 PM - 04:00 PM | Mon Thu

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Jesper Nils Tegner	jesper.tegner@kaust.edu.sa	+966128082388 8082388	4232, 2, Ibn Al-Haytham (bldg. 2)	

Teaching Assistant(s)

Name	Email
Manjula Thimma Robert Lehman	manjula.thimma@kaust.edu.sa robert.lehman@kaust.edu.sa

Course Information

Comprehensive Course Description	The course includes (i) eight teaching modules, (ii) three integrated computer exercises, and (iii) a research presentation & opposition. Historical Overview and Conceptual Positioning. Major events in Biology, Genomics, and Computing: Turing, the Action Potential, Neumann, DNA, Cybernetics, Different modes of Computing, rise of Molecular Biology, Complex Systems, sequencing of the Human Genome, sequence based technology explosion and production of multiple data-types, Systems Biology, high-performance Computing, CRISP/CAS editing, Synthetic Biology, machine intelligence. **(1)** Genomes, Transcriptomes and Epigenomes. Genome Alignment, Bioinformatics data processing and pipelines for RNA-seq, DNA methylation, microRNA, ATAC-seq, and Chip-Seq. The lectures cover the conceptual steps in data-processing as well as their respective practical implementation. This type of analysis ends with clean, normalized, and statistical relevant data. (*From Data to Information/Signals*) **(2)** Bioinformatics downstream analysis. This includes differential analysis (time, samples), enrichment, pathways, clustering and PCA. Illustrated using transcriptomics and epigenomics. (*From Information to Knowledge/biological insight*). **(3)** Overview of Available Databases, Resources, and Software. Alignment tools, Bioconductor, Github, IPA, STRING, cytoscape, NAR databases, tools, and software, Google Scholar. (*Practical Resources*) **(4)** Transcriptional Networks. Definition, how to find them (prior knowledge, reverse-engineering), how to analyse them (network analysis), using case studies including E-Coli, Yeast, Cell lines, and primary cells. (*Beyond lists of molecular entities*) **(5)** Metabolic Modelling, Analysis and Protein Bioinformatics and Analysis. Metabolite and protein networks/charts, protein sequences and 3D structure, network analysis, pathways, modelling and flux analysis. **(***Interface between the inside and outside of cells)* **(6)** Data Science - Concepts, Ideas, and Tools. Multiple testing, cross validation, bootstrapping, normalization, uni/multi variate statistics, networks, supervised and unsupervised learning, Dimension reduction, clustering techniques, PCA, SVM, ICA, MDS, and tSNE. **(***Conceptual Resources and Challenges*) **(7)** Machine Learning, Mathematical Modelling, Computing, and Artificial Intelligence – Genomics, Biomedicine, and Health Care. Motivation, models, inverse problems, forward simulation models, ODE/Boolean models, dynamical systems, and uncertainty. Machine Learning & Deep Neural Networks. **(***From Associations to Causality*) **(8)** Single Cell Biology. The emerging field of single cell genomics, single cell sequencing (sc-seq), scRNAseq, CyTof, scATACseq, and analytics of data. **(***Towards precision biology and data-driven predictive analytics)* Summary and Outlook. Summary of current and emerging concepts and trends. Data to Models, Data Integration, Multi-scale Biology, Biomedical and Industrial applications, Synthetic Biology, and Artificial Intelligence. Some Example areas/systems/topics from which we will select a few during the relevant course modules to illustrate the use of integrative systems analysis, bioinformatics, and machine learning. Development Biology (Sea Urchin, Davidson), Eukaryotic Cell Cycle (Tyson/Novak), Action Potential (Hodgkin/Huxley), C-Elegance, Genomic Circuits in Stem Cells, Systems Immunology, Systems Medicine (P4 Medicine), Drug Development, Multi-scale Heart Modelling (Noble, VPH). Deep Learning for genomics and health care, AI in Medicine **(***Real world examples and research programmes)* There will be 3 topics that will be augmented with practical computer exercises. These are (a) genome alignment, (b) analysis of transcriptional (RNAseq) data, and (c) single cell genomics.
Course Description from Program Guide	The Course provides a broad and practical overview of selected techniques and concepts in rapidly developing areas such as bioinformatics, computational biology, systems biology, systems medicine, network biology, synthetic biology, data analytics, predictive modelling, machine learning, and machine intelligence. Topics are selected to be of relevance for the computer scientist, working biologist, computational scientist, and applied investigator (Biotechnology and engineering).
Goals and Objectives	The course provides a broad and practical overview of available techniques and concepts in the area of computational bioscience and machine learning. This includes areas referred to as bioinformatics, computational biology, systems biology, systems medicine, mathematical biology, network biology, synthetic biology, data analytics, predictive modelling, computer simulations, learning algorithms, machine learning, AI, and machine intelligence. Topics are selected to be of relevance for the computer scientist, working biologist, computational scientist, and applied investigator (biotechnology and engineering). The aim is to support the development of bilingual students. Either coming from a bioscience or computer science background, thus having a firm ground in his/her area of expertise the student will develop a practical and conceptual language supporting the student with a working knowledge to be able to communicate and collaborate effectively with experts trained outside his/her domain of expertise. From the standpoint of biology, the course addresses the analysis of complex biological systems at difference scales, ranging from molecules, cells, organs, to organisms. The computer scientist/data scientist will acquire a working knowledge on which techniques from machine learning, network theory, modelling, and statistics are useful and how to apply them in the analysis of biological systems also in the context of drug development and biomedicine. The biologist is provided with concepts and practical tools how to advance from data to information about significant correlated features hidden in the data and advancing to insights into biological mechanisms. The scope of the course is intentionally broad since the aim is to provide the student with a comprehensive conceptual and in part practical overview of this inter-disciplinary area. Yet we include practical computer exercises, targeting three core areas, in order to ground the lectures in hands-on-knowledge. The course is relevant for the researchers in life-science in academia or industry, and data-scientists developing new analytic techniques in research or industry. Upon the completion of the course the student can readily advance to further studies in selected sub-areas, which in part are offered in the BESE or CEMSE programs. Alternatively, the student develops the background and ability to collaborate and identify the relevant specialists in the appropriate sub-area. Computational techniques are necessary, not yet sufficient, to understand and analyse data produced from living systems. Eo ipso, or by the same token, complex living systems are a rich source of inspiration for developing new techniques to decipher patterns, learning algorithms, and causal mechanisms in big data.

Required Knowledge	Knowledge and skills corresponding to course B204 Genomics and basic mathematics and statistics facilitate the understanding of the course. Entry level includes MSc or PhD students. In case you are not sure please contact the instructor for an individual consultation.
Reference Texts	**Examples of Recommended Books (for reference only):** 1. Selected chapters from the books referred to below will be recommended and distributed 2. The instructor compiles slides and course notes. - -**"Handbook of Systems Biology: Concepts and Insights"** Edited by Walhour, Vidal, and Dekker. Academic Press 2013 -**"A first Course in Systems Biology"** Berhard O. Voit. Garland Science, 2013. -**"Computational and Statistical Epigenomics"** Edited by Andrew E. Teschendorff, Springer, 2015. -**"Uncertainty in Biology"** Edited by Liesbet Geris & David Gomez-Cabrero, Springer, 2016. -**"Systems Medicine"** Edited by Ulf Schmitz & Olaf Wolkenhauer, Methods in Molecular Biology, Springer Protocols, Humana Press, 2016. -**"Computational Biomedicine"** Edited by Coveney, Diaz-Zuccarini, Hunter, & Viceconti. Oxford University Press, 2014. -**"Handbook of Statistical Systems Biology"** Edited by Stumpf, Balding, Girolami. Wiley, 2011.
Method of evaluation	**25.00**% - Course Project(s) **25.00**% - Homework /Assignments **50.00**% - Final exam
Nature of the assignments	**Final exam:** Includes all topics covered in the course. A total of one exam will be given, the final exam during the final exam period. **Research Presentation and Opposition:** Each student – the *presenter* - will be assigned a topic (research question/challenge) and required to give a 15-minute presentation in the second half of the course. The student presents a strategy and implementation meeting the question/challenge and another student – the *opponent* – leads the critical discussion following the presentation. The presenter shares a ½-1 page written outline to the opponent prior to the presentation. This task supports training in presenting and motivating a scientific project, crafting a sound analysis plan, and the opponent is challenged with leading a scientific discussion. **Computer Exercises:** Three topics for practical sessions - running scripts, evaluate results. The topics are (a) how to perform genome alignment, (b) how to analyse transcriptional (RNAseq) data using a pipeline of scripts and steps, and (c) Single Cell Genomics – dropouts, normalization, and tSNE plots.
Course Policies	Attendance to the classes is mandatory.
Additional Information	The course webpage is on the Blackboard system. Any course announcements will be posted on this site. In addition, the lecture notes, problem sets, solutions, and any extra material will be available on the Blackboard.

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Mon 01/28/2019 Thu 01/31/2019	Historical Overview and Conceptual Positioning. Major events in Biology, Genomics, and Computing: Turing, the Action Potential, Neumann, DNA, Cybernetics, Different modes of Computing, rise of Molecular Biology, Complex Systems, sequencing of the Human Genome, sequence based technology explosion and production of multiple data-types, Systems Biology, high-performance Computing, CRISP/CAS editing, Synthetic Biology, machine intelligence.
2	Mon 02/04/2019 Thu 02/07/2019	**(1)** Genomes, Transcriptomes and Epigenomes. Genome Alignment, Bioinformatics data processing and pipelines for RNA-seq, DNA methylation, microRNA, ATAC-seq, and Chip-Seq. The lectures cover the conceptual steps in data-processing as well as their respective practical implementation. This type of analysis ends with clean, normalized, and statistical relevant data. (*From Data to Information/Signals*)
3	Mon 02/11/2019 Thu 02/14/2019	**(1)** Genomes, Transcriptomes and Epigenomes. Genome Alignment, Bioinformatics data processing and pipelines for RNA-seq, DNA methylation, microRNA, ATAC-seq, and Chip-Seq. The lectures cover the conceptual steps in data-processing as well as their respective practical implementation. This type of analysis ends with clean, normalized, and statistical relevant data. (*From Data to Information/Signals*)
4	Mon 02/18/2019 Thu 02/21/2019	**Computer Exercises:** Three topics for practical sessions - running scripts, evaluate results. The topics are (a) how to perform genome alignment, (b) how to analyse transcriptional (RNAseq) data using a pipeline of scripts and steps
5	Mon 02/25/2019 Thu 02/28/2019	**(2)** Bioinformatics downstream analysis. This includes differential analysis (time, samples), enrichment, pathways, clustering and PCA. Illustrated using transcriptomics and epigenomics. (*From Information to Knowledge/biological insight*).
6	Mon 03/04/2019 Thu 03/07/2019	**(3)** Overview of Available Databases, Resources, and Software. Alignment tools, Bioconductor, Github, IPA, STRING, cytoscape, NAR databases, tools, and software, Google Scholar. (*Practical Resources*)
7	Mon 03/11/2019 Thu 03/14/2019	**(4)** Transcriptional Networks. Definition, how to find them (prior knowledge, reverse-engineering), how to analyse them (network analysis), using case studies including E-Coli, Yeast, Cell lines, and primary cells. (*Beyond lists of molecular entities*)
8	Mon 03/18/2019 Thu 03/21/2019	**(5)** Metabolic Modelling, Analysis and Protein Bioinformatics and Analysis. Metabolite and protein networks/charts, protein sequences and 3D structure, network analysis, pathways, modelling and flux analysis. **(***Interface between the inside and outside of cells)*
9	Mon 03/25/2019 Thu 03/28/2019	Spring Break
10	Mon 04/01/2019 Thu 04/04/2019	**Research Presentation and Opposition:** Each student – the *presenter* - will be assigned a topic (research question/challenge) and required to give a 15-minute presentation in the second half of the course. The student presents a strategy and implementation meeting the question/challenge and another student – the *opponent* – leads the critical discussion following the presentation. The presenter shares a ½-1 page written outline to the opponent prior to the presentation. This task supports training in presenting and motivating a scientific project, crafting a sound analysis plan, and the opponent is challenged with leading a scientific discussion.
11	Mon 04/08/2019 Thu 04/11/2019	**Research Presentation and Opposition:** Each student – the *presenter* - will be assigned a topic (research question/challenge) and required to give a 15-minute presentation in the second half of the course. The student presents a strategy and implementation meeting the question/challenge and another student – the *opponent* – leads the critical discussion following the presentation. The presenter shares a ½-1 page written outline to the opponent prior to the presentation. This task supports training in presenting and motivating a scientific project, crafting a sound analysis plan, and the opponent is challenged with leading a scientific discussion.
12	Mon 04/15/2019 Thu 04/18/2019	**(6)** Data Science - Concepts, Ideas, and Tools. Multiple testing, cross validation, bootstrapping, normalization, uni/multi variate statistics, networks, supervised and unsupervised learning, Dimension reduction, clustering techniques, PCA, SVM, ICA, MDS, and tSNE. **(***Conceptual Resources and Challenges*)
13	Mon 04/22/2019 Thu 04/25/2019	**(7)** Machine Learning, Mathematical Modelling, Computing, and Artificial Intelligence – Genomics, Biomedicine, and Health Care. Motivation, models, inverse problems, forward simulation models, ODE/Boolean models, dynamical systems, and uncertainty. Machine Learning & Deep Neural Networks. **(***From Associations to Causality*)
14	Mon 04/29/2019 Thu 05/02/2019	**(8)** Single Cell Biology. The emerging field of single cell genomics, single cell sequencing (sc-seq), scRNAseq, CyTof, scATACseq, and analytics of data. **(***Towards precision biology and data-driven predictive analytics)*
15	Mon 05/06/2019 Thu 05/09/2019	**Computer Exercises:** (c) Single Cell Genomics – dropouts, normalization, and tSNE plots.
16	Mon 05/13/2019 Thu 05/16/2019	Summary and Outlook. Summary of current and emerging concepts and trends. Data to Models, Data Integration, Multi-scale Biology, Biomedical and Industrial applications, Synthetic Biology, and Artificial Intelligence.

17	Mon 05/20/2019 Thu 05/23/2019	Final Exam Week

Note

The instructor reserves the right to make changes to this syllabus as necessary.