**Data Analysis, Modeling, Visualization (Algorithms)**

|  |
| --- |
| **Computational techniques:**  Bayesian inference,  deep learning,  hidden Markov models (HMMs),  random forests,  convolutional neural networks,  auto-encoders,  recurrent neural networks,  clustering and classification,  k-means,  hierarchical clustering,  model complexity selection,  network structure,  PCA,  SVD,  network diffusion kernels,  quantitative trait mapping,  mediation analysis,  causality inference,  string matching,  sequence alignment,  tree data structures,  rapid database search,  hashing,  data integration,  pattern finding,  expectation maximization,  Gibbs sampling.  **Biological applications:**  Genetic associations, common/rare variants, GWAS, PheWAS, multi-trait mapping, EHR mining, cancer genomics, CRISPR, biological sequence analysis, gene finding, comparative genomics, RNA structure folding, sequence alignment, gene expression analysis, motifs, epigenomics, single-cell genomics, evolutionary analysis, gene/species trees, coalescent, personal genomics, population genomics, human ancestry, recent selection, disease mapping, genetic association analysis, population genetics, regulatory genomics, dissecting disease mechanism. |

|  |
| --- |
| **Inferential Statistics, Analytics, Machine Learning** |
| |  | | --- | | **Inferential Statistics, Analytics, Machine Learning (R) Videos** | | **P Values, clearly explained**  <https://www.youtube.com/watch?v=5Z9OIYA8He8&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=16>  **Linear Regression in R**  <https://www.youtube.com/watch?v=u1cc1r_Y7M0&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=24> **Linear Models Pt.1 - Linear Regression**  <https://www.youtube.com/watch?v=nk2CQITm_eo&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=23>  **Linear Models Pt.2 - t-tests and ANOVA**  <https://www.youtube.com/watch?v=NF5_btOaCig&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=25>  **Linear Models Pt.3 - Design Matrices (old version)**  <https://www.youtube.com/watch?v=2UYx-qjJGSs&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=26>  **Linear Models Pt.3 - Design Matrix Examples in R**  <https://www.youtube.com/watch?v=Hrr2anyK_5s&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=27>  **PCA main ideas**  <https://www.youtube.com/watch?v=HMOI_lkzW08&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=21>  **Principal Component Analysis (PCA) clearly explained**  <https://www.youtube.com/watch?v=_UVHneBUBW0&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=22>  **Principal Component Analysis (PCA), Step-by-Step**  <https://www.youtube.com/watch?v=FgakZw6K1QQ&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=3>  **PCA Practical Tips**  [**https://www.youtube.com/watch?v=oRvgq966yZg&feature=youtu.be**](https://www.youtube.com/watch?v=oRvgq966yZg&feature=youtu.be)  **PCA in R**  <https://www.youtube.com/watch?v=0Jp4gsfOLMs&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=4>  **PCA in Python**  <https://www.youtube.com/watch?v=Lsue2gEM9D0&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=5>  **RPKM, FPKM, TPM**  <https://www.youtube.com/watch?v=TTUrtCY2k-w&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=6>  **MDS and PCoA**  <https://www.youtube.com/watch?v=GEn-_dAyYME&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=7>  **MDS and PCoA in R**  <https://www.youtube.com/watch?v=pGAUHhLYp5Q&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=15>  **t-SNE, Clearly Explained**  <https://www.youtube.com/watch?v=NEaUSP4YerM&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=8>  **K-means clustering**  <https://www.youtube.com/watch?v=4b5d3muPQmA&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=28>  **Hierarchical Clustering**  <https://www.youtube.com/watch?v=7xHsRkOdVwo&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=10>  **Drawing and Interpreting Heatmaps**  <https://www.youtube.com/watch?v=oMtDyOn2TCc&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=11>  **False Discovery Rates, FDR, clearly explained**  <https://www.youtube.com/watch?v=K8LQSvtjcEo&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=17>  **Fisher's Exact Test and the Hypergeometric Distribution**  <https://www.youtube.com/watch?v=udyAvvaMjfM&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=18>  **Logs (logarithms), clearly explained**  <https://www.youtube.com/watch?v=VSi0Z04fWj0&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=20> |  |  | | --- | | **A Matrix Algebra Companion for Statistical Learning (matrix4sl)**  <https://www.gastonsanchez.com/matrix4sl/types-of-tables.html> |  |  | | --- | | **StatQuest: (Josh Starmer)**  **U North Carolina Chapel Hill**  [**https://statquest.org/video-index/**](https://statquest.org/video-index/)  **Videos** | | **High-throughput Sequencing Analysis: StatQuest (Josh Starmer)**  **Introduction to RNA-seq**  <https://www.youtube.com/watch?v=tlf6wYJrwKY&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=1>  **Introduction to ChIP-Seq**  <https://www.youtube.com/watch?v=nkWGmaYRues&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp&index=2> **RNA-seq**  <https://statquest.org/tag/rna-seq/>  **edgeR, part1: Library Normalization**  <https://www.youtube.com/watch?v=Wdt6jdi-NQo&feature=youtu.be>  **DESeq2, part1: Library Normalization**  <https://www.youtube.com/watch?v=UFB993xufUU&feature=youtu.be>  **edgeR and DESeq2, part2: Independent Filtering (removing genes with low read counts)**  <https://www.youtube.com/watch?v=Gi0JdrxRq5s&feature=youtu.be>  **FDR and the Benjamini-Hochberg Method**  <https://statquest.org/statquest-fdr-and-the-benjamini-hochberg-methoc-clearly-explained/>  **Linear Discriminant Analysis (LDA)**  <https://statquest.org/statquest-linear-discriminant-analysis-lda-clearly-explained/>  **Heatmaps how to draw and interpret them**  <https://statquest.org/heatmaps-how-to-draw-and-interpret-them/>  **RNA-seq: The Pitfalls of Technical Replicates**  <https://statquest.org/rna-seq-replicates-clearly-explained/>  **PCA**  <https://statquest.org/pca-clearly-explained/>  **RPKM, FPKM and TPM**  <https://statquest.org/rpkm-fpkm-and-tpm-clearly-explained/> |  |  | | --- | | **Knowledge of Inferential Statistics, Analytics, Machine Learning**  **Statistics for Genomics**  <https://www.youtube.com/playlist?list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr> | | **17 Videos:** 2017 May  **Statistics for Genomcs: Distances and Clustering**  <https://www.youtube.com/watch?v=wQhVWUcXM0A&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=2>  **Statistics for Genomics Lab: Quick Introduction to R and Bioconductor**  <https://www.youtube.com/watch?v=J5h5WxOn3Gw&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=11>  **Statistics for Genomics Lab: Distances and Clustering RStudio**  <https://www.youtube.com/watch?v=PArRvqLUP6o&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=7>  **Statistics for Genomics: Introduction to RNAseq**  <https://www.youtube.com/watch?v=C8RNvWu7pAw&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=12>  **Statistics for Genomics: Advanced Differential Expression**  <https://www.youtube.com/watch?v=QINX3cI7qgk&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=15>  **Statistics for Genomics: Useful plots and bad plots**  <https://www.youtube.com/watch?v=46-t2jOYsyY&list=PLdl4u5ZRDMQQpUcSDRKN3V2vvx3_SmMbr&index=17> |      |  | | --- | | **Inferential Statistics, Analytics, Machine Learning (Reading)** | | **5 Questions which can teach you Multiple Regression (with R and Python)** October 15, 2015  <https://www.analyticsvidhya.com/blog/2015/10/regression-python-beginners/>  **Going Deeper into Regression Analysis with Assumptions, Plots & Solutions** July 14, 2016  <https://www.analyticsvidhya.com/blog/2016/07/deeper-regression-analysis-assumptions-plots-solutions/>  **7 Regression Techniques you should know** August 14, 2015  <https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>  **Statistics for Analytics and Data Science: Hypothesis Testing and Z-Test vs. T-Test** June 18, 2020  <https://www.analyticsvidhya.com/blog/2020/06/statistics-analytics-hypothesis-testing-z-test-t-test/>  **Commonly used Machine Learning Algorithms (with Python and R Codes)** September 9, 2017  <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> |  |  | | --- | | **Relevant Journal Publications:**  **Example of good search** | | [**https://journals.plos.org/ploscompbiol/search**](https://journals.plos.org/ploscompbiol/search)  **All Fields: bioconductor**  **Publication Date: 2016-01-01 - 2020-10-09**  **Sort By: Most Bookmarked**  **Subject Area: Principal component analysis**  >> results  **Ten quick tips for effective dimensionality reduction**  20 Jun 2019 PLOS Computational Biology  <https://doi.org/10.1371/journal.pcbi.1006907>  <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006907>  Citations: 19  **Context Specific and Differential Gene Co-expression Networks via Bayesian Biclustering**  28 Jul 2016 PLOS Computational Biology  <https://doi.org/10.1371/journal.pcbi.1004791>  Citations: 22  **Machine learning-based microarray analyses indicate low-expression genes might collectively influence PAH disease**  12 Aug 2019 PLOS Computational Biology  <https://doi.org/10.1371/journal.pcbi.1007264>  Citations: 2 | |

|  |
| --- |
|  |
| |  | | --- | | **Bioinformatics Courses** | | **Computational Biology: Genomes, Networks, Evolution** (MIT Open Courseware) <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-047-computational-biology-fall-2015/>  Textbook 547 pages  <https://ocw.mit.edu/ans7870/6/6.047/f15/MIT6_047F15_Compiled.pdf>  Assigned Reading:  <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-047-computational-biology-fall-2015/readings/>  Included  Gene Regulation 1 – Gene Expression Clustering 225 Gene Regulation 2 – Classification 239 | | **Analysis of single cell RNA-seq data (**University of Cambridge Bioinformatics)  2-day Course, 16 hours of YouTube video  <https://biocellgen-public.svi.edu.au/mig_2019_scrnaseq-workshop/public/index.html#docker-image>  quality control, visualisation, data normalisation, exploratory data analysis, clustering, trajectory (pseudotime) inference, differential expression, batch correction, combining datasets, data integration, confounders, latent spaces, cell annotation, case studies  Day 1: <https://www.youtube.com/watch?v=thHgPqQpkE4&feature=emb_err_woyt>  Day 2: <https://www.youtube.com/watch?v=7dQ_pleDO2Y&feature=emb_err_woyt> |      |  | | --- | | **Machine Learning in Genomics** 6.047/6.878 Fall 2019 (MIT)  **Algorithms, Statistical Visualizations (No Code)**  Professor: Manolis Kellis  <http://stellar.mit.edu/S/course/6/fa19/6.047/> | | **Lectures: (26 videos)**  <https://www.youtube.com/playlist?list=PLypiXJdtIca6U5uQOCHjP9Op3gpa177fK>  1:15:38 Lecture 01 - Introduction  1:19:28 Lecture 02 - Dynamic Programming  1:23:05 Lecture 03 - Hashing BLAST Database Search  1:17:20 Lecture 04 - HMMs Hidden Markov Models I  1:15:45 Lecture 05 - HMMs Hidden Markov Models II  1:17:19 Lecture 06 - Expression Analysis Clustering Classification  1:16:22 Lecture 07 - RNA world, RNA-seq, RNA folding  1:11:44 Lecture 08 - Epigenomics I  1:22:42 Lecture 09 - Epigenomics II  1:18:33 Lecture 10 - Regulatory Genomics  1:16:12 Lecture 11 - Network inference and analysis  1:30:43 Lecture 12 - Deep Learning  1:18:54 Lecture 13 - Population Genetics  1:17:49 Lecture 14 - GWAS  1:18:37 Lecture 15 - eQTLs Mediation  1:14:19 Lecture 16 - Systems Genetics  1:22:58 Lecture 17 - Comparative Genomics  1:21:02 Lecture 18 - Genome Evolution  1:20:37 Lecture 19 - Phylogenetics  1:22:44 Lecture 20 - Phylogenomics  1:25:30 Lecture 21 - Single-cell genomics  1:26:33 Lecture 22 - Cancer Genomics  1:20:36 Lecture 23 - Multi-Phenotype analyses  1:18:40 Lecture 24 - Genome Engineering  1:20:29 Lecture 25 - How to Present - Papers, Figures, Presentations Materials: <http://stellar.mit.edu/S/course/6/fa19/6.047/materials.html>  Textbook: <https://ocw.mit.edu/ans7870/6/6.047/f15/MIT6_047F15_Compiled.pdf>  Assigned Reading: <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-047-computational-biology-fall-2015/readings/> | | **Computational techniques:**  Bayesian inference, deep learning, hidden Markov models (HMMs), random forests, convolutional neural networks, auto-encoders, recurrent neural networks, clustering and classification, k-means, hierarchical clustering, model complexity selection, network structure, PCA, SVD, network diffusion kernels, quantitative trait mapping, mediation analysis, causality inference, string matching, sequence alignment, tree data structures, rapid database search, hashing, data integration, pattern finding, expectation maximization, Gibbs sampling.  **Biological applications:**  Genetic associations, common/rare variants, GWAS, PheWAS, multi-trait mapping, EHR mining, cancer genomics, CRISPR, biological sequence analysis, gene finding, comparative genomics, RNA structure folding, sequence alignment, gene expression analysis, motifs, epigenomics, single-cell genomics, evolutionary analysis, gene/species trees, coalescent, personal genomics, population genomics, human ancestry, recent selection, disease mapping, genetic association analysis, population genetics, regulatory genomics, dissecting disease mechanism. |  |  | | --- | | **Machine Learning for Genomics** 6.047/6.878 Fall 2020 (MIT)  Professor: Manolis Kellis  11 videos (Lecture 6 - No Video)  <http://stellar.mit.edu/S/course/6/fa20/6.047/>  Foundations and frontiers of machine learning, statistical, and algorithmic techniques for understanding the human genome, epigenome, circuitry, evolution, and cancer/disease mechanism. | | Lecture 1 - **Introduction**  <https://www.youtube.com/watch?v=CTPs5HELLpo>  Lecture 2 - **Dynamic Programming**  <https://www.youtube.com/watch?v=AuXp5IZNq70>  Lecture 3 - **Local alignment Hashing BLAST alignmentScores**  <https://www.youtube.com/watch?v=pO3GCbmfUKQ>  Lecture 4 - **HMMs 1**  <https://www.youtube.com/watch?v=IcSou8bdKIA>  Lecture 5 - **HMMs 2**  <https://www.youtube.com/watch?v=uBZAWM612_E>  Lecture 07 - **RNA folding, RNA world, RNA structures**  <https://www.youtube.com/watch?v=s3nMNAa-CdQ>  Lecture 8 - **Epigenomics 1**  <https://www.youtube.com/watch?v=ywJep35QnjY>  Lecture 9 - **Epigenomics 2 and 3D genome**  <https://www.youtube.com/watch?v=6Ay5_NGCtPY>  Lecture 10 - **Regulatory Genomics and Motifs**  <https://www.youtube.com/watch?v=6dFUdFPUsTs>  Lecture 11 - **Networks**  <https://www.youtube.com/watch?v=rjFBdm_fhZg>  Lecture 12 - **Deep Learning**  <https://www.youtube.com/watch?v=wwKzgstpkes> | |

**bioinformatics.stackexchange.com**

<https://bioinformatics.stackexchange.com>

<https://bioinformatics.stackexchange.com/questions>