Applied Human Computational Genomics
Tuesday and Thursday, 9:35 AM – 10:55 AM in Cherry Emerson 206

Professor: **Fredrik O. Vannberg** and T.A.: **Cai Huang**
School of Biology, Office: IBB 1308
fredrik.vannberg@biology.gatech.edu   cai.huang@biology.gatech.edu

This course fills the niche of a heavily applied course in the modern interpretation of genomic data, focused on understanding genomic features important in human disease. Following closely in the format of Computational Genomics, taught by Professor Jordan, this class seeks to utilize raw sequencing reads to inform in clinical decision making as pertains to three areas of interest:
1-**Clinical metagenomics (CM)**, 2-**Mendelian genomics (MG)** and 3-**Cancer genomics (CG)**.

Within these research areas students will be assigned to subgroups within each of these topic areas:
1-**Database team (DT)**; *creating gold standard annotation for subsequent analysis*,
2-**Algorithm/coding team (AT)**; *creating pipelined analysis with appropriate analytic strategy*,
3-**Report/Visualization team (VT)**; *creating appropriate clinical reports for clinicians and patients*.

Students will be mentored in all stages of these projects, with an emphasis on how to think about the implementation of a minimally viable product (MVP) useful for clinical interpretation. The framework for this course culminates in the final student presentations at the end of the class, with 50% of the course grade allocated to this final project presentation.

This applied course is designed to broaden the student's understanding of how genomics will transform medicine in the future, and each of the subgroups (DT, AT, and VT) will add important skills to the student's broadening bioinformatics knowledge. This course will also introduce aspects of cloud computing and its increasing role in clinical medicine.

This class meets on Tuesdays and Thursdays from 9:35 AM to 10:55 AM in Cherry Emerson Room 206. Attendance is **mandatory** and missing any one of the classes without an excused absence from the Dean of Students will automatically reduce your participation points to zero. Although there is no assigned textbook for this course students will be assigned reading from the primary literature.

Learning Outcomes: Students will be able to:
1. Have a detailed understanding of next generation sequencing data, especially <.fastq> format from DNA and RNA studies
2. Accurately use and apply a large set of Unix/Linux utilities and commands
3. Create Unix/Linux shell scripts for data analysis and pipelining
4. Utilize higher level programming languages to interpret clinically relevant features within the data (programming languages include Python, R, Matlab, and C++)
5. Understand the concepts of modern clinical genetics in the core areas of: 1) cancer genomics, 2) Mendelian genomics and 3) clinical metagenomics

6. Annotate and create large-scale data sets from major bioinformatics databases, including data from curated clinical genetic databases. Be able to coordinate and harmonize data across different genomic builds

7. Understand the data models that underlie, and readily convert among, widely used bioinformatics file formats

8. Understand the fundamental concepts and principles that underlie utilizing k-mers (short stretches of equal length DNA) in interpreting <.fastq> files for DNA mutations and RNA expression

9. Write Unix/Linux shell scripts and/or R, python and Matlab to parse large-scale bioinformatics data sets and extract useful subsets of information

10. Contemplate and create bioinformatic pipelines using de novo code derived from first principle understanding of concepts (i.e. k-mer based SNP caller)

11. Understand how to create cloud based implementations of the pipelines created

12. Understand principles of cloud based computing

13. Create visualization and reports from the pipelines to portray the data simply for clinicians and patients

Students are required to uphold Georgia Tech's Academic Honor Code (www.honor.gatech.edu).

We will make classroom accommodations for students with documented disabilities. These accommodations must be arranged in advance and in accordance with the ADAPTS office (http://www.adapts.gatech.edu).

Grading

Class participation	10%
Group Introduction Presentation	10%
Group Strategy Presentation	10%
Group Preliminary Data Presentation	10%
Class Assignments	10%
Final Results & Documentation	50%

The grading rubric for the presentation involves the clarity of presentation, the robustness of the analysis, statistical appropriateness and the supporting evidence provided with each update leading up and including the final presentation. More specific details on grading will be provided during the initial lectures.

Date	Topic	Presenter
Tue 8/18/2014	Course Introduction and Logistics	Fredrik Vannberg and Cai Huang
Thu 8/20/2014	Compute Infrastructure Background	Fredrik Vannberg and Cai Huang
Tue 8/25/2014	Introduction to Research Topics (CM, MG, CG)	Fredrik Vannberg
Thu 8/27/2014	Introduction to Subtopic Areas (DT, AT, VT)	Fredrik Vannberg
Tue 9/1/2014	Clinical Metagenomics	Guest Lecture
Thu 9/3/2014	Introduction to Clinical Metagenomics	Students – CM group
Tue 9/8/2014	Initial Clinical Metagenomics strategy (DT, AT, VT)	Students – CM group
Thu 9/10/2014	Mendelian Genomics	Guest Lecture
Tue 9/15/2014	Introduction to Mendelian Genomics	Students – MG group
Thu 9/17/2014	Initial Mendelian Genetics strategy (DT, AT, VT)	Students – MG group
Tue 9/22/2014	Cancer Genomics	Guest Lecture
Thu 9/24/2014	Introduction to Cancer Genomics	Students – CG group
Tue 9/29/2014	Initial Cancer Genomics strategy (DT, AT, VT)	Students – CG group
Thu 10/1/2014	Preliminary Data - Clinical Metagenomics Database Team	Students – CM group
Tue 10/6/2014	Preliminary Data - Mendelian Genomics Database Team	Students – MG group
Thu 10/8/2014	Preliminary Data - Cancer Genomics Database Team	Students – CG group
Tue 10/13/2014	Holiday	
Thu 10/15/2014	Preliminary Data - Clinical Metagenomics Database Team	Students – CM group
Tue 10/20/2014	Preliminary Data - Mendelian Genomics Database Team	Students – MG group
Thu 10/22/2014	Preliminary Data - Cancer Genomics Database Team	Students – CG group
Tue 10/27/2014	Preliminary Data - Clinical Metagenomics Algorithm Team	Students – CM group
Thu 10/29/2014	Preliminary Data - Mendelian Genomics Algorithm Team	Students – MG group
Tue 11/3/2014	Preliminary Data - Cancer Genomics Algorithm Team	Students – CG group
Thu 11/5/2014	Preliminary Data - Clinical Metagenomics Visualization Team	Students – CM group
Tue 11/10/2014	Preliminary Data - Mendelian Genomics Visualization Team	Students – MG group
Thu 11/12/2014	Preliminary Data - Cancer Genomics Visualization Team	Students – CG group
Tue 11/17/2014	Cloud Computing Implementation	Fredrik Vannberg
Thu 11/19/2014	Preparation for Final Presentations (in class workshop)	All invited
Tue 11/24/2014	Final Presentation of Clinical Metagenomics Project / Demo	Students – CM group
Thu 11/26/2014	Holiday	
Tue 12/1/2014	Final Presentation of Mendelian Genomics Project / Demo	Students – MG group
Thu 12/3/2014	Final Presentation of Cancer Genomics Project / Demo	Students – CG group