Summary of the Independent study:
The independent study titled: Bioinformatics (WGS NGS and R Focus) was primarily designed to educate me on the basics of handling and analyzing large biological data sets. The over-all goal of the independent study was to train me to attain expertise in using R-programing language to analyze genomic data. My overall career goal is use my data analyzing skills to design clinical trials using genomic information. Therefore, this course was not only integral to my thesis project but also as my career as physician scientist.

My research workflow involves downloading the genomic data from the European Genome Phenome Archive, identifying the variants and analyzing them. During this process, I intend to learn the basic applications of command line and Linux, followed by the principles of using various command line tools to call and filter variants. There are a wide variety of tools available for these applications, therefore a part of the learning process was to gain insights to the factors that helps to identify the best tool for project specific application.

The course was primarily focused on the basic concepts of R, and also complex applications to analyze the data statistically and/or otherwise and also train me in making publication quality figures using R for my thesis and publication. There are numerous R-software packages available for scientific use. It is very important to understand how to make the selection of the right kind of analysis tools for the data under investigation.

At the end of the course I intended to attain the following the goals:

1) Understanding Good practices in coding: Understand the importance of properly documenting the code for reproducible bio-informatics research.
2) Understanding the basics of Linux: The basic command line command.
3) Gaining the basic skills of R programming language.
4) How to use R programming language for data analysis: How to import data and manipulate it.
5) The important tools for data visualization; what is data visualization
6) The basic understanding of the GATK pipeline. What are the filters used and why? What is a good variant? What is VCF file?
7) What is variant annotation? How to use annotated data?
8) To manipulate a data file by myself.

I feel with practice and with the completion of my CAP stone I will be able to achieve most of the objectives of this course. The most fun part of the course was learning something completely new and learning a computing language and to understand how to look at and understand genomic data sets. The most challenging part was to wrap head around working through programming, often I will get stuck at one point and it would be hours before I can figure out what was really wrong and most often I will not figure out what was wrong and need help from Dr. Bowman.