# workshop.mutation.2019

In this workshop you will learn about the fundamentals of mutation identification in cancer. By the end of the course you should be able to generate a full R markdown for your assigned TCGA ID with plots and identifying actionable mutations. Below is an outline of the course. We will start off with a 20 minute lecture talking about historical and general workflows. This will be follow by a workshop designed to completely interactive and will consist of an R script pre written that you can follow along. For each line of code we will describe what it’s doing and if necessary talk about the subject in detail. The workshop will also contain “bonus/challenge” questions that you are expected to try on your own. This will include using your own assigned TCGA ids. We will walk around during this time in case there are questions. Lastly we will also have homework questions that will be supply at the end of the course.

20 min Introduction:

Section I

1. Impact of next generation sequence in cancer biology.
2. The size of modern data sets (TCGA) and their current applications, e.g. in healthcare
3. Types of sequencing, e.g. WES and WGS, and also DNA vs RNA
   1. Pros and cons and how each modality can be integrated.
4. Basic bioinformatics workflow: from sample to identifying potentially targetable mutations. ( the workshop will be based on the later part )
5. Standard pipelines to go from reads to actionable information
   1. Nuisances to consider
   2. thresholding and filters.
   3. Differences between germline and somatic mutations.
6. What does a standard genome look like and how it informs the analysis.

Section II Basic vocabulary and concepts.

1. Classes of somatic mutations
2. The differences between synonymous vs nonsynonymous mutations

* # Pt mutations
  + Coding
  + Synonymous ( Silent )
  + Missense
  + Nonsense
  + Noncoding ( eg UTR )
  + Intronic
  + Intergenic
  + Splice site variants
* # Small regional mutations
  + Insertion
  + Deletions
  + Duplications

1. Deciphering nomenclature of sequence variations.
   1. DNA
   2. Protein

Section III Workshop

* We will start off with basic data mining. You will learn how to access cbioportal’s API through R and access all the dataset available to the public.
* From here we will download the BRCA dataset relevant to this course.
* Although this is not a course in R per se, but you will learn how to manipulate/wrangle mutation within the context of a data.frame.
* Learn how to identify potential cancer driver for your sample.
* Learn how to import and cross reference your mutations.
* Learn how to filter for germline mutations.
* Learn how to calculate allelic fraction for a given mutation.
* Study the mutation specific to the BRCA dataset.
  + This includes most frequent gene and types of mutation.
  + From here we will also determine
* Learn to query specific variants
* Learn to tabulate mutations.
* Learn how to identify what could be potentially be pathogenic and cross reference it with existing data.
* Learn how to take existing mutation data and predict possible actionable targets for either druggability, diagnostic or prognosis.
* You will also learn a few ways to plot the data.
* Basic plotting of your mutation table.
* Finally we will wrap up with a few other ways to cross reference your data with other external databases.

Summary and further exercises.

Here we will summarize and talk about the strength and shortcomings of the current bioinformatics field. There will also be some "homework" or extra take home exercises.