**Interview Data Challenge**

As part of our research into clear-cell renal cell carcinoma (ccRCC), we are of course very interested in the somatic mutational profile of tumour samples. Since whole genome or even whole exome next generation sequencing of every single biopsy sample would be prohibitively expensive, we frequently consider a so-called “driver panel” which targets the exons of a relatively small number of genes associated with ccRCC.

In order to provide an opportunity for you to display your data analysis skills, we have performed a similar experiment *in silico* using publically available mouse sequencing data. Specifically, we have taken whole genome sequencing data from a single mouse strain made available as part of the Sanger Institute’s Mouse Genomes Project and have “targeted” a selection of genes. This reduced set of reads will be made available to you in FastQ format and we would like you to process them as though they were a “real” sample.

As part of your interview, we would like you to prepare a short (~10 minutes) presentation discussing your analysis of this data and any results you obtained. As part of this, you may wish to include some of the following:

* Quailty Control
* Pre- and Post- alignment processing of the data
* Identification of which genes we may have targeted (you don’t need to provide an exhaustive list; an estimate of the total number and comments on any which you think may be of particular interest will be sufficient)
* Identification of point mutations and indels relative to the C57BL/6J strain (from which the mm9 and mm10 builds are derived)
* Examination of copy number variation events
* Any further analyses or observations you feel may be appropriate