Understanding the molecular complexity of human retroperitoneal liposarcoma

Introduction

Human retroperitoneal liposarcoma (RPLS) is a rare (1 per 100,000 people) tumour that develops slowly in the fat surrounding the kidneys near the back. Although rare, it is a devastating tumour with the potential for rapid spread to other organs. Very little is known about the genetic make up of this tumour however a recent collaborative project with Illumina UK and BGI Inc. has carried out the first panomics profiling of RPLS.

Data available

- Whole genome sequencing data (tumour cells 45x coverage, normal cells 30x coverage) from Illumina HiSeq 2500
- Total RNA sequencing data (35M reads) from Illumina Hiseq 2000
- Illumina HumanMethylation450 array data

Number of samples available: 12 paired human tumour:normal samples across all data types (WGS, RNAseq, Methylation)

Overall aim

To explore the molecular complexity of RPLS by a systems based analysis of NGS datasets to understand the drivers underlying this tumour type

Specific questions

1) What are the primary pathways that are dysregulated in liposarcoma?
2) How does synonymous mutation in the most frequently mutated driver genes in this cohort correlate with RNA change?
3) What alternative splicing events are occurring in liposarcoma?
4) What are the potential "druggable" pathways in liposarcoma?
5) Is there any evidence of initiation by a virus by examining the unmapped reads?

Prospects for a follow-on PhD project

During the mini-project we can only aim to achieve a first-pass analysis for a subset of the above questions. A PhD project follows naturally from expanding and deepening these analyses and there will be opportunities to develop topics of interest based on the findings in the mini-project. In addition, other types of sarcoma datasets may become available depending on the results of funding applications. This funding will also co-fund the costs of a MathSys studentship.

Approach in mini-project

We are going to use available software tools to identify mutations, differentially expressed genes, differentially spliced genes, and differentially methylated genomic loci. We will then do a statistical analysis to identify links between these features looking to detect the major signals in the data.