Characterizing the Cancer Genome

Cancer is a disease of the genome that is driven by a combination of possible germline risk-alleles together with a set of ‘driver’ somatic mutations that are acquired during the clonal expansion of increasingly fitter clones. Mutations occur at all levels and scales, including DNA point mutations, small insertions and deletions, larger genomic rearrangements and copy-number alterations, as well as epigenetic, transcriptional and proteomic changes. To generate a comprehensive list of all germline and somatic events that occurred during and prior to development of the cancer, we are developing and applying highly sensitive and specific tools for detecting these events in sequencing data. The complexity of the underlying cancer genomes requires the use of state-of-the-art statistical and machine learning approaches to most efficiently extract the signal from the noise (tools we developed include MuTect, Indelocator, SegSeq, CapSeg, dRanger, BreakPointer and others).

Detecting Cancer-Associated Genes

Once we detect the events in the cancer genomes, we analyze them across a cohort of samples searching for genes (and pathways) that show significant signals of positive selection, e.g. the number of mutations exceeds what is expected by chance. To do so, we construct a detailed statistical model of the background mutational processes and detect genes that deviate from this model. We have developed tools for detecting genes which are significantly gained or lost in cancer (GISTIC) and genes with increased density or irregular patterns of mutations (MutSig, CLUMPS). We recently reported the importance of modeling the heterogeneity of these mutational processes across patients, sequence contexts and along the genome, when searching for cancer-associated genes. We are improving
Somatic mutation frequencies across cancer.

Each dot represents the total frequency of somatic mutations (in the exome) in each tumor–normal pair. Tumor types are ordered by their median somatic mutation frequency, from haematological and paediatric tumors (left), to tumours induced by carcinogens such as tobacco smoke and ultraviolet light (right). Mutation frequencies vary more than 1,000-fold between lowest and highest across different cancers and also within several tumour types. The bottom panel shows the relative proportions of the six different possible base-pair substitutions. Taken from Lawrence et al. (2013).

Selected Publications:

*Co-corresponding authors

Heterogeneity and clonal evolution of cancer

Cancer samples are heterogeneous, containing a mixture of normal (i.e. non-cancer) cells and a population of cancer cells that often represents multiple subclones. Keeping in mind that cancer is a dynamic system, these subclones may represent the remaining cells of less-fit clones which have not yet been overtaken by the expanding most-fit clone or they may represent interacting sub-clones that co-evolved to support each other and reached an equilibrium or a combination of these scenarios. Our lab has been developing tools (ABSOLUTE) for characterizing the heterogeneity of cancer samples using copy-number, mutational and other data measured on bulk samples and now also getting into the analysis of single cells. Using these tools, we can infer which mutations are clonal (i.e. exist in all cancer cells) or sub-clonal (i.e. exist in subclones), as well as estimate the number of subclones and monitor their evolution over time or space by studying multiple samples from the same patient. Recently, we demonstrated that sub-clonal driver mutations are associated with outcome, emphasizing the importance of including clonal information in clinical trials.

Mutational processes

Mutations are the product of multiple processes that damage, repair, replicate and deliberately alter DNA. We use mutation data to study these processes, understand their mutational signatures, infer their molecular mechanisms and identify alterations that are associated with their activity. We recently found an association between mutations in ERCC2 and a specific mutational signature. By studying asymmetries in mutational processes we were able to detect that a mechanism that works on the lagging strand of DNA while it is replicated and a new mutational process that generates mutations on the non-transcribed strand.