PhD Academy Scholarship Proposal
[redacted]
5/1/21

Colorectal Cancer (CRC) is a common cancer with more than 1.2 million newly diagnosed patients each year. Survival rates vary greatly depending on progression of the disease upon diagnosis, where patients with localized disease having a five-year relative survival rate of 90%, while patients with metastatic disease experiencing a rate of 11%. Several molecular markers are used for prognostic purposes, such as *BRAF* and *KRAS* mutations, and mismatch repair deficiency, however, a greater understanding of the interconnections between these heterogeneous molecular backgrounds, prognosis, and response to treatment is needed in order to increase relative survival rates (Brenner 2013).

There has been interest in the role of transposable elements (TEs) in CRC carcinogenesis. Recent work has shown that retrotransposons, typically repressed in healthy tissue, become activated in CRC, with recurrent insertions observed in known oncogenes such as *APC*, and a high number of insertions has been associated with poor survival (Cajuso 2019). It has been hypothesized that these TEs drive expression of oncogenes, in particular by reactivation of early developmental TEs to repurpose these developmental pathways to promote carcinogenesis, but the mechanisms of action, in particular whether or not these TE activation events represent initial drivers of dedifferentiation, remains unclear (Lynch-Sutherland 2020). There is evidence that TEs may also play a role in evading immune surveillance. Our research group has found TEs to be suppressed in leukemia stem cells, as well as in high risk cases of myleodisplastic syndrome relative to low risk cases (Colombo 2017). There remains a need to understand the role of TEs in CRC.

We propose to explore the nature of TE activation in CRC by analyzing single cell RNA sequencing data (scRNA) from cells derived from primary CRC tumor samples as well as CRC cell lines. We will determine important parameters such as the degree of TE activation, whether or not these TE insertion events occur in developmentally important pathways or known oncogenes, as well as associate these events with the presence of other mutations known to be prognostic in the clinic, such as in the mismatch repair genes (de la Chapelle 2010). We will also use structural equation modelling (Nuzhdin 2012) in order to understand the effects of abberate gene expression or TE activity on downstream gene activity. We will also investigate whether these patterns of TE activity and downstream gene expression observed in the scRNA data hold true in a separate cohort of colorectal cancer RNA sequencing samples from whole tissue (Muzny 2012). We will use available clinical data with this cohort, such as demographic information, treatment, and survival, in order to model how these variables might relate to observed gene expression.

By exploring TEs at the single cell level, we intend to expand the understanding of the mechanism of these activation events in CRC tumorigenesis and identify cleaner signatures of cell specific activity than what would be available with whole tissue RNA sequencing. We hope to connect these patterns observed in this single cell data to a separate cohort comprised of whole tissue RNA sequencing in order to relate these patterns to clinically relevant variables. Understanding how gene activity and clinical information relate to each other is necessary to develop personalized treatment options that will improve patient outcomes.

References:

Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *The Lancet* **383**, 1490–1502 (2014).

Cajuso, T. *et al.* Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nature Communications* **10**, 1–9 (2019).

Colombo, A. R. *et al.* Suppression of Transposable Elements in Leukemic Stem Cells. *Scientific Reports* **7**, 7029 (2017).

de la Chapelle, A., and Hampel, H. (2010). Clinical relevance of microsatellite instability in colorectal cancer. Journal of Clinical Oncology *28*, 3380–3387.

Lynch-Sutherland, C. F., Chatterjee, A., Stockwell, P. A., Eccles, M. R. & Macaulay, E. C. Reawakening the Developmental Origins of Cancer Through Transposable Elements. *Front. Oncol.* **10**, (2020).

Muzny, D.M., Bainbridge, M.N., Chang, K., Dinh, H.H., Drummond, J.A., Fowler, G., Kovar, C.L., Lewis, L.R., Morgan, M.B., Newsham, I.F., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Nuzhdin, S.V., Friesen, M.L., and McIntyre, L.M. (2012). Genotype-phenotype mapping in a post-GWAS world. Trends Genet. *28*, 421–426.

Budget:

$500 of research support will be used to fund high performance computation activities at USC's Center for Advanced Research Computing. These funds will be used to cover the costs of compute and condo storage resources for this work on the endeavour cluster system.

Timeline:

Data access requests are already completed or in process for the RNA sequencing and scRNA sequencing cohorts. The computational analysis for this work along with the writing up of the eventual manuscript, which will happen in parallel with the computational analysis, will take one year.