Research Abstract – Mutational Capture Sequencing

Gene mutations directly impact human, animal, and ecosystems health. Mutations affects how living organisms respond to environmental stressors and is a driver of evolution. Manmade and naturally occurring agents such as cosmic radiation, hydrocarbons, and cadmium can directly or indirectly increase mutation rate, causing numerous pathologies including cancer. Thus, understanding the mechanisms by which exposure to these agents drive mutations is key in the development of interventions. One of the biggest and most direct impediments in mutation research is the rare and transient nature of DNA damage, the event that precedes a DNA mutation. When DNA sustains damages, three events may proceed. Cells could: (1) Repair the damage by activating DNA repair pathways. (2) Undergo programed cell death to prevent the conversion of a damaged locus into a new mutation site, or (3) Undergo DNA replication, converting the damaged locus into a new mutation site. As a result, it is very difficult to track DNA mutations from carcinogen induced DNA damage to the eventual DNA mutation. We have a developed a novel mutational capture sequencing method (MutC-Seq) to solve this major impediment in one of the most prevalence forms of DNA damage, base-mismatches, which drive single nucleotide changes. In a pilot study, we were able to identify and compare TG mismatches in an isogenic pair of Hereditary leiomyomatosis and renal cell cancer (HLRCC) cell line: UOK262-EV, which harbors biallelic inactivating mutation to the gene encoding Fumarate Hydratase (*FH*), and UOK262-RES (FH-rescued line). The outline of the draft MutC-Seq method is shown in figure 1. Briefly, total gDNA was sonication and size selected into 500 bp fragments. Fragmented DNA was then blunt ended and circularized. Non-circularized DNA was then removed. DNA mismatch sites were then digested with mismatch endonuclease I (New England Biolabs), resulting in the linearization of DNA fragments that contain mismatches. This is then followed by the annealing of Illumina's p5 and indexed p7 adaptors. Sequencing libraries were then amplified by PCR. DNA of cellular origins (non-PCR origin) is removed by *Dpn*I digestion. The amplified library was size selected into 500 bp fragments to remove any unligated adaptors, and the purified library was sequenced in a pair end configuration, with standard Illumina Read1 (R1), Read2 (R2), and indexing primers. Sequencing reads was then preprocessed to make it compatible with legacy aligners, and then mapped as a single end read, and the mismatch endonuclease digestion site will be easily identifiable from the overlapping Read1 and Read2 regions. The proposal will build on our pilot work with the overall aim to optimize MutC-Seq and develop a software package to handle and process the sequencing results. We will achieve this with the following aims

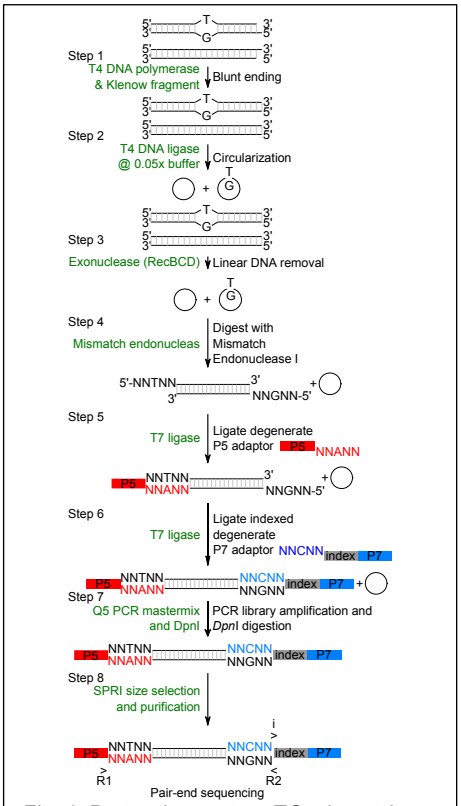

Fig. 1. Protocol to capture TG mismatches

Aim 1. Further optimize the MutC-Seq library preparation method. I will optimize each library prep stage of MutC-Seq to increase the number of damaged loci identified and assess the quality of the constructed library by cloning the DNA fragment into a pUC19 vector through Gibson cloning. Twenty individual clones will be picked, miniprepped and Sanger sequenced with M13-forward and M13-reverse primers to ensure the incorporation of Illumina p5, p7, and index sequences into the captured DNA segments.

Aim 2. To develop a software suite and a C++ library to handle and process data generated by MutC-Seq. Upon optimization of the sequencing libraries, we will identify DNA mismatches by next generate illumina sequencing. The mutational captured sequencing method splits and flips the DNA fragment at the site of a base mismatch following linearization of circularized DNA. We have developed the first phase of a C++ program to unflip these DNA fragments. I will further develop this C++ program and implement it into a software package that will regenerate new fastq files that work with legacy genome alignment tools. The software package will be made freely available to research community, and source code will be shared through GitHub.

Aim 3. Apply the method to characterize mutational processes driven by environmental exposure. To identify DNA mismatches caused by environmental carcinogens, I will culture expose primary skin, lung and bladder epithelial cells (purchased from ATCC) to environmentally important carcinogens including arsenic, cadmium, chromium, and UV light. Exposed cells will be compared to their passage controlled unexposed counterparts and DNA mismatch events will be identified using the optimized MutC-Seq technique.