Challenges in renal cell carcinoma biomarker discovery.

Customer case study

The McGill Genome Centre successfully applies the NxSeq AmpFREE Kit for whole genome and targeted sequencing in renal cancer biomarker discovery.

Introducing renal cell carcinoma

Renal cell carcinoma (RCC) is the most common form of kidney tumours, and the incidence of RCC is increasing worldwide. At present, a biological marker has not been identified for routine clinical use, making identification and treatment of RCC a challenge. In addition, RCCs are resistant to conventional chemotherapy and radiotherapy resulting in a need to identify novel therapies to enable effective treatment of patients. There are several different subtypes of RCC, with clear cell RCC (ccRCC) being the most frequently diagnosed and accounting for over 75% of cases. Figure 1 shows a high magnification micrograph of ccRCC.

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

The ccRCC subtype is characterised by heterogeneity in tumour biology and in clinical outcomes. An improved understanding of the genetic determinants underlying this heterogeneity is crucial in facilitating personalised patient care. The Riazalhosseini laboratory at the McGill Genome Centre, Canada, utilises a combination of next generation sequencing (NGS) functional genomics and computational analysis in their RCC research. The team recently generated genome, epigenome and transcriptome landscapes of ccRCC which are being comprehensively analysed to identify molecular variants that drive RCC tumorigenesis and progression or can serve as potential biomarkers for clinical applications.

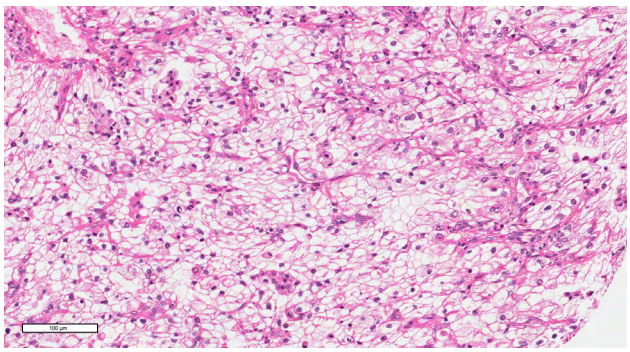

Figure 1. High magnification micrograph of a clear cell renal cell carcinoma. Note the classic clear cytoplasm in the cells, which are typically arranged in nests. 100 µm scale bar (bottom left). Image credit: McGill Genome Centre

Multi-step next generation sequencing workflows

An essential step in any NGS workflow is the preparation of an NGS library, within which template nucleic acid is prepared for sequencing on an NGS instrument. Typical NGS library preparation protocols are time consuming, involving multiple steps (e.g. enrichment, fragmentation, end-repair and ligation) and sample transfers between tubes. Using a kit specifically designed for the preparation of NGS libraries can reduce the risks of user error and improve consistency of reactions.

Multiple centre cohort studies

To deliver statistical significance in studies of complex biomarker interactions, it is crucial to have high sample numbers. Availability of samples within individual clinics are limited, hence multiple centre cohort studies are required. Rolling out a study over multiple centres requires extra focus on standardisation of workflows. For biomarker clinical research, such as the work presented here, the NGS library preparation workflow should deliver high quality data with low failure rates across samples. These requirements are of utmost importance when working within multiple centre cohort studies.

The nature of samples in cancer research

In cancer research, the nature of the biological samples creates further challenges; it can often be difficult to obtain good quality DNA. For this reason, there is a need for generating robust libraries for a range of DNA input types, including crude DNA. In RCC research, DNA types include high quality genomic DNA (gDNA), low quality gDNA purified from formalin-fixed paraffin-embedded tissues (FFPE gDNA), and circulating cell-free DNA (cfDNA) which are degraded fragments of DNA, ~165 bp long, that have been released into biofluids such as blood plasma.

The NxSeq AmpFREE kit solution

The efficiency of library preparation drives the success of NGS projects. LGC, Biosearch Technologies™ developed the NxSeq™ AmpFREE kit to facilitate high efficiency, low input, PCR-free library preparation. Each step of the protocol is optimised to ensure robust library construction, peak performance on Illumina sequencing instruments, with a fast and easy workflow to decrease user error risk. The NxSeq AmpFREE kit demonstrates considerable efficiency advantages over its two closest competitors, with a greater percentage of library DNA containing correctly ligated adaptors (Figure 2). The protocol is also considerably faster, thus streamlining customer's NGS workflows (Figure 3).

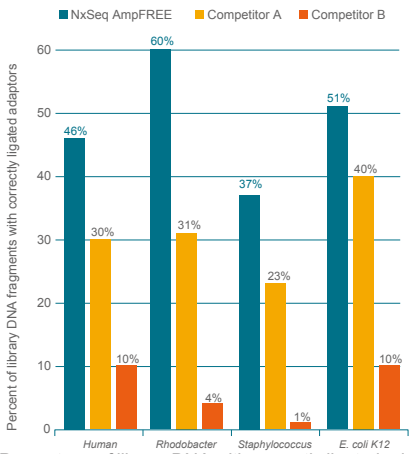

Figure 2. Percentage of library DNA with correctly ligated adaptors measured by qPCR. Duplicate libraries were prepped per kit/organism (Human, *Staphylococcus aureus*, *Rhodobacter sphaeroides* (1 library only), and *Escherichia coli*) according to the manufacturer's recommended input amounts and protocols. Adaptor ligation efficiency was measured by qPCR using the Competitor A kit and matching amplified library as an internal standard.

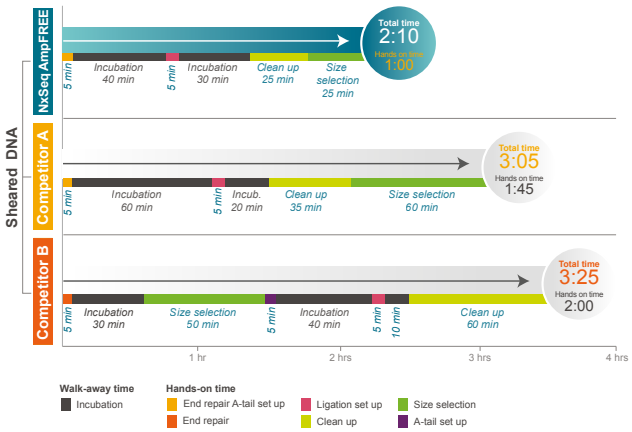

Figure 3. Comparison on NGS workflows using the NxSeq AmpFREE kit and two competitor kits. Hands on time and total time required to complete the protocol is much shorter for the NxSeq AmpFREE kit than both Competitor A and Competitor B.

The McGill Genome Centre initially utilised the NxSeq AmpFREE kit in a whole genome sequencing (WGS) study, and found it to give consistent results across a large number of blood-derived gDNA received from various labs around the globe. The kit's simplicity, robustness and successful implementation into their laboratory workflows enabled the group to adapt the kit for their more recent project involving interrogation of the mutational status of commonly mutated genes in ccRCC and their potential relationship to clinical outcomes. In addition to gDNA isolated from fresh-frozen tumours, template DNA within this project included both low quality FFPE gDNA and circulating tumour DNA (ctDNA, a specific component of cfDNA), both of which can be challenging to obtain as high quality samples. The adapted workflow included 6-8 cycles of amplification for FFPE gDNA and 8-10 cycles of amplification for ctDNA. As shown in Figure 4, NGS libraries were successfully generated using FFPE gDNA, with the aligned insert size proving comparable to libraries generated from high quality DNA extracted from fresh tumour tissue. Targeted NGS libraries (hybridisation-based enrichment) were also successfully generated from minute amounts of ctDNA in the cfDNA samples, further illustrating the efficiency of library preparation using

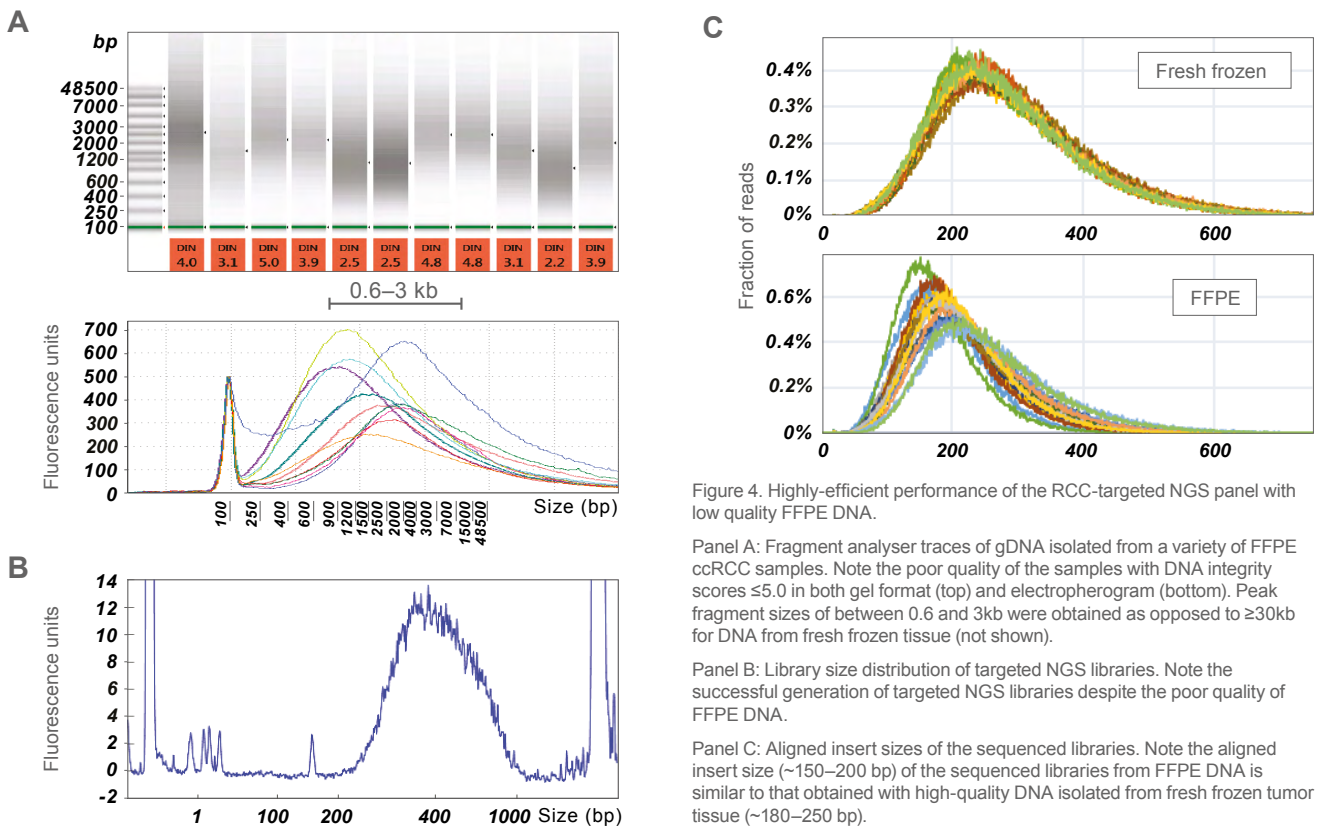

Figure 4. Highly-efficient performance of the RCC-targeted NGS panel with low quality FFPE DNA.

Panel A: Fragment analyser traces of gDNA isolated from a variety of FFPE ccRCC samples. Note the poor quality of the samples with DNA integrity scores ≤5.0 in both gel format (top) and electropherogram (bottom). Peak fragment sizes of between 0.6 and 3kb were obtained as opposed to ≥30kb for DNA from fresh frozen tissue (not shown).

Panel B: Library size distribution of targeted NGS libraries. Note the successful generation of targeted NGS libraries despite the poor quality of FFPE DNA.

Panel C: Aligned insert sizes of the sequenced libraries. Note the aligned insert size (~150–200 bp) of the sequenced libraries from FFPE DNA is similar to that obtained with high-quality DNA isolated from fresh frozen tumor tissue (~180–250 bp).

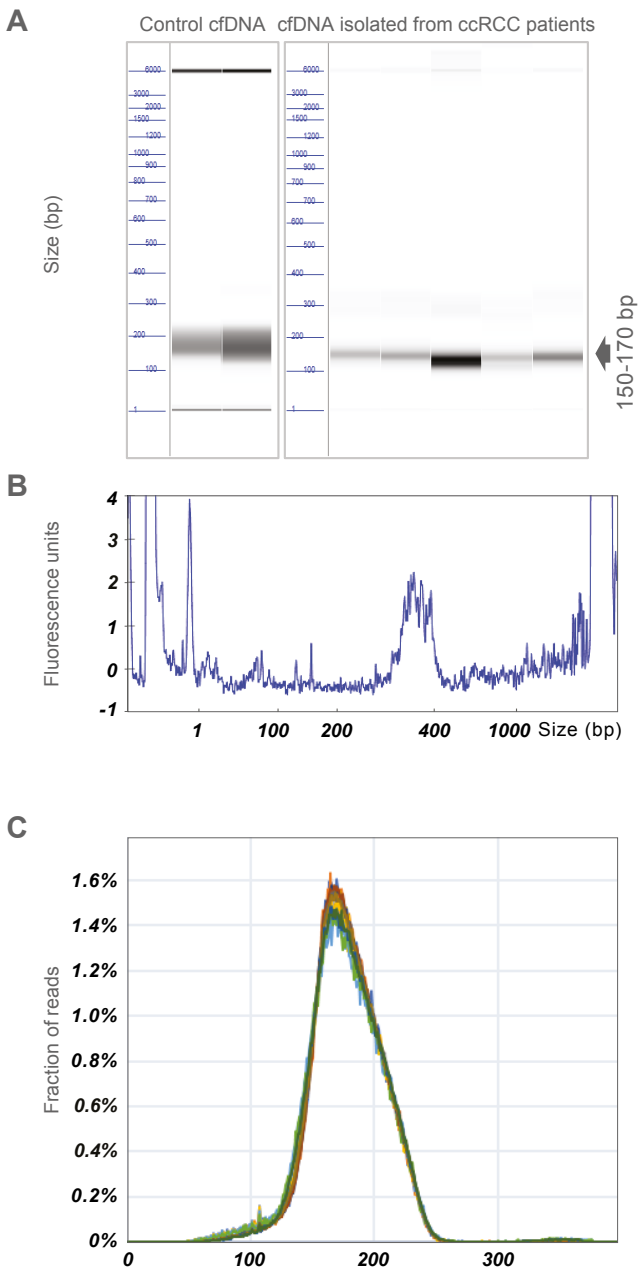

Figure 5. Highly-efficient performance of the RCC-targeted NGS panel with minute amounts of cfDNA.

Panel A: Size distribution of cfDNA extracted from commercial control samples and ccRCC patient samples. Note the typical 150–170 bp size distribution obtained from both sample types.

Panel B: Library size distribution of targeted NGS libraries. Note the successful generation of targeted NGS libraries from as little as 1 ng cfDNA.

Panel C: Aligned insert sizes of the sequenced libraries. Note the aligned insert sizes of the sequenced targeted libraries yields a single peak at ~165 bp as expected for cfDNA.

the NxSeq AmpFREE kit (Figure 5).
The team captured target regions from the amplified libraries using a standard probe-based hybridisation approach with custom-designed capture probes targeting coding regions of candidate genes commonly mutated in RCC, and sequenced the captured regions. This custom panel covers 17 genes with 534 probes with a total size of ~120,000 bp. The assay was able to specifically capture targets of interest in all examined sample types. Deep sequencing and unique molecular identifiers (UMI) facilitated the identification of mutations with ultra-low frequency in small portion of ctDNA present in the cfDNA samples. UMIs were successfully used during data analysis to identify unique, sequenced fragments and build consensus sequences for each region. This process enabled the elimination of sequencing artefacts, PCR-duplicates and facilitated the identification of high-confidence true mutations. An illustration of this process for a representative target of the RCC-panel is shown in Figures 6A-C.

The full process utilised by the Riazalhosseini laboratory in this project is represented in Figure 7.

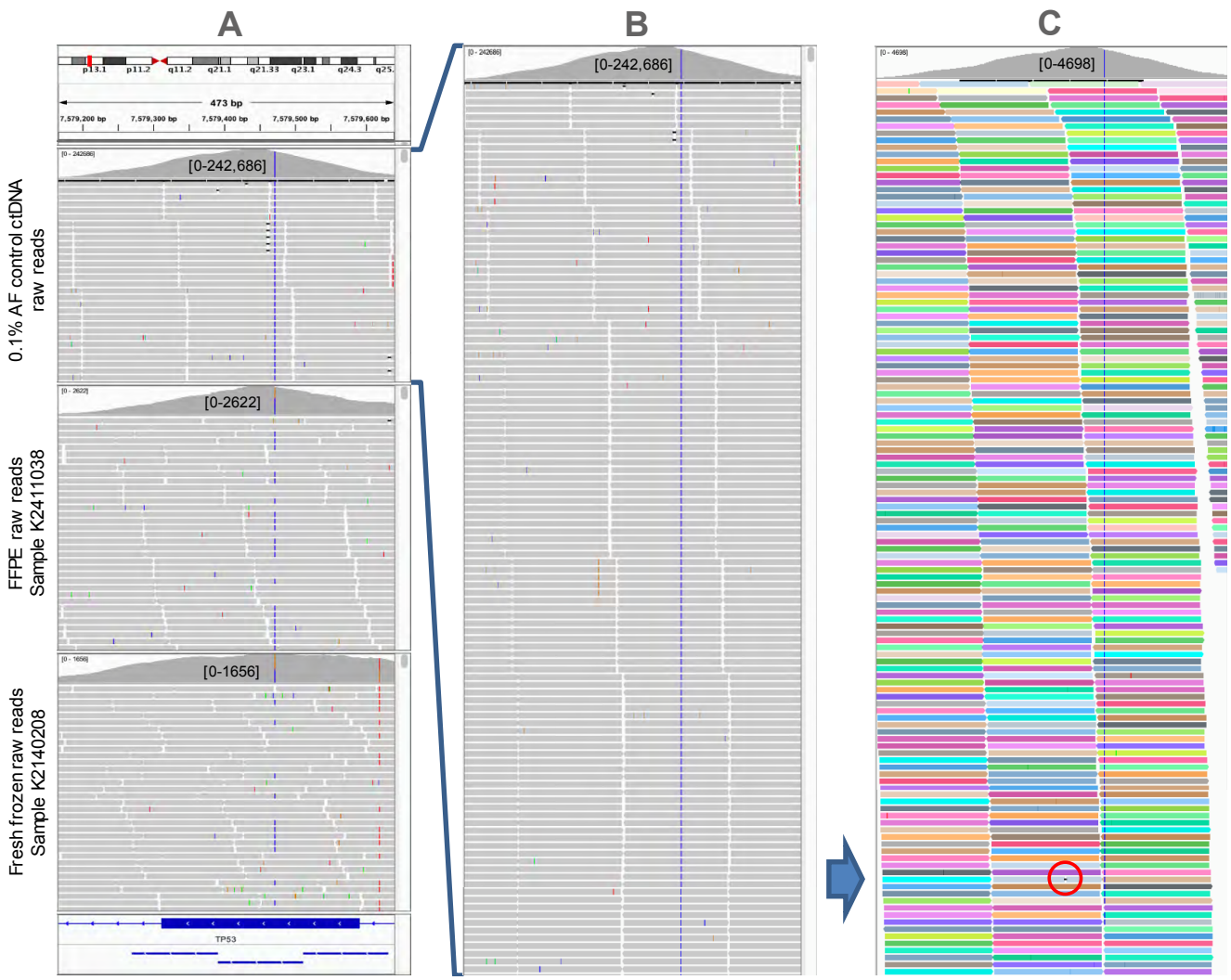

Figure 6. Leveraging the power of UMIs and deep sequencing to identify mutations with ultra-low frequency in ctDNA. Actual sequencing data of NGS libraries generated for the RCC-panel is visualised with help of Integrative Genomics Viewer (IGV) software. (A) Sequencing reads mapping to a 473 bp region of chromosome 17 spanning exon 4 of TP53, as a representative target of the RCC-panel, are shown for a control cfDNA sample with 0.1% VAF (top), FFPE DNA (middle), and fresh-frozen DNA (bottom). These data demonstrate the efficacy of the assay to specifically capture targets of interest in all examined sample types. The three capture probes targeting this region of interest are shown at the bottom below the gene tract depicting the exon in question. Each IGV panel shows a coverage histogram at the top with the scale reiterated in enlarged font for legibility purposes. (B) An expanded view of individual sequencing reads for the cfDNA sample are shown. The coloured vertical lines within reads represent sequencing errors. (C) The effect of collapsing individual reads from panel B into consensus sequences of high confidence using the UMI tags. The different coloured reads depict different UMI-tagged molecules. This molecular tag-based consensus building allows the elimination of artifacts and identification of high-confidence, true mutations (highlighted by red circle at the level of the blue arrow). Compare the much cleaner signal of a frameshift-inducing 1-base deletion in panel C as compared to the noisy background of errors in panel B.

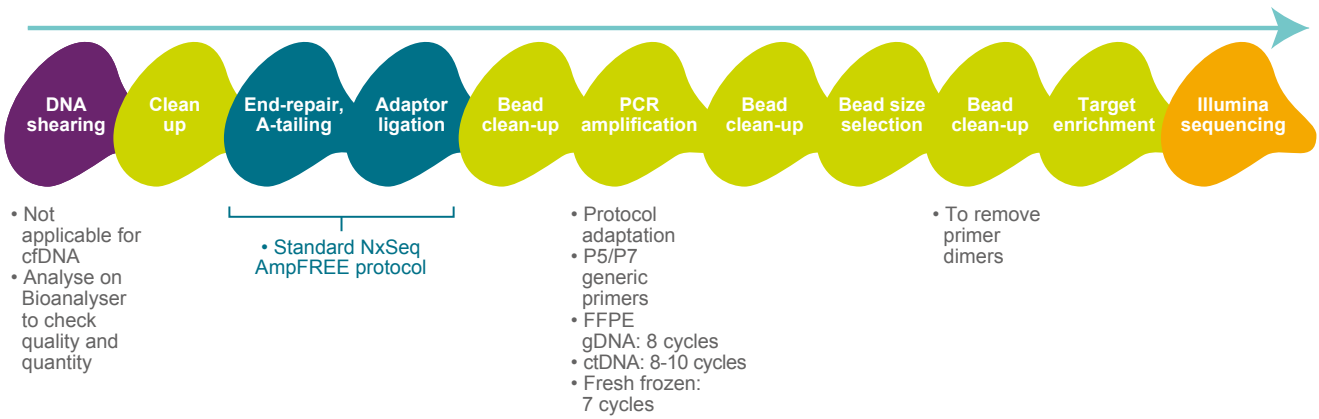

Figure 7. Schematic of the full laboratory process employed in this project. The Riazalhosseini laboratory utilised the standard NxSeq AmpFREE protocol within their laboratory workflow to facilitate generation of high quality NGS libraries from gDNA from fresh frozen samples, FFPE gDNA and cfDNA starting material. A PCR amplification step was employed following adapter ligation and reaction clean-up due to the difficult nature of the starting materials.

Outcome

This case study, focusing on research conducted by Riazalhosseini laboratory at the McGill Genome Centre, clearly demonstrates successful implementation of the NxSeq AmpFREE kit in clinical research, in particular ccRCC biomarker discovery. High quality NGS libraries were generated using challenging FFPE gDNA and cfDNA starting materials, producing consistently excellent results from more than 80% of the samples. The robust performance of the NxSeq AmpFREE kit has facilitated progress in this customer's laboratory, enabling targeted sequencing of an international multiple centre cohort of 750 ccRCC patients.

The researchers involved in this project have identified novel genetic drivers of poor outcome in ccRCC clinical samples. Some of these identified mutations appear to function in a sex-dependent context in-line with the previously established fact that a patient's sex has considerable effect on the risk of ccRCC development, as well as on clinical progression and outcomes. The outcomes of this work have the potential to significantly impact patient management allowing accurate risk stratification for recurrence and progression in a sex-specific manner, timely identification of patients experiencing failure, better adapted follow up regimens, and potentially better selection of candidates for adjuvant treatments.

In summary, the NxSeq AmpFREE kit demonstrated a robust workflow, delivery of high quality, highly informative data, with high confidence in results from challenging sample sources. The application of this kit within the complex world of biomarker discovery enables the development of new diagnostic assays for improved therapies for both this disease across the wider molecular diagnostics field.

The McGill Genome Centre

The McGill Genome Centre, Canada, was founded in 2002 with the principal aim of advancing genomics in medical research. The Centre comprises over 215 faculty members, staff and students, and conducts research in genomics and quantitative analysis to understand the causes and cures of human diseases. In addition, the Centre supports the scientific community at large through its core platforms and the provision of advanced training in new methodologies for generating and analysing massive biological datasets.

Dr Yasser Riazalhosseini's laboratory focuses on generating and analysing large-scale cancer genomics data to study carcinogenesis and tumour progression at the molecular level, and also to identify molecules that may have clinical applications for better diagnosis, prognosis or therapeutic use. Dr Riazalhosseini's team currently comprises two postdoctoral fellows, two research assistants, a research associate and four graduate students.