**Shicheng Guo**

**Center for Precision Medicine, Marshfield Clinic Research Institute**

**Circulating cell-free DNA based low-pass genome-wide bisulfite sequencing aids non-invasive surveillance to Hepatocellular carcinoma**

Haikun Zhang1#, Peiling Dong2#, **Shicheng Guo**3#, Chengcheng Tao1, Wenmin Zhao2, Jiakang Wang4, Ramsey Cheung5, Augusto Villanueva6, **Steven J. Schrodi**3,7\*, Dake Zhang1\*, Changqing Zeng1\*

1Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100101, China 2Department of Hepatology, Beijing You’an Hospital Affiliated with Capital Medical University, Beijing 100069, China 3Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, USA 4Biology Department, Stonybrook University, Stonybrook, NY, USA 5 Department of Gastroenterology and Hepatology, VA Palo Alto Health Care System and Stanford University, Palo Alto, CA, USA 6Liver Cancer Research Program, Division of Liver Diseases, Tisch Cancer Institute, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA 7Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI, USA 8

# These authors contributed equally to this work \* Corresponding Author

Circulating cell-free DNA has been demonstrated to provide a promising opportunity for non-invasive cancer diagnosis. DNA methylation in circulating cell-free DNA has exhibited particularly promising signals for cancer diagnosis and tissue-of-origin mapping. It is now well-recognized that genome-wide DNA hypo-methylation is a hallmark feature of the human cancer genome and therefore may be usefully applied to cell-free DNA-based cancer diagnosis. However, the amount of circulating cell-free DNA is typically too limited for interrogation with conventional high-depth/coverage genome-wide bisulfite sequencing (WGBS). Here we proposed a novel method in which we utilized long-region methylation (MethylLRM) in low-pass WGBS data (~5 million reads) generated from cfDNA to detect methylation changes. We applied the method to investigate dynamic methylation changes in cfDNA from blood samples of patients with hepatitis, cirrhosis, early and advanced hepatocellular carcinoma (HCC). We found a significant enrichment of differential methylation loci in intergenic and repeat regions, especially in HBV integration sites. Moreover, methylation profiles nearby HBV integration sites (MethyHBV) were found to enhance the prediction performance. Multiple machine learing models based on MethylLRM, MethyHBV, cfDNA fragment size (cfDNAsize) with five-fold cross-validation demonstrated low-pass cfDNA methylation data provided powerful discriminating ability. Our study shows that low-pass WGBS provides a stable and powerful diagnostic tool for HCC. Furthermore, our approach enables an evaluation of the efficacy of surgical intervention for HCC. Interestingly, we present evidence of over-representation of differentially methylated CpGs in HBV integration regions based on our low-pass WGBS approach, providing additional insights into the mechanisms of HCC molecular pathophysiology and may aid with HCC diagnosis and clinical decisions. Implementation of this approach is favored by the low cost compared to conventional techniques. Using machine learning, we show that HBV integration-based DNA methylation in cell-free DNA (MethylHBV) exhibited excellent predictive performance in distinguishing HCC from other liver diseases. Finally, using the same data, we introduced cell-free DNA fragment size distribution effects into our predictive model yielding a powerful HCC discriminating ability.

Shicheng Guo acknowledges support of the NLM training grant: 5T15LM007359.