Abstract Title: A framework for evaluation of new or modified sequencing technologies for use in human genomics

Author Name: Mr Michael Gatzen, Senior Computational Associate

Author Institute: Broad Institute of MIT and Harvard, Cambridge, MA, USA

Abstract: The rapid pace of development and innovation in sequencing technologies accelerates the genomics community's ability to understand the genetic underpinnings of disease, develop and deploy new screening and diagnostic assays, and advance the creation of new therapeutic modalities. When a new sequencing technology is released, or an existing technology is significantly changed (new instrument, new chemistry) there is a need to quickly understand the performance and limitations of the new platform so that decisions can be made about what applications may be most appropriate. The NIST Genome in a Bottle curated truth reference samples HG001 and HG002 are a critical resource as they represent a standard against which we can compare all new technologies in an unbiased manner. This is particularly applicable to understanding the performance of the platforms for human genomics but also highlights performance characteristics for microbial, somatic, and other applications. We have created a framework for assessment of new or updated technologies that gives a first pass insight into the performance of the technology across variant types and genomic contexts. Genomic contexts used are those defined by the GA4GH group and include tandem repeats and homopolymers, segmental duplications, low mappability regions and a range of GC content regions. We have also included the 273 genes included in the 'challenging medically-relevant genes' set that has been curated by NIST in HG002. We have applied that framework to a variety of currently available technologies including Illumina, Element Biosciences, PacBio and Oxford Nanopore. We share the results of those evaluations as well as the code and contexts required for researchers to apply this framework in their own hands. Leveraging these methods, we also present a framework for evaluating functional equivalence between sequencing and analysis modalities to mitigate the risk of batch effects when switching from one configuration to another.

Area of expertise: Computational Biology