Estimating Microsatellite Instability (MSI) Score with MSIsensor2

Note: Text is adapted from the MSIsensor2 manual

MSI stands for “Microsatellite Instable,” and is a classification of mutations within microsatellite regions of the genome. There are two possible classifications for MSI status, MSI-H and MSS. The classification MSI-High or MSI-H means that there is a high amount of instability. MSI-H samples are associated with a buildup of somatic microsatellite mutations in tumor cells and can lead to a spectrum of molecular and biological changes including high tumor mutational burden. Tumors with MSI-H (Deficient Mismatch Repair (dMMR)) are sensitive to immune checkpoint blockade (ICB), particularly to PD-1 and PD-L1 inhibitors. The opposite classification MSS or “Microsatellite Stable,” implies a low degree of mutation within microsatellite regions.

Msisensor2 is an extension of the tool MSIsensor. MSISensor2 incorporates a novel algorithm based on machine learning, featuring a large upgrade in the microsatellite instability (MSI) detection for **tumor only** sequencing data, including target gene sequencing data. Furthermore, it is applicable for WES, WGS, panel and cfDNA data. The original MSIsensor was specially designed for tumor/normal paired sequencing data.

Additional details on the tool from the developer:

Given tumor-only sequencing data, MSIsensor2 relies on a set of machine learning models to determine the MSI status for a distribution per microsatellite. The MSI score is calculated as follows:
number of msi sites / all valid sites. Where all valid sites is defined as a subset of all microsatellite regions with sequencing coverage exceeding a user defined threshold. In developer testing, results from the tumor-only module were comparable with matched tumor-normal sequencing data input. The developers recommend an MSI score cutoff value of 20% [MSI-H: MSI score >= 20%]. The developers also tested TCGA and EGA data with known MSI status, and showed the accuracy of the tumor-only module is up to 99%, and demonstrated that the tumor-only module in MSIsensor2 is comparable to the original MSIsensor tumor/normal paired module.

Citation

Msisensor2 is available on Github. This tool is yet to be published; however, the previous version of the algorithm is described in the following paper:

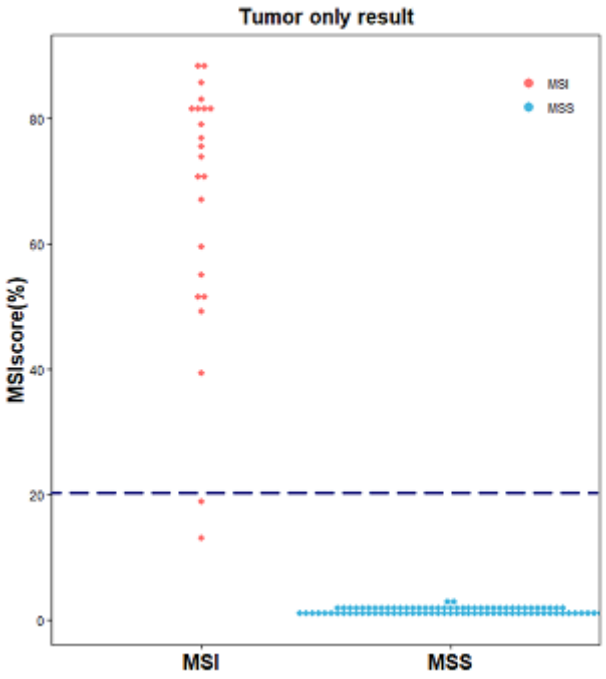
Beifang Niu, Kai Ye, Qunyuan Zhang, Charles Lu, Mingchao Xie, Michael D. McLellan, Michael C. Wendl and Li Ding. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics 30, 1015–1016 (2014).

Workflow overview

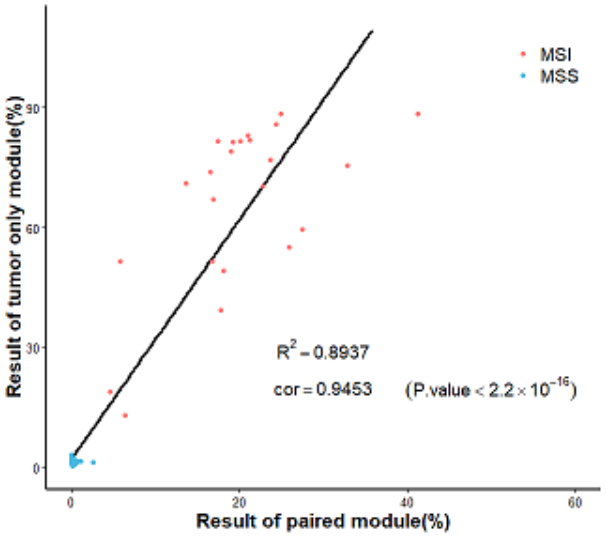
1. Align reads against reference, and apply appropriate alignment post-processing steps.
2. Determine MSI score with Msisensor2

MSI Threshold Score

The developers of the tool conducted a number of tests of the tumor-only module of their data, using TCGA and EGA data. They provided the following justification for the 20% threshold standard and demonstrated clear differentiation between MSI-H and MSS.



The developers also demonstrated a high correlation coefficient (0.94) between the matched tumor-normal and tumor-only modules. Although the two modules used different algorithms, the results were highly consistent, thus demonstrating that the tumor-only module is a reliable tool to use for determining the MSI status.



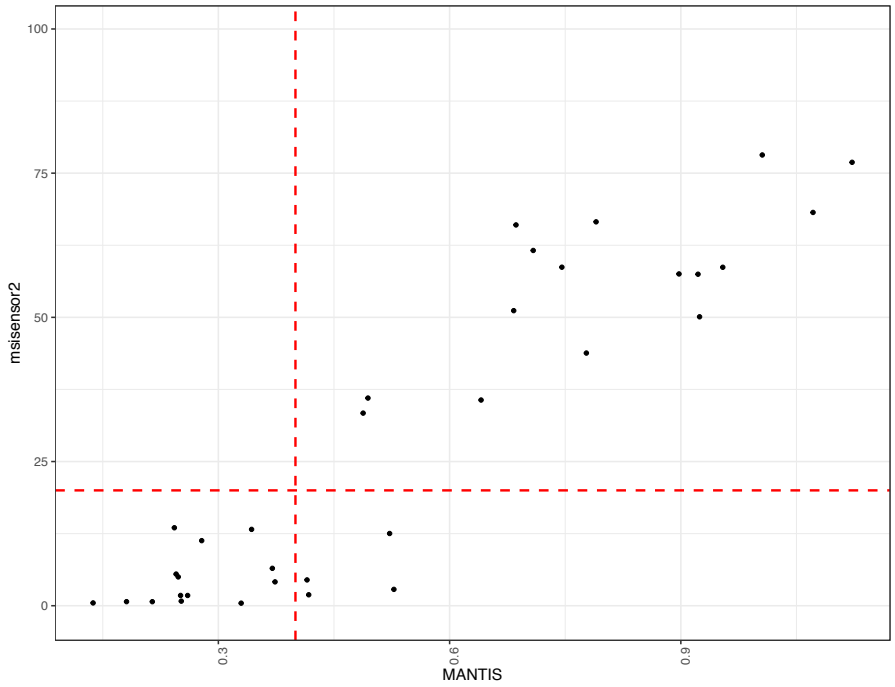
The Jackson Laboratory (JAX) Internal Benchmarking

JAX conducted a number of internal benchmarking tests of various MSI classifications tools including MSIsensor2, MSI-seq, msiNGS, and MANTIS.

The first test involved recreating a set of matched normal MSI calls generated by the tool MANTIS. This tool is recognized as a community standard for MSI detection for the samples with matched normal data. A number of MSI-H and MSS matched normal TCGA data samples, as determined by MANTIS in Kautto EA, et al. 2017, were selected and then JAX's internal whole exome pipeline with MANTIS was run on this dataset. For matched-normal TCGA samples, JAX was able to use MANTIS to correctly re-classify a set of 20 MSI-H and 14 MSI-stable tumors. This step was performed to ensure that our in-house data preprocessing and variant calling matches the TCGA standards and can successfully replicate the results of TCGA.

The majority of JAX PDX data comes from tumor-only samples, thus it was critical to identify a tool which could work effectively with tumor-only data. The following tools were tested to address this issue: MSIsensor2, MSI-seq, and msiNGS.

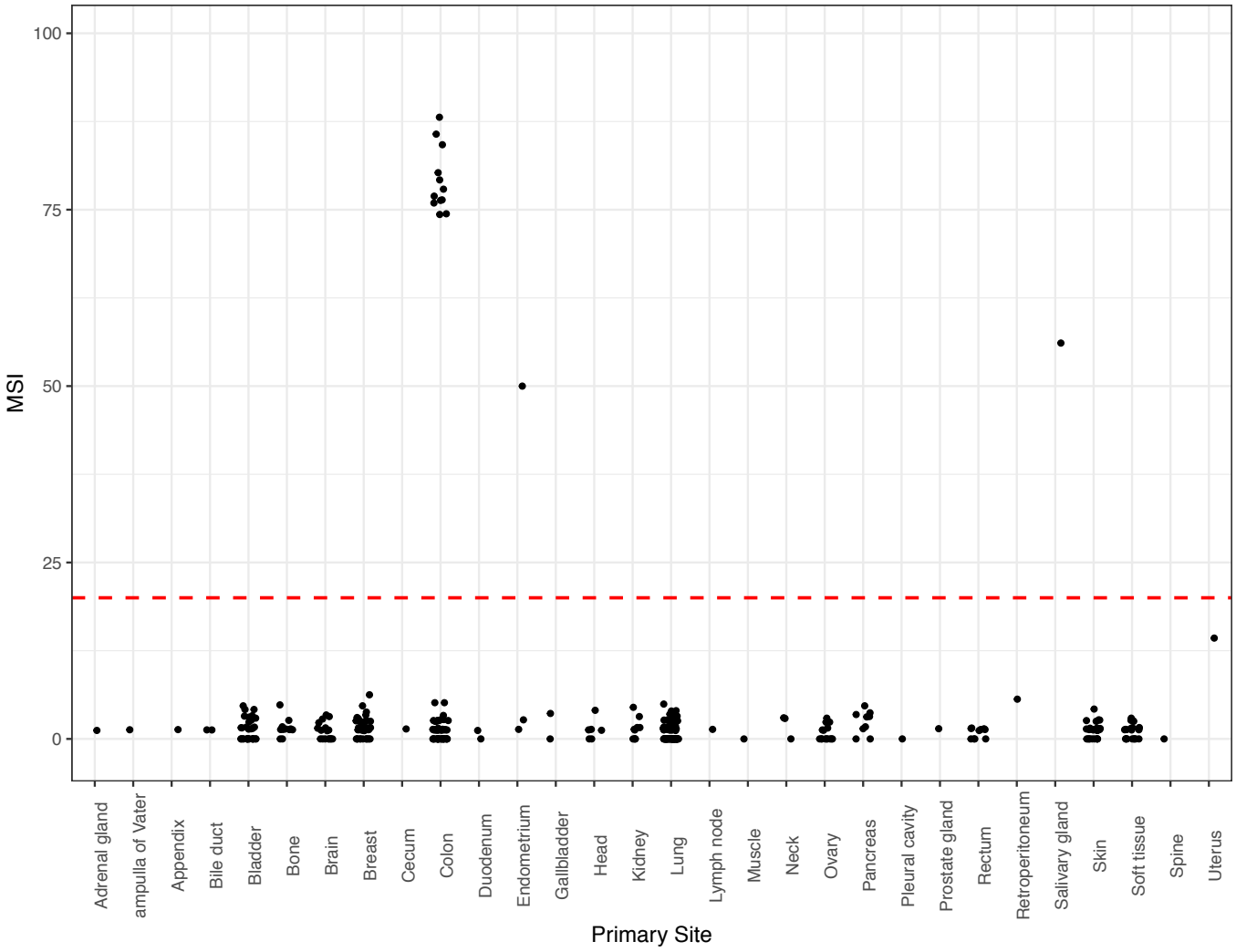
MSi-seq and msiNGS showed poor performance in determining the MSI status of the aforementioned TCGA samples in tumor only mode. However, MSIsensor2, was able to correctly classify all MSS samples as well as most of the MSI-H samples (16 out of 20). The four samples that MANTIS classified as MSI-H, were all borderline cases as determined by the tool.



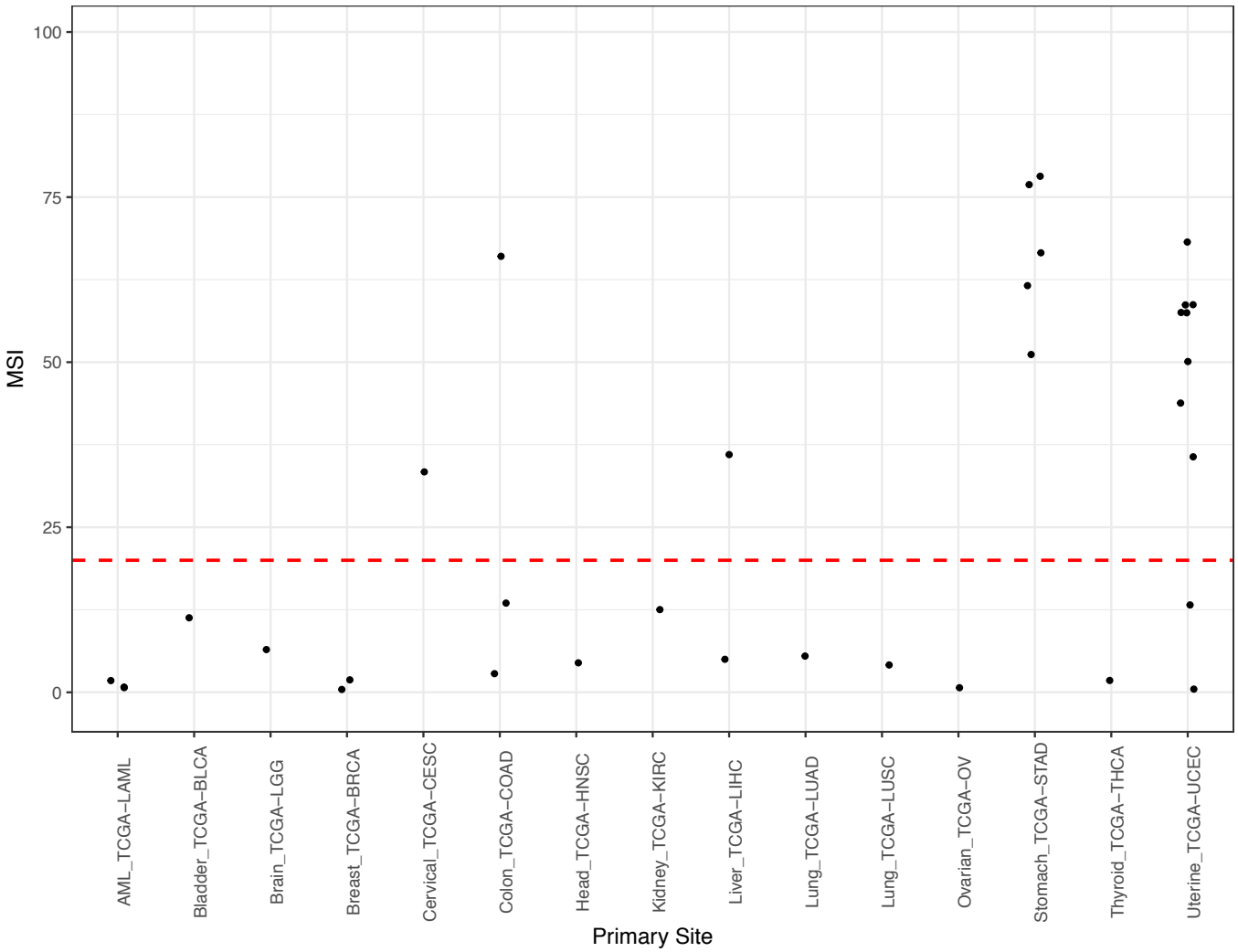
MANTIS matched normal scores on the X-axis with an MSI-H threshold of 0.4.
MSIsensor2 tumor-only scores on the Y-axis with a threshold of 20.

JAX PDX MSI Scores

Cancer Treatment Profile (CTP) data derived from 402 JAX PDX models were passed through MSIsensor2 to calculate MSI scores. The distribution of MSI scores by primary site shows that 14 of 402 models are classified as being above the score threshold of 20, with mainly JAX colon models being classified as MSI-H.



To ensure the sensitivity of MSIsensor2 in calling MSI status in other primary cancer types besides colon, additional tumor-only TCGA samples from a range of cancer types were run. MSIsensor2 was able to properly classify MSI status across various cancer types relative to published MSI status calls from MANTIS.



References

Kautto EA, et al. 2017 Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. Oncotarget. 2017;8(5):7452–7463. doi:10.18632/oncotarget.13918