Microsatellite instability detection with cell-free DNA next-generation sequencing

Ariane Lozac'hmeur, Ashraf Hafez, Jason Perera, Denise Lau, Aly A. Khan
Immunology and Data Science Groups, Tempus Labs

TEMPUS

BACKGROUND

Microsatellite instability is a clinically actionable genomic indication for cancer immunotherapy. In microsatellite instability-high (MSI-H) tumors, defects in DNA mismatch repair (MMR) can cause a hypermutated phenotype where alterations accumulate in the repetitive microsatellite regions of DNA. MSI detection is typically performed by subjecting tumor tissue ("solid biopsy") to clinical next-generation sequencing or specific assays, such as MMR IHC or MSI PCR. Circulating cell-free tumor DNA (cfDNA) testing ("liquid biopsy") is rapidly emerging as a less invasive method for detecting cancer and monitoring disease progression. **Here, we explore the possibility of detecting MSI in cfDNA using the Tempus xF cfDNA liquid biopsy assay.**

AIM

Develop a novel cfDNA MSI detection assay with specificity greater than 95%.

CHALLENGES

- cfDNA samples generally have a low tumor content (in the 1%-10% range).
- The germline phenotype of the patient (ie. the length of its microsatellite repeats) is unknown.
- The cohorts used to develop this assay were small thus increasing the risk of overfitting.

PATIENT COHORTS

	MSS (Training / Validation)		MSI-H (Training / Validation)	
Colorectal Cancer	7	8	5	4
Gynecologic Cancer	1	1	6	8
Other cancer type	30	39	6	4
Total	**38**	**48**	**17**	**16**

METHODS

1. Reads containing repetitive sequences are selected for analysis, including 5 base pairs on each flank.
2. The number of repeat units in each read are counted to determine relative frequency and distribution. Unstable loci typically result in **bimodal distribution** due to shorter reads originating from the tumor.
3. The following metrics are generated:
   - **Percent lower** – percent of reads with less repeat units than the most frequently occurring number of repeat units
   - **Mean lower** – average number of repeat units in reads with less repeat units than the most frequently occurring number of repeat units
   - **Mean loglikelihood** – the mean loglikelihood that each read originated from a stable locus based on the number of repeat units it contains. This probability model was built from a reference of 350 MSS solid tumors and blood samples.
4. Predict the probability of the locus being unstable using the 3 normalized metrics and a k Nearest Neighbors model (with k=100).
5. Calculate percentage of unstable loci in each patient. **Patients with more than 50% unstable loci are classified MSI-H.**

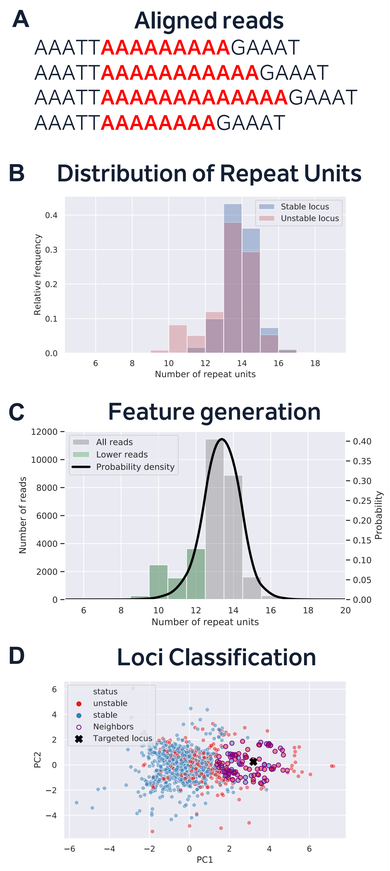

Figure 1. A. Illustration of aligned sequencing reads. **B**. Examples of distribution of number of repeat units. **C**. Illustration of the feature generation methods. The green lower reads represent the subset of reads considered in **percent lower** and **mean lower**. The probability density is used to score each read to calculate **mean loglikelihood**. **D**. Illustration of the k Nearest Neighbor model. To classify the targeted locus, a majority vote weighted by distance is performed among its 100 closest neighbors in the training dataset.

PREDICTION RESULTS

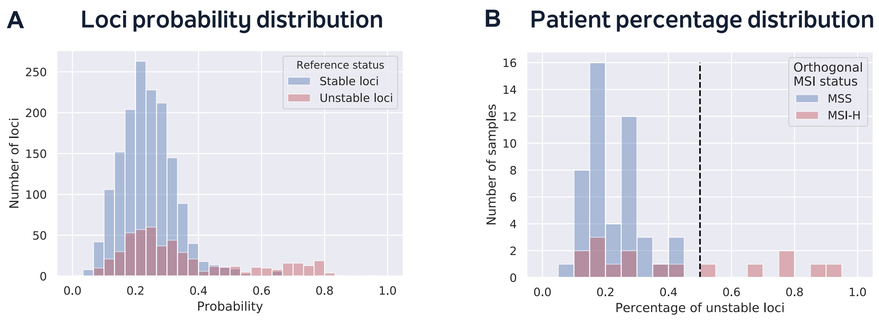

Figure 2. A. Distribution of the predicted probability of the validation dataset loci. Our model detects unstable loci with high specificity. The reference status of each locus is determined through comparing distribution of repeat units in sequencing results of patient's solid tumor and blood cells (germline). **B**. Distribution of the percentage of unstable loci in the validation dataset patients. We detected microsatellite instability in 37.5% (6/16) patients at 100% (6/6) positive predictive value. Orthogonal MSI status of the patient was determined by MMR IHC or Tempus' clinically validated solid tumor MSI test.

LIMIT OF DETECTION

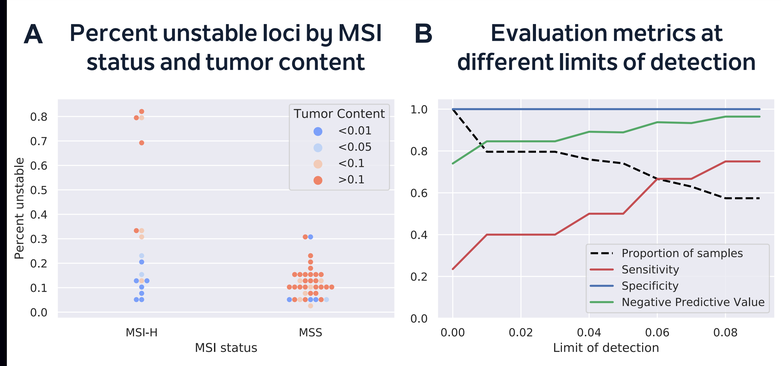

Figure 3. A. Percent unstable loci by MSI status and tumor content. 70% of the false negative samples have a tumor content inferior to 6%. Tumor content is estimated using Low Pass Whole Genome Sequencing and IchorCNA (Adalsteinsson, *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8,** 1324 (2017)) **B**. Sensitivity, Specificity and Negative Predictive Value at different limit of detection thresholds. For each threshold, only samples with a higher tumor content than the limit of detection are considered. By setting a limit of detection at 6%, our assay reaches a sensitivity of 66% and a Negative Predictive Value of 94%.

CONCORDANCE WITH SOLID TUMOR ASSAY

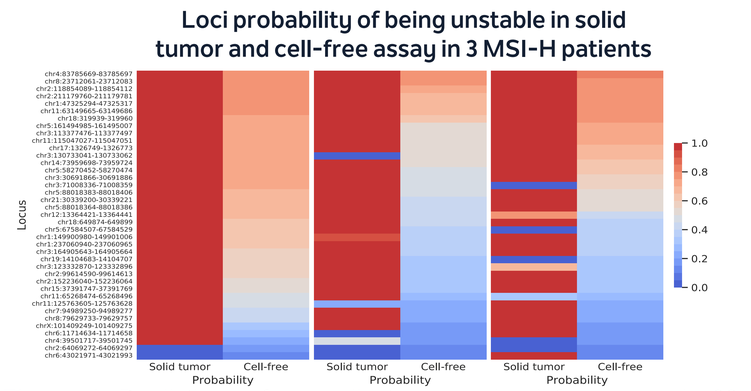

Figure 4. Predictions of our solid tumor and cell-free assay in 39 microsatellites loci of 3 MSI-H patients with more than 6% tumor content in their cell-free sample. All but two loci classified unstable by the cell-free assay are also unstable by the solid tumor assay, indicating the Tempus xF cell-free assay detects unstable loci with high specificity.

CONCLUSIONS

These results demonstrate the ability of our assay to detect MSI in cfDNA with high specificity, providing a transformative opportunity to report a clinically actionable insight alongside other somatic changes detected from cfDNA.

Sensitivity of our assay is improved by setting a limit of detection of 6% tumor content. Increasing the size of our training cohort will further improve sensitivity.