

## US - F, value in segment

### Tracks

1. Track 1: unmarked segment
2. Track 2: function

### Question

In the unmarked segments of track 1, is the average value of the function in track 2 smaller/different/larger than expected by chance?

Comment:

- We assume the function in track 2 is fixed and that the segments in track 1 are independent of the function values in track 2 under the null hypothesis. The segments in track 1 are preserved and randomized with different algorithms in order to determine whether to reject the hypothesis.
- Significance is determined by means of p-values. Small p-values identify bins where the function values are smaller/different/larger than expected in the segments of track 1.
- The p-values are found by simulation.

### Bins

The genome (or the areas of the genome under study) are divided into small regions, called bins. The tests are performed in each bin.

### Hypothesis tested

For each bin  $i$  we have one null hypothesis

$\mathbf{H}_0$ : *In the unmarked segments of track 1, the average value of the function in track 2 is the same as the average function value in the bin.*

There are three alternative hypotheses:

$\mathbf{H}_1$ : *In the unmarked segments of track 1, the average value of the function in track 2 is smaller than the average function value in the bin.*

or

$\mathbf{H}_2$ : *In the unmarked segments of track 1, the average value of the function in track 2 is different than the average function value in the bin.*

or

$\mathbf{H}_3$ : *In the unmarked segments of track 1, the average value of the function in track 2 is larger than the average function value in the bin.*

## Statistics and rejection of the null hypothesis

Let  $X$  be the average function value in track 2 evaluated in the base pairs inside segments of track 1. The distribution for this test statistics is not know for any of the permutation and randomization of the segments of track 1. It is necessary with MC simulation in order to decide whether to reject the hypothesis.