**Introduction**

Are these two features correlated?

Some measure is necessary!!

The idea of Kernel

Allows to account for correlation of neighbour positions rather than the same only!

**Problem genome is too long!!!**

We have used the value of the distance between the features.

Some control is necessary!!

**Correlation**

Cohesine vs CTCF

H3K36me3 vs H3K4me1

**Thee-way correlation: projection.**

Projection correlation is intended for analysis correlations of two profiles \(fg\) with exclusion of correlation of these profiles with third one (confounder)

\[
f(x) = f(x) - a(x) \leq af 
\]

Correlation of H3K4me3 with mRNA-Seq, both projected on H3K4me1:

**Tool**

- Very fast (3 min per genome)
- Works with quantitative and qualitative data
- The kernel-based approach allows complex geometry (shifts, smoothing, etc)
- Along with predefined kernel, calculates the results for set of standard shifts.
- Produce correlation track that can be used as input for further correlation (liquid association)
- Allows also to scale and sum profiles and compare profile combinations

http://stereogene.bioinf.fbb.msu.ru

https://github.com/favorov/stereogene

**Tissue clustering (the tree)**

- 9 marks, 111 tissues
- For each mark we build a distance matrix, based on pairwise correlation between all tissues.
- Based on the matrix, we build hierarchical cluster tree
- For each pair of tissues, we count the maximal level of common subtree containing them both (divergence level), or the minimal path length inside the tree
- To count the mean divergence level for the pair of tissues, we average the DL in all trees that contain the pair
- Then we build the new distance matrix
  1) mean(trees) level and run the hierarchical clustering again

**Method**

Correlation:

An equivalent transformation of the previous integral:

We can use another kernel function instead of delta-function:

For the whole genome: Mann–Whitney U test p-value

For each real frame: permutation p-value and FDR q-value.

Cross-correlation function on coordinate shift shows typical picture of mutual positioning of two tracks

**Used marks:**

H3K4me1, H3K4me3, H3K9me3, H3K9ac, H3K27me3, H3K27ac, H3K36me3
Problem: genome is too long!!!

We have a set of values!
Q1 Q2 Q3...
Are the features correlated?
Some control is necessary!!!