

Test for clustering of 3 loci

Here we consider testing whether 3 genes (A, B, and C) cluster together in a diploid nucleus. As in the case of two genes (5.2.3), we defined a “distance measure” for heterologous clustering of genes. If heterologous clustering occurs, we would expect at least one of the triangles of the type A_1BC and one of the type A_2BC to be small. Let $f(XYZ)$ denote the area of a triangle with vertices at the points X, Y and Z.

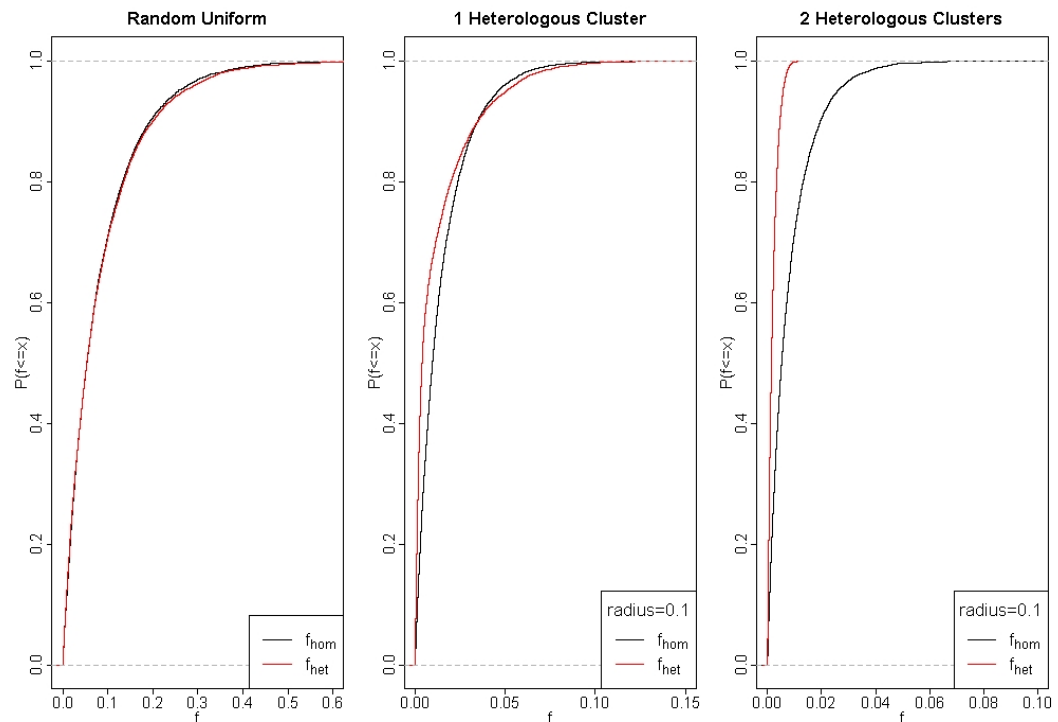
We use

$$f_{het} = 0.5 \left(\min(f(A_1B_1C_1), f(A_1B_1C_2), f(A_1B_2C_1), f(A_1B_2C_2)) + \min(f(A_2B_1C_1), f(A_2B_1C_2), f(A_2B_2C_1), f(A_2B_2C_2)) \right)$$

to measure the proximity within strictly heterologous triplets of genes and

$$f_{hom} = 0.25 \left(\min(f(A_1A_2B_1), f(A_1A_2B_2), f(A_1A_2C_1), f(A_1A_2C_2)) + \min(f(B_1B_2A_1), f(B_1B_2A_2), f(B_1B_2C_1), f(B_1B_2C_2)) + \min(f(C_1C_2A_1), f(C_1C_2A_2), f(C_1C_2B_1), f(C_1C_2B_2)) \right)$$

to measure the proximity of mixed heterologous and homologous triplets of genes¹.



Cumulative distribution functions for f_{hom} and f_{het} for 10000 simulated spherical nuclei with radius 1.

As in the case discussed in section 5.2.3 (Comparison within the same cells), we can test the presence of heterologous clusters by testing whether the mean of f_{het} is significantly smaller than f_{hom} . For simulated data, the corresponding Wilcoxon-Signed-Rank test

¹ For convenience of computations, the area f of a triangle with side lengths x , y and z may be calculated using Heron's formula: $f = 0.25\sqrt{(x+y+z)(y+z-x)(z+x-y)(x+y-z)}$.

has a power of about 0.8 to detect the presence of two heterologous clusters with maximal distances between cluster members of up to 20% of the nuclear radius (significance level 5%, 10 replicates). The power of the t-test is only about 0.6.

Note that neither the Wilcoxon-Signed-Rank test, nor the t-test are very successful in finding a single cluster of ABC genes (max. distance as above) if the remaining three copies are located completely randomly within the nucleus: For simulated data of this kind, the corresponding tests only achieved a power of 20% (Wilcoxon) or 10% (paired t-test) (10 replicates, significance level 5%). Power is notably improved by testing whether the ratio f_{hom} / f_{het} is equal to 1 (power/sensitivity approx. 30 % (Mann-Whitney), 20% (T-test)) instead of testing whether the difference $f_{hom} - f_{het}$ is 0. Unfortunately this test is slightly anti-conservative (14% (Mann-Whitney) and 12% (t-test) false positives for a nominal level of 10%).

A permutation test analogous to the one described in 5.2.4 using the ratio f_{hom} / f_{het} has a better performance: its sensitivity for detecting a single heterologous cluster is substantially larger than that of the other methods (ca. 50% true positives for samples with 10 replicates, significance level 5%) and its sensitivity to detect two clusters is approximately the same as that of the Wilcoxon-Signed-Rank-Test.