Biostatistics

In-class exercise week 5
Topic: multiple testing, interpretation of a p-value histogram

A biologist has developed a cell based assay to investigate the effect of a certain compound on the gene expression of human blood cells. He extracted blood cells from 20 randomly chosen subjects and treated 10 blood samplest with the compound solved in DMSO and the other 10 blood samples only with DMSO. Then he performed a whole genome expression analysis and tested for each of the ~20'000 known human genes if it is significantly differently expressed in compound-treated blood cells compared to the solution-treated blood cells.
For each of the 20'000 tests he got one p-value. The distribution of the p-values is shown in the histogram below.

Use the histogram to roughly estimate the following quantities:

a) How many of the 20'000 genes were unchanged by the compound treatment?
   Briefly explain your procedure.

b) How many significant tests did he get when working with a significance level $\alpha=0.05$?

c) How large is the false discovery rate FDR (=proportion of false positives among significant test results)?