Biological Data Analysis

Carl Herrmann
IPMB - Universität Heidelberg

IPMB
Institut für Pharmazie und
Molekulare Biotechnologie

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

8. Power of a test

Reliability of statistical test

- A **reliable test** should have a small number of false-positives and false-negatives
- Increasing significance level leads to ???? false-positives and ??? false-negatives

	H_0 is valid	H_0 is NOT valid	
H_0 rejected ($p < \alpha$)	**False-positive (type 1 error)**	**True positive**	**test positive**
H_0 not rejected ($p > \alpha$)	**True negative**	**False-negative (type 2 error)**	**test negative**
	negative	**positive**	

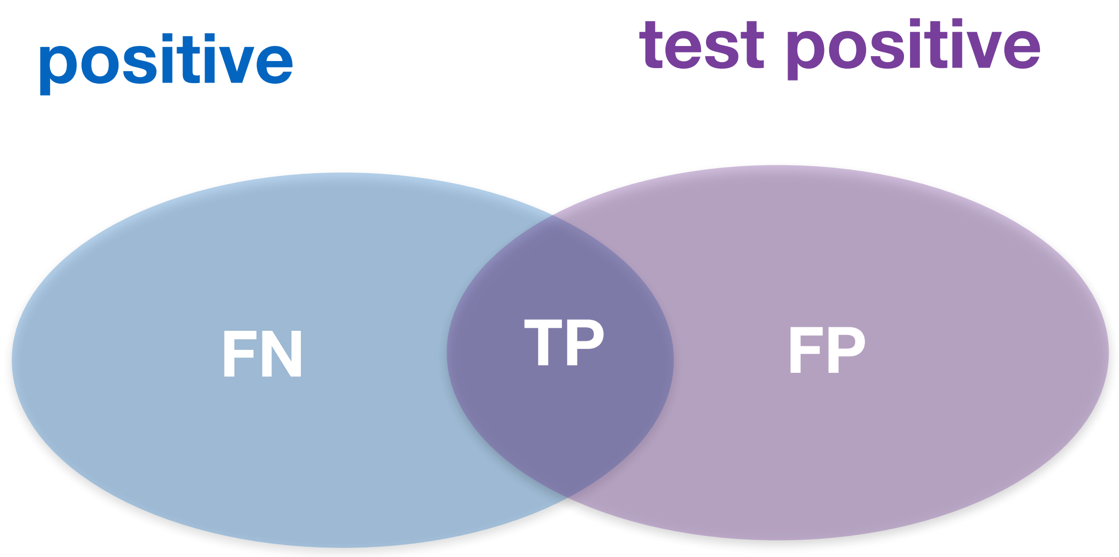
Reliability of statistical test

	H_0 is valid	H_0 is NOT valid	
H_0 rejected ($p < \alpha$)	FP	TP	test positive
H_0 not rejected ($p > \alpha$)	TN	FN	test negative
	negative	positive	

$$\text{false-negative rate (FNR)} = \frac{FN}{\text{positives}} = \frac{FN}{FN + TP}$$

$$\text{false-positive rate (FPR)} = \frac{FP}{\text{negatives}} = \frac{FP}{FP + TN}$$

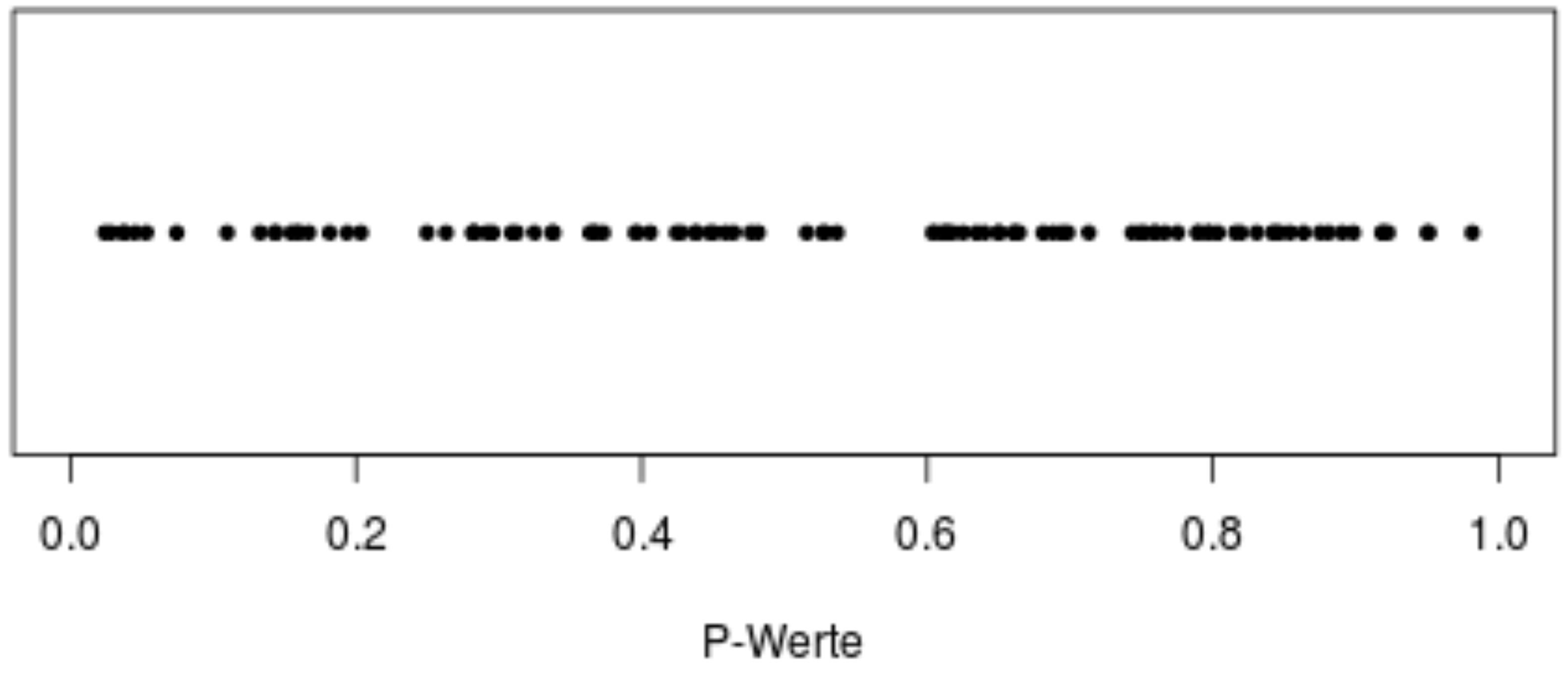
$$\text{false-discovery rate (FDR)} = \frac{FP}{\text{test positives}} = \frac{FP}{FP + TP}$$

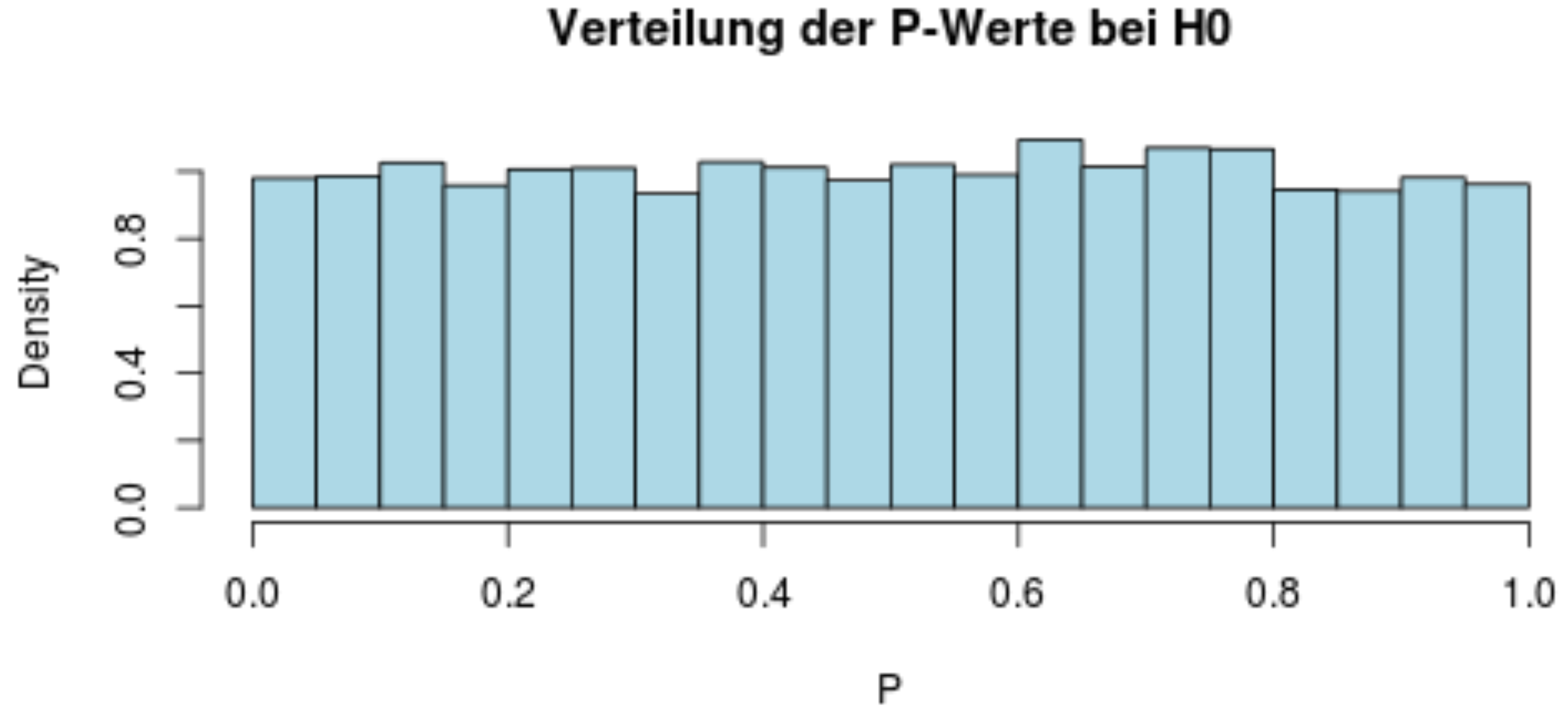


$$\text{precision} = \frac{TP}{\text{test positives}} = \frac{TP}{FP + TP} \qquad \text{recall} = \frac{TP}{\text{positives}} = \frac{TP}{FN + TP}$$

P-value distribution under H_0

- ***What are typical p-values under H_0?***
- **Experiment**: draw 2 sets (S_1 & S_2) of 50 random numbers each **from the same distribution**
- ***H_0: the expectation of both distributions are equal (TRUE!)***
- Compute t-test between S_1 and S_2, and determine P-value
- Repeat this experiment 1000 times, and plot the distribution of the 1000 p-values





Distribution of p-values under H_0 = uniform distribution

Type 1 errors

- **Red area:**

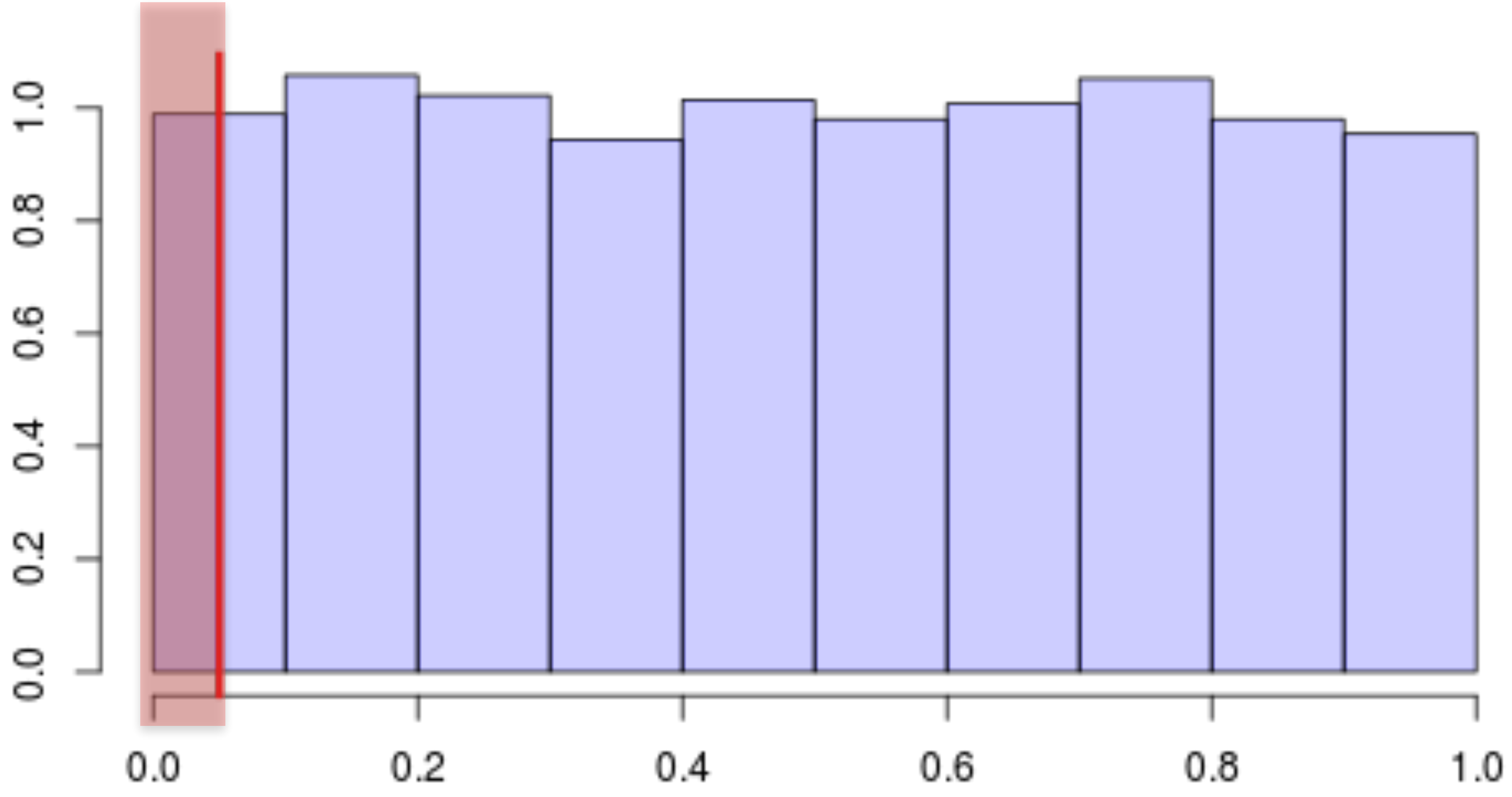
  with α = 5%, we would have wrongly rejected H0

  → ***FALSE POSITIVE***

- ***How often would that occur?***

  → red area compared to the total area = 5% because uniform distribution

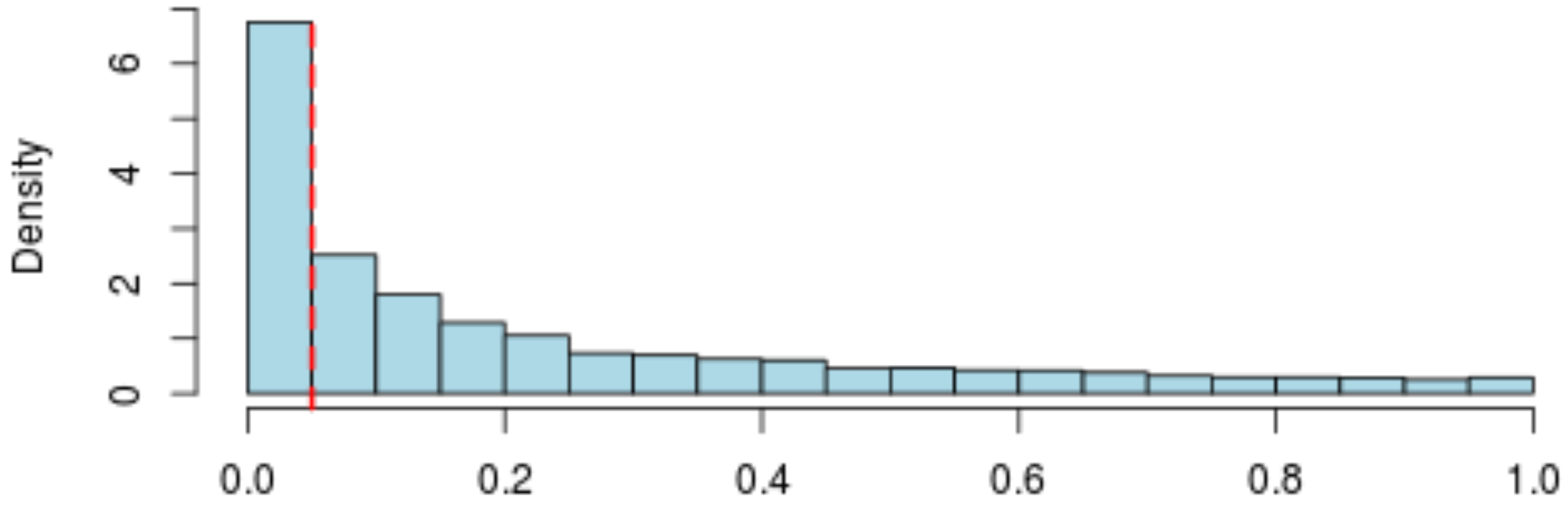
Distribution of p-values under H_0



α is the FALSE-POSITIVE RATE (FPR)

P-value distribution under H_1

- Experiment: draw 2 sets (S_1 & S_2) of 50 random numbers each **from two distributions with different expectation**
- ***H_0: the expectation of both distributions are equal (FALSE!)***
- compute p-value using a 2 sample t-test
- Repeat 1000 times and plot distribution of p-values



Many small p-values
→ H_0 would have been rejected

🙂

Some large p-values
→ H_0 would have NOT been rejected

☹

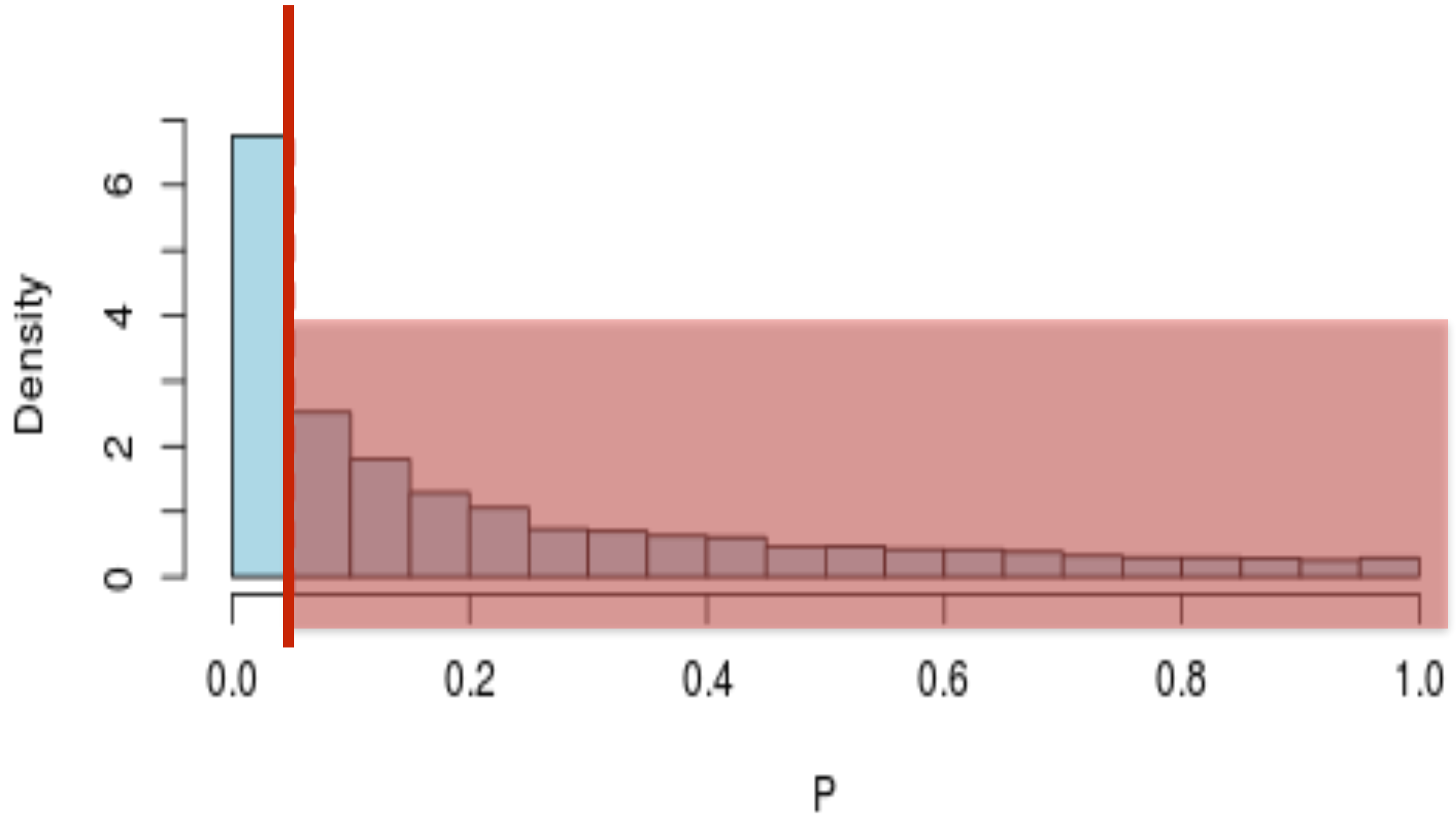
Type 2 errors

- Occur when a false H_0 hypothesis is **NOT rejected** by the test
  → False-negative (Type 2 errors)

- Probability of a type 2 error:
  β - value

- Probability for a type 2 error NOT to occur
  → **power of a test = 1- β**

α = 5%

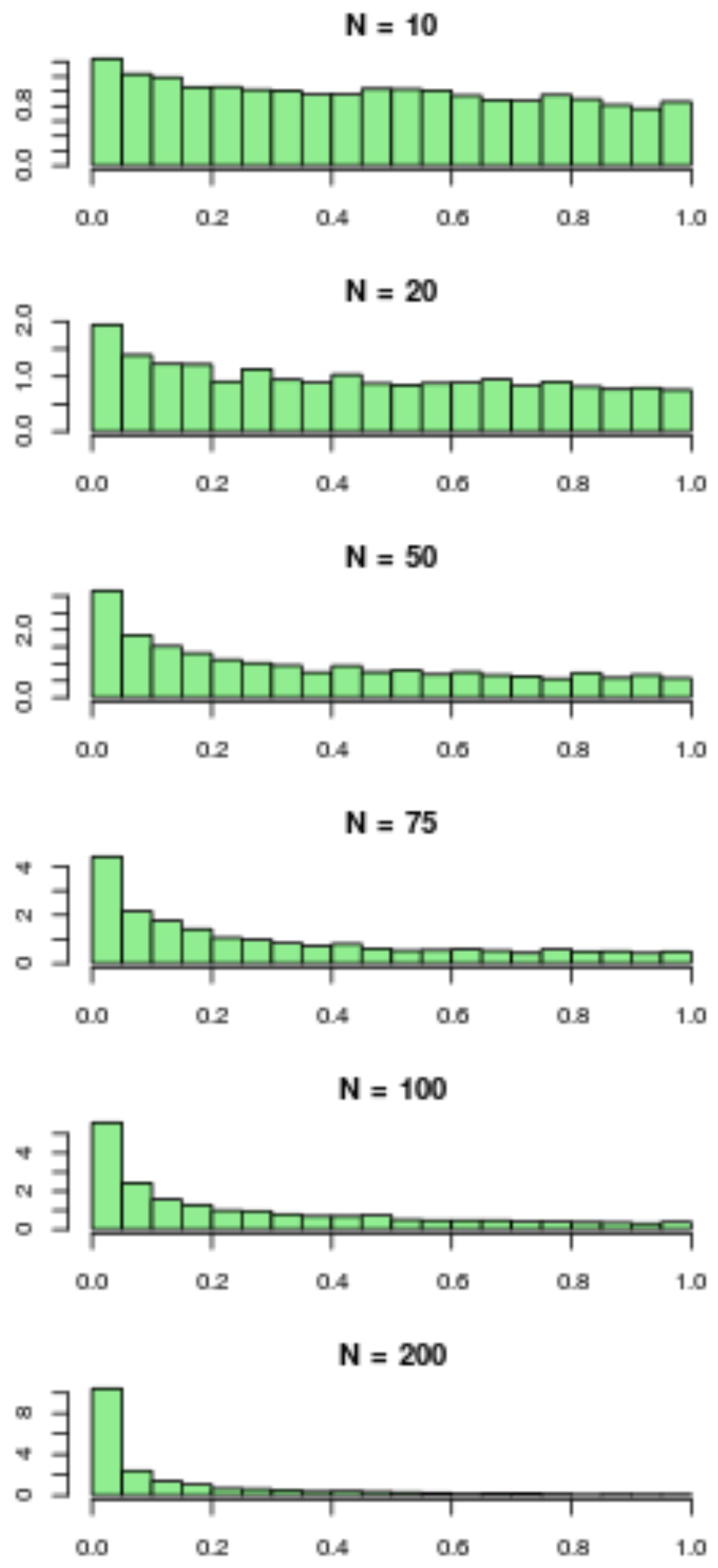


This area represents the cases for which H0 will not be rejected
→ ***false-negatives***

Power of a test

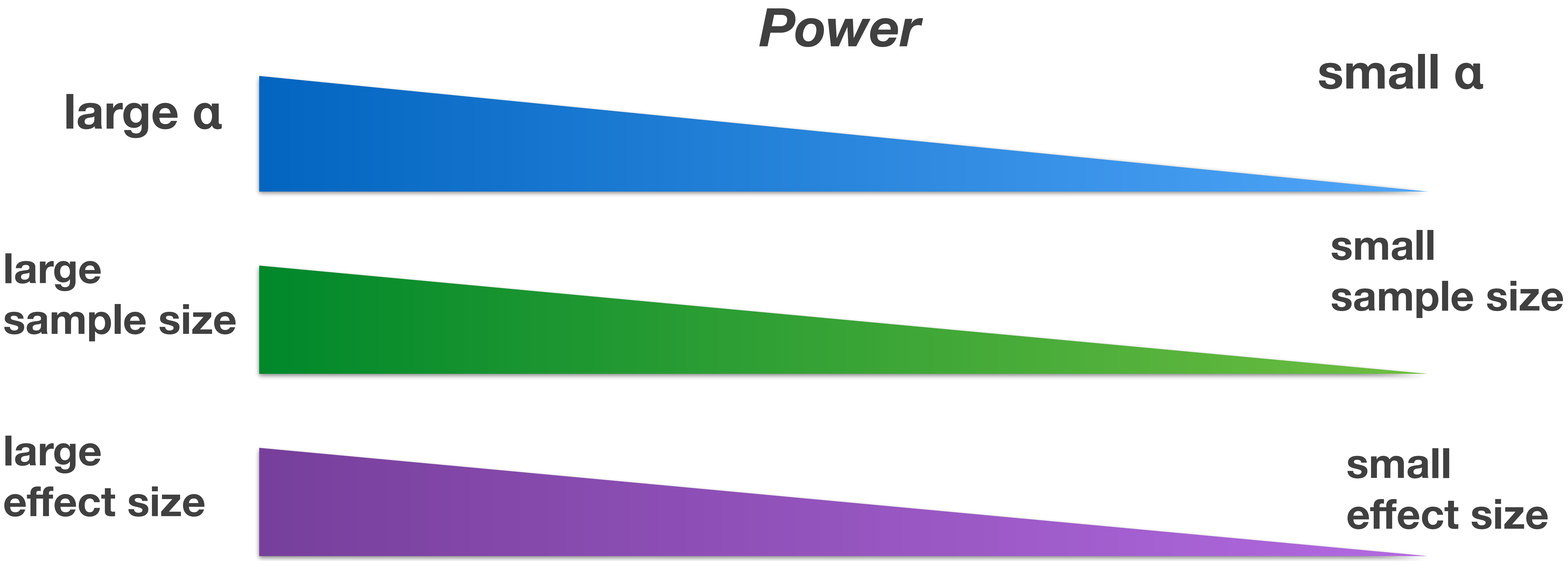
- Generate 2 datasets of length *n*
  - one from a normal distribution with mean 0
  - one from a normal distribution with mean 0.2
- **H_0: expectation of both underlying distributions is identical** (False!)
- perform t-test, compute p-values for various values of *n*

$$\beta \xrightarrow{n\to\infty} 0$$



Power of a test

- The power depends on:
  - **Significance level α**
  - **Sample size n**
  - **Effect-size**: how strong is the observed effect?



Power

large α

small α

large sample size

small sample size

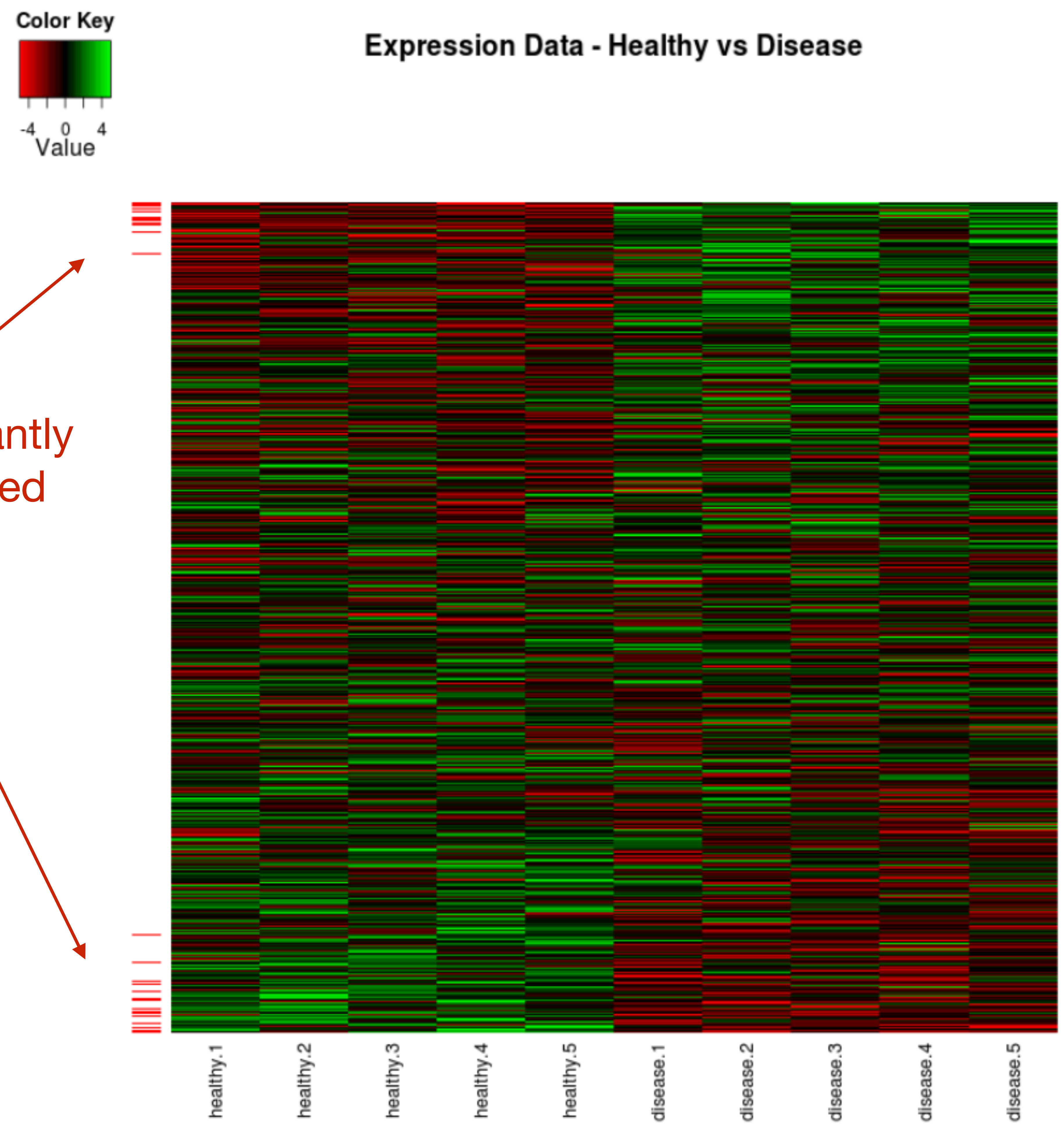
large effect size

small effect size

9. Correction for Multiple Testing

Gene expression data

- Finding differentially expressed genes between healthy and disease patients
- t-test with α = 5%
- H_0: non-significant expression difference between the two groups



445 significantly diff. expressed genes

Fake news ...

This dataset contains only random numbers

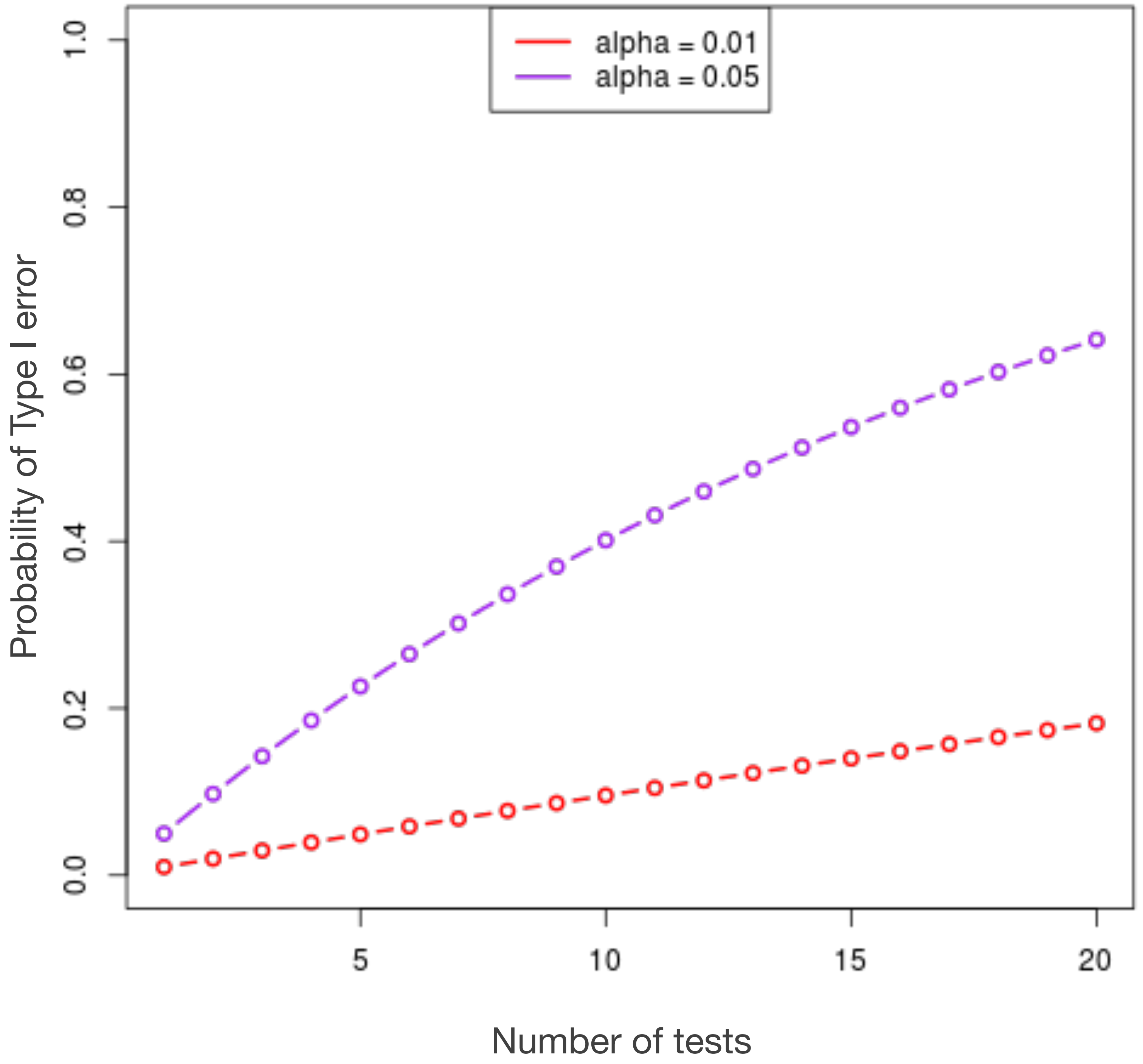
→ H_0 holds for all 10.000 “genes"

→ all the 445 genes are false-positives

```
X <- matrix(rnorm(n=100000,sd=3),nrow=10000)
```

Pitfalls of multiple testing

- We have repeated **10.000 independent tests**
- the p-value indicates the probability to obtain a more extreme test statistics if H_0 holds true
- α is the risk to call a positive event ("reject H_0") even if H_0 is true
- Probability of calling at least one false-positive across all tests:
  - 2 tests: $1-(1-\alpha)^2$
  - k tests: $1-(1-\alpha)^k$
  - 10.000 tests: $1-(1-\alpha)^{10000} \sim 1$

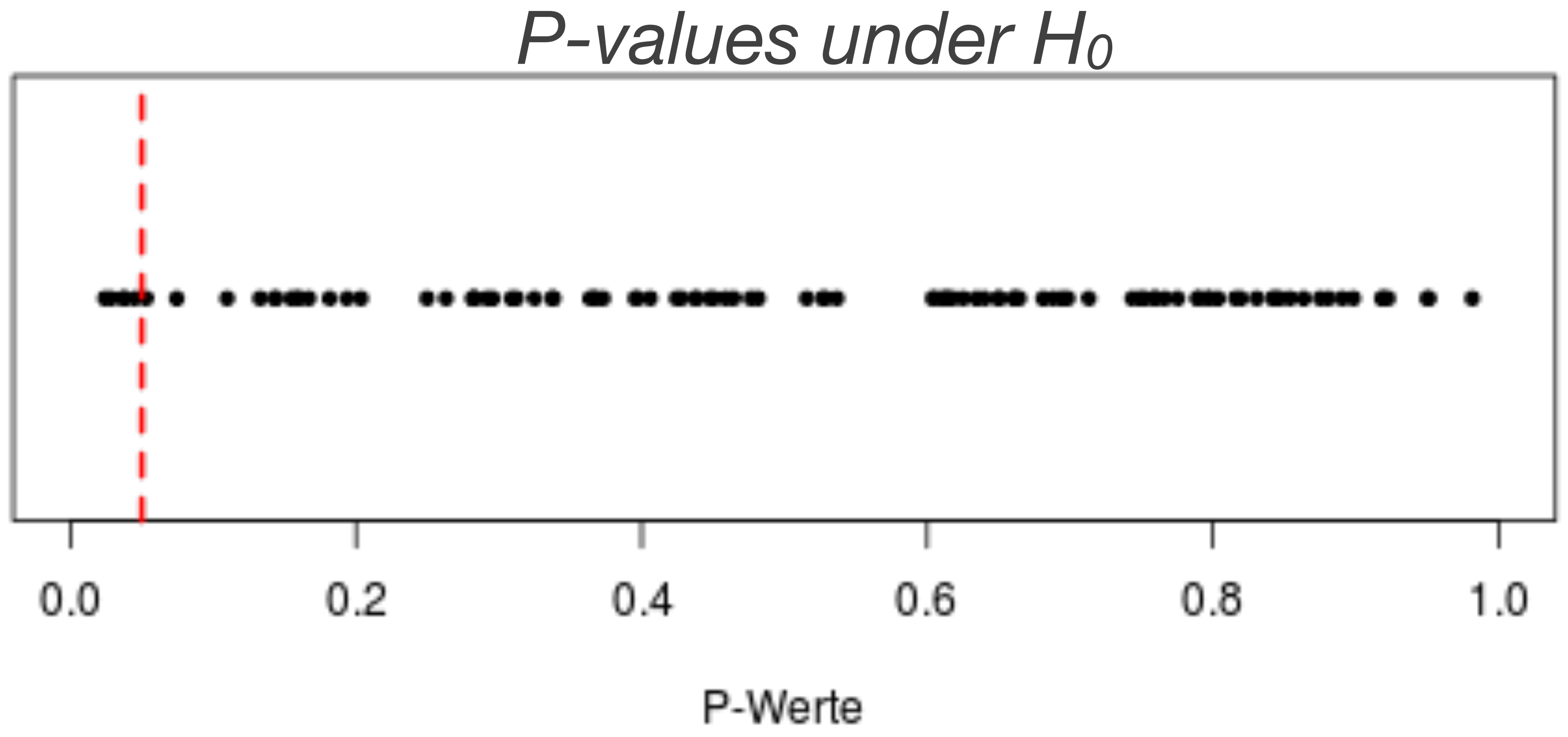


Beware of confusions!

- $1-(1-\alpha)^k$ is the probability to have at least one false-positive across all the tests = **family-wise error rate (FWER)**

- α is the **False Positive Rate (FPR)** i.e. the proportion of false positives if H_0 holds true

FWER = Probability to obtain at least one point below this threshold = $1-(1-\alpha)^k$



FPR = Proportions of tests below the threshold = α

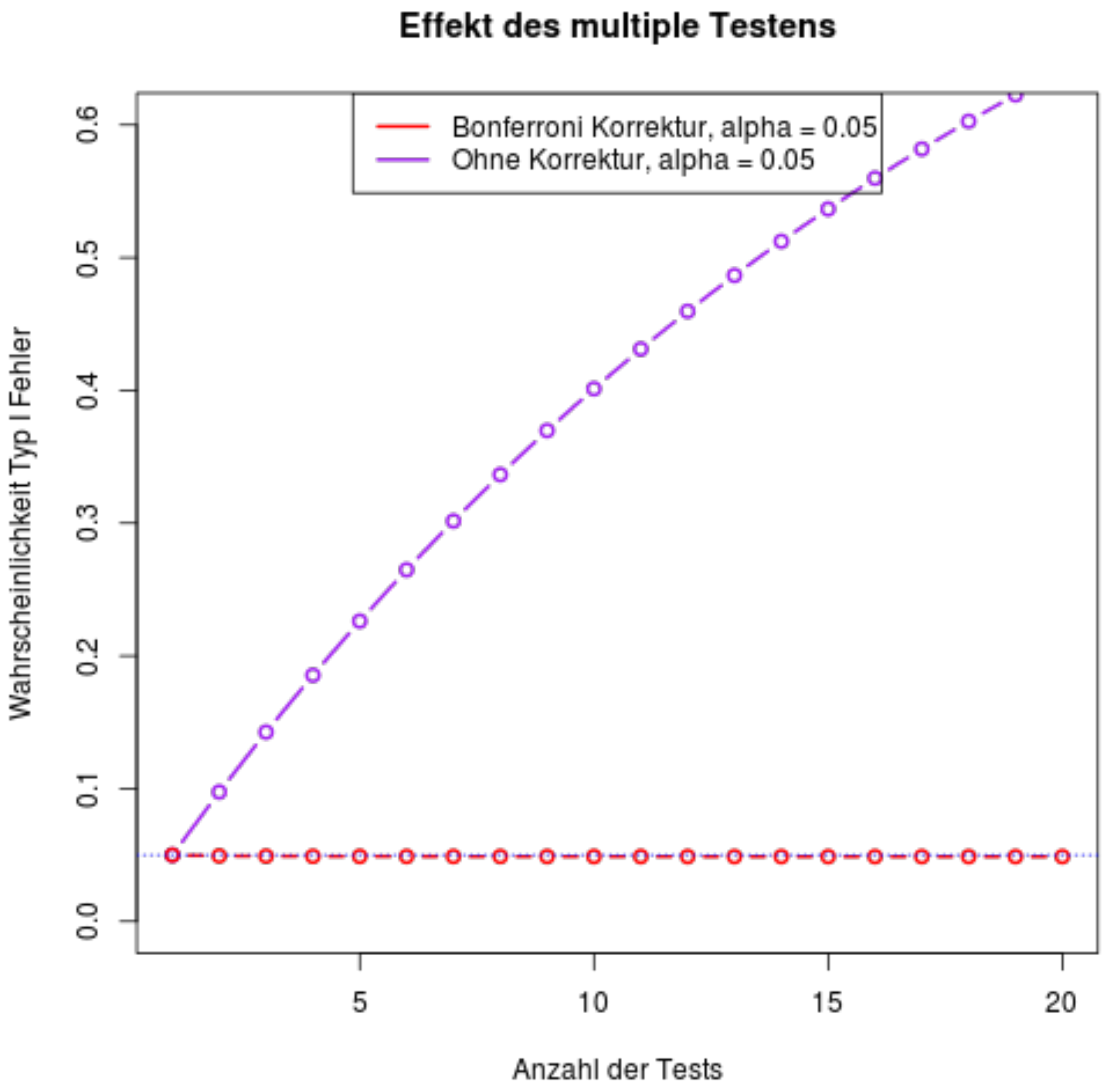
Type I errors

- Total number of tests
  = “Family" (m tests)
- Probability of a type I error over all tests
  = **Family wise error rate (FWER)**
  FWER = P(V > 0)
- Proportion of false positive reported to all negatives
  = **False positive rate (FPR)**
  FPR = V / m0
- Proportion of false positives reported to all significant ones
  = **False discovery rate (FDR)**
  FDR = V / R

	H_0 is valid	**H_0 is NOT valid**	
H_0 rejected ($p < \alpha$)	**V**	**S**	**R**
H_0 not rejected ($p > \alpha$)	**U**	**T**	**m-R**
	m_0	**$m-m_0$**	m

Control of the FWER

- **Bonferroni** correction
- adapt the significance level α to the number of tests
- when n tests are performed
  - $\alpha \rightarrow \alpha / n$
  - $p \rightarrow p_{adj} = \min(np,1)$
- **Probability of having a type I error remains constant at α**
- Very stringent correction!
  → increased type II error rate (false negatives)
- Example gene expression:
  - n = 10.000 tests
  - $\alpha = 0.05 \rightarrow \alpha / n = 5e\text{-}6$



Control of false-discovery rate (FDR)

- When a large number of tests is performed (typically for genomics data), Bonferroni correction is too stringent (too many Type II errors!)
- We can live with some false positives, as long as we can control their proportions within the significant test = false discovery rate (FDR)
- FDR = proportion of false-positives within the significant results
- FDR = 10% : 10% of the test which I consider to be significant ($p < \alpha$) are false positives