

# Mindless Statistics – a Freudian Unconscious Conflict and its Implications for GWAS



1910

**KNUT M. WITTKOWSKI**

The Rockefeller University Hospital  
Center for Clinical and Translational Research  
New York, NY 10065

2013

# Mindless statistics

Gerd Gigerenzer\*

*Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany*

---

## Abstract

Statistical rituals largely eliminate statistical thinking in ~~the social~~ sciences. Rituals are indispensable for identification with social groups, but they should be the subject rather than the procedure of science. The “null ritual” consists of three steps:

- (1) set up a statistical null hypothesis, but **do not specify your ... alternative hypothesis**,
- (2) use the **5% significance level** for rejecting the null hypothesis ..., and
- (3) always perform this procedure.

I report evidence of the resulting collective confusion and fears about sanctions on the part of students and teachers, researchers and editors, as well as textbook writers.

... no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.

Sir Ronald A. Fisher (1956)

Fisher, R.A., 1956. *Statistical Methods and Scientific Inference*. Oliver & Boyd, Edinburgh

## P-Values and Confidence Intervals: Two Sides of the Same Unsatisfactory Coin



**KNUT M. WITTKOWSKI**

The Rockefeller University Hospital  
Center for Clinical and Translational Research  
New York, NY 10021

June 28, 2012

### Confidence Level $1-\alpha$

#### Disadvantages of $\alpha = .05$

"Fisher: 'no scientific worker has a fixed level of significance at which in all circumstances he rejects hypotheses; he rather gives his mind to each particular case in light of his evidence and his ideas.'" [1956, p. 358]

0.050	0.025	0.408	0.753	0
0.040	0.020	0.402	0.695	0
0.030	0.015	0.409	0.658	0
0.020	0.010	0.405	0.611	0
0.010	0.005	0.400	0.574	0
0.000	0.000	0.416	0.577	0
0.050	0.025	0.471	0.680	0
0.040	0.020	0.465	0.644	0
0.030	0.015	0.459	0.788	0
0.020	0.010	0.453	0.743	0
0.010	0.005	0.445	0.680	0
0.000	0.000	0.436	0.650	0
0.050	0.025	0.472	0.810	0
0.040	0.020	0.466	0.769	0
0.030	0.015	0.461	0.724	0
0.020	0.010	0.455	0.689	0
0.010	0.005	0.448	0.654	0
0.000	0.000	0.440	0.619	0

# Who is to blame?

A smart graduate student told me that he did not want problems with his **thesis advisor**.

[As] a post-doc, his concern was **to get a real job**.

[As] an assistant professor ... he still felt he could not afford statistical thinking because he needed to publish quickly to get tenure. **The editors** required the ritual, he apologized, but after tenure, ... he would be a free man.

Years later, he found himself tenured, ... he had been asked **to teach a statistics course**, featuring the null ritual. He did. [*He could not prevent his students from comparing his teaching to his publications – kmw*]

# Neyman-Pearson Decision Theory

Gigerenzer (2004):

*“In Neyman–Pearson theory, the meaning of  $\alpha = 0.02$  is the following: If [a hypothesis]  $H$  is correct, and*

***the experiment is repeated many times,***  
*the experimenter will wrongly reject  $H$  in 2% of the cases ...*  
*[however,] in science, Fisher argued,*

***one does not repeat the same experiment again and again,***  
*as is assumed in Neyman and Pearson’s interpretation of*  
*the level of significance as an error rate*

***in the long run [of repeated experiments].**”*

Neyman-Pearson (1933):

*“We may search for rules to govern our behavior with regard to*  
*[hypotheses of a given type], in following which [rules]*  
*we insure that,*

***in the long run of experience [with different experiments],***  
*we shall not be too often wrong.”*

# The Unconscious Conflict

## Über-Ich / Superego (Neyman-Pearson)

Two hypotheses;  
alpha/beta determined by [risk of false decisions];  
sample size computed; aim is decision;  
no statements about truth of hypothesis

## Ich / Ego (Fisher)

Null hypothesis only;  
significance computed after the experiment; beta ignored;  
sample size by rule of thumb; [aim is to] get paper published;  
abundant epistemic statements about confidence in particular hypotheses

## Es / Id (Bayes)

[Continuum of hypotheses;]  
blocking the intellect from understanding that  $p(D|H) \neq p(H|D)$   
wishful thinking  
desire for [posterior believe as] probabilities of hypotheses

# Elements of Social Rituals

Elements of social rituals include (Dulaney, Fiske 1994)

- (i) the repetition of the same action,
- (ii) a focus on special numbers  
[5%],
- (iii) fears about serious sanctions for rule violations  
[teachers, mentors, editors, ...],
- (iv) wishful thinking and  
delusions that virtually eliminate critical thinking.

*Only in social sciences?*

# “Genome-Wide Significance”

$$0.05 / 1,000,000 = 5.00 \times 10^{-8} \quad (\text{Bonferroni 1936})$$

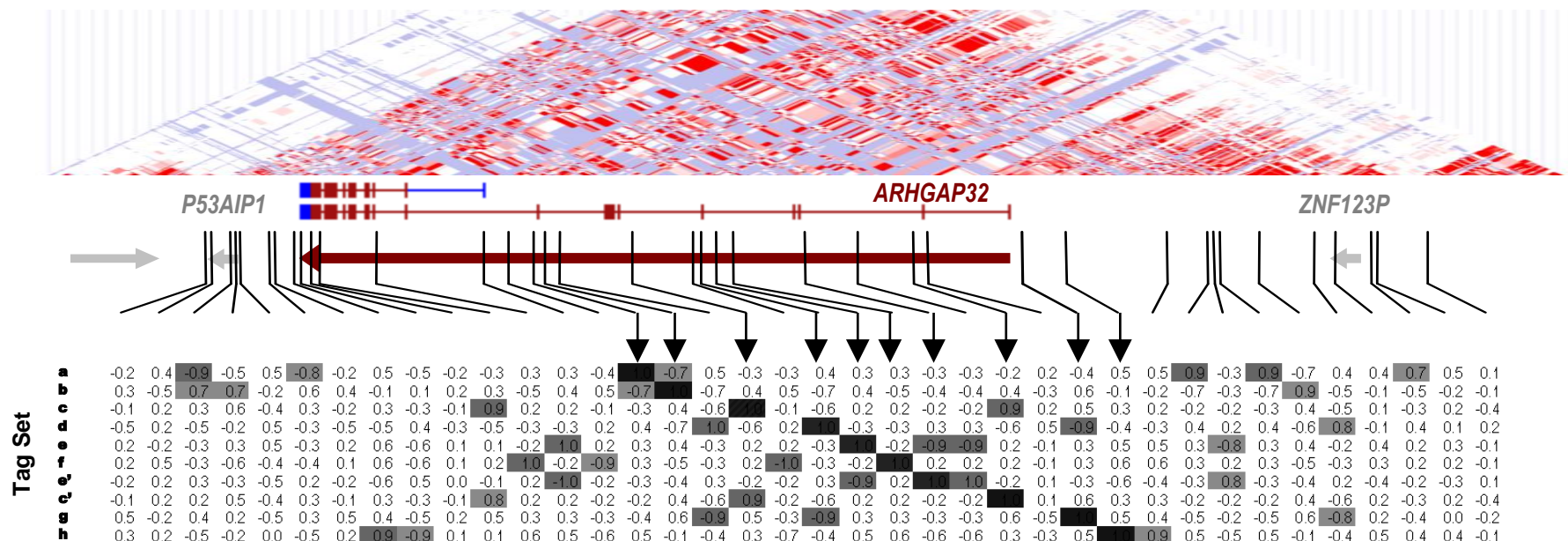
$$\approx 3.16 \times 10^{-8} = 10^{-7.5}$$

Why 1,000,000 ?

HapMap: 2,500,000 SNPs in the human population

300K Chip: 250,000 SNPs (83%) informative (NA, MAF, exp LD ...)

1M Chip: 600,000 SNPs (60%) informative (... obs LD, HWE, ...)





# “Genome-Wide Significance”

$$0.05 / 1,000,000 = 5.00 \times 10^{-8} \quad (\text{Bonferroni 1936})$$
$$\approx 3.16 \times 10^{-8} = 10^{-7.5}$$

Why 1,000,000 ?

HapMap: 2,500,000 SNPs in the human population

300K Chip: 250,000 SNPs (83%) informative (NA, MAF, exp LD ...)

1M Chip: 600,000 SNPs (60%) informative (... obs LD, HWE, ...)

LD blocks:  $\approx$  200,000 ‘tag sets’ of genetically indistinguishable SNPs

Why 0.05 ?

“no scientific worker” (Fisher 1958)  
Social ritual (Gigerenzer 2004)

Exploratory screening:  $\alpha > \beta$ ; power =  $1 - \beta = ?$

$$0.10 / 200,000 = 5.00 \times 10^{-7} \approx 10^{-6.5}$$

$$0.20 / 200,000 = 1.00 \times 10^{-6} = 10^{-6.0}$$

$$0.50 / 200,000 = 2.50 \times 10^{-5} \approx 10^{-4.5}$$

# Meehl's Conjecture (1978)

**Rationale:** Some difference exists between any natural groups. Therefore, with sufficient statistical power, one will almost always find a significant result.

**Conjecture:** In **non-experimental** settings with **large samples**,

- the probability of rejecting the null hypothesis of nil group differences in favor of a directional alternative is about 0.50.
- If we do not commit to the direction, we get [a significant result] virtually 100% of the time.

**Example:** Waller (2004) had access to data from >81,000 individuals. For each of 511 items (unrelated to sex) he postulated a sex difference in a randomly determined direction. 46% of these predictions were confirmed, often with very impressive p-values.

**Conclusion:** Large sample sizes give **low p-values**, but do not guard against model misspecification (**Type III Error**).

# Type-III Error

The attempt to give an “optimal” answer to the wrong question has been called “Type-III error”. [Kimball 1957]

*“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.”* [Tukey 1962, p. 13–14]


## Genetics:

Common diseases often have several risk factors within a functional intragenic region, each in LD with different markers (through recombination)

A “significant” result in a single SNP is more likely to be an artifact than an indication of disease association.

Meehl's  
conjecture

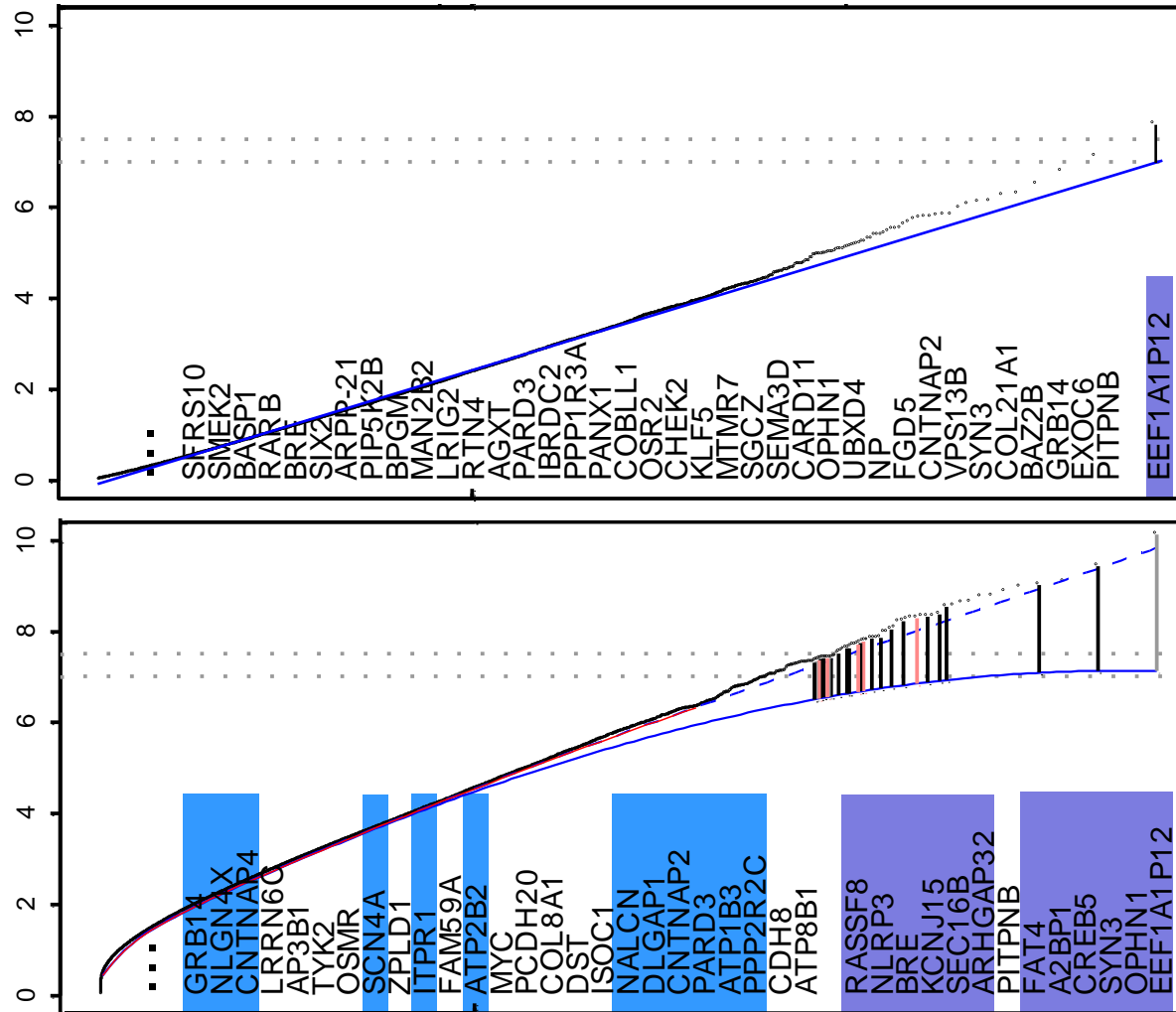
Small sample size  
Single-SNP GWAS

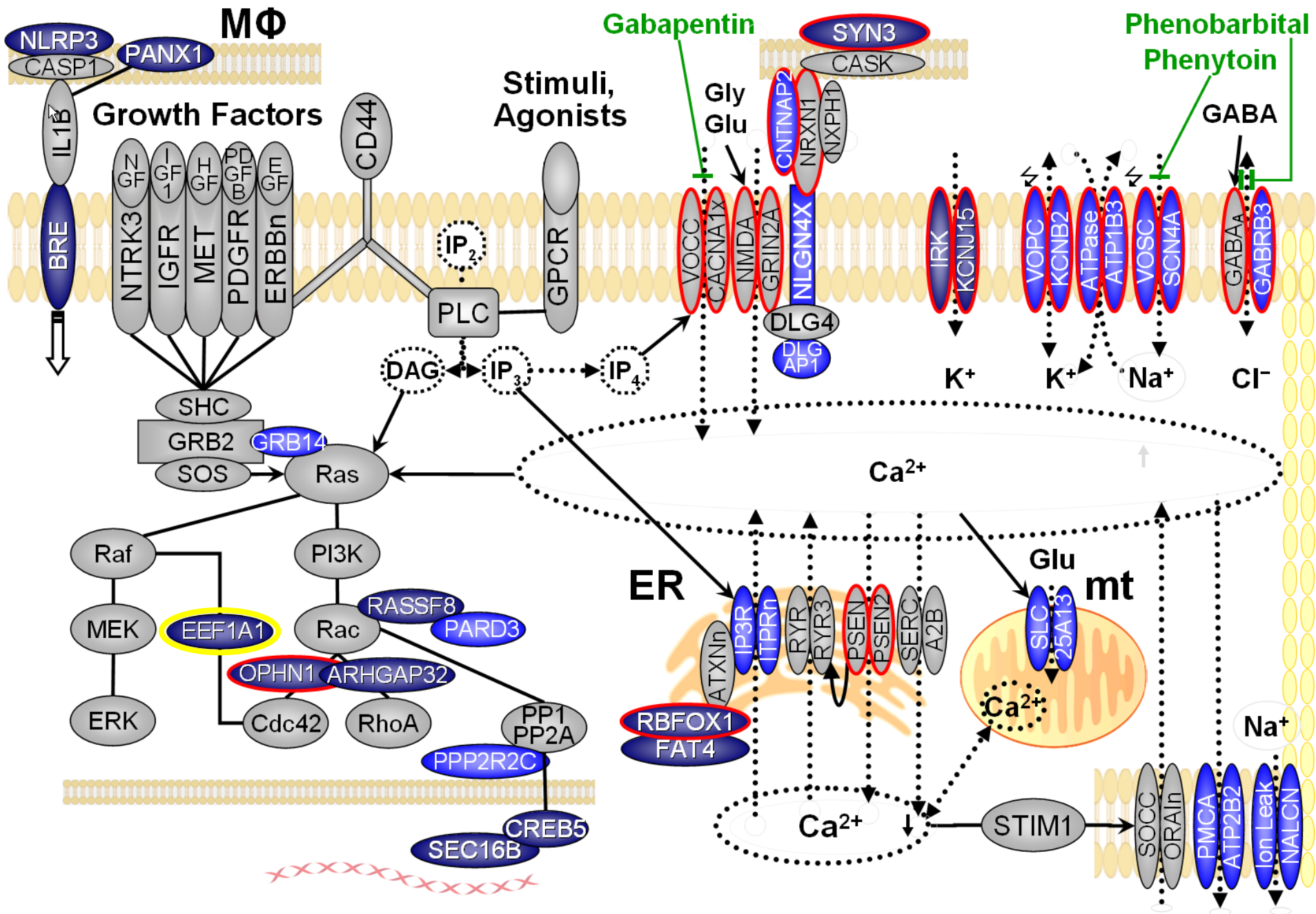


~~Larger sample size / smaller p-value~~  
wide-locus hypothesis / approx. result

# Single-SNP Hypothesis / Exact QQ Plots

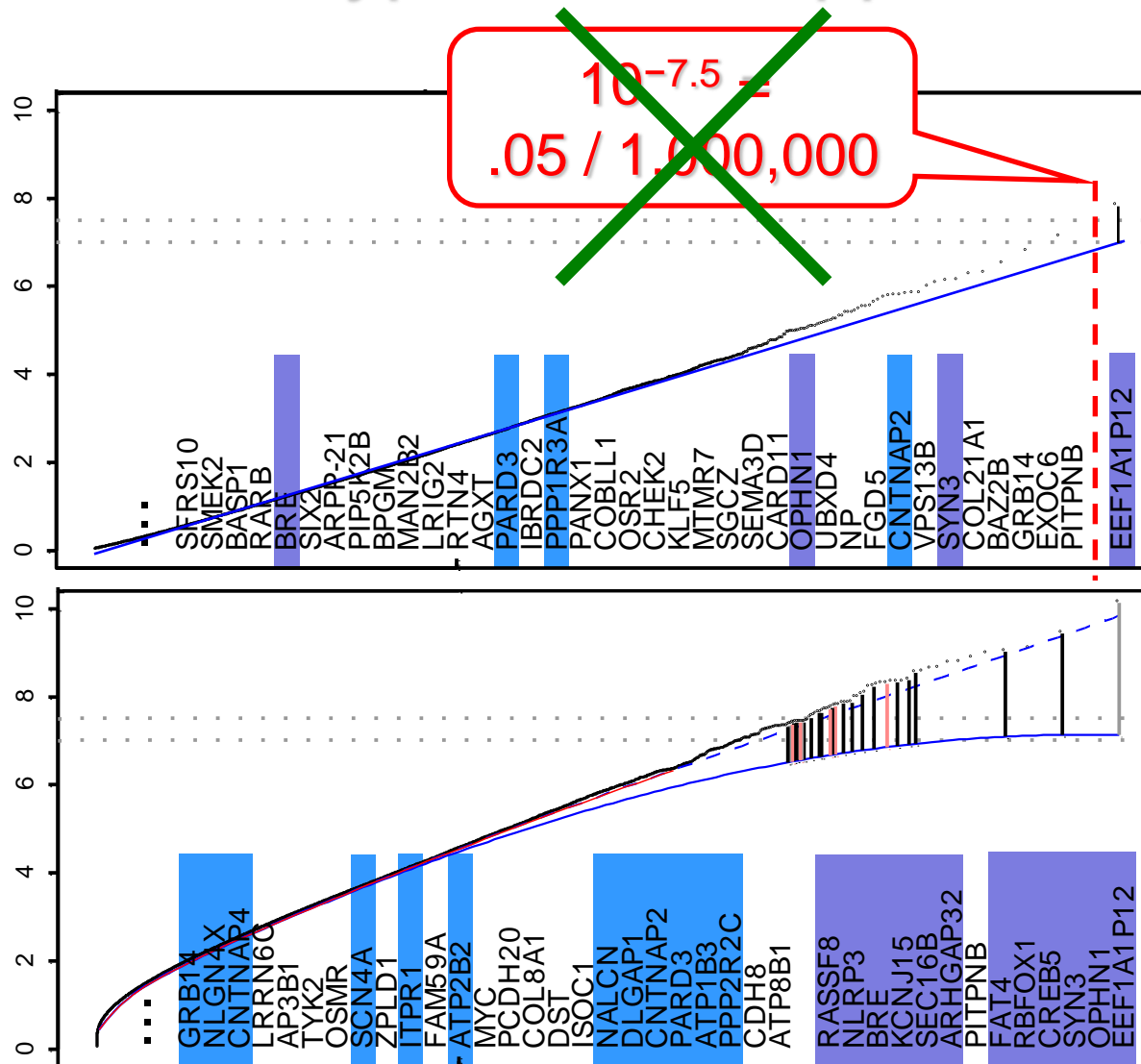
## Wide-locus Hypothesis / Approx. QR Plots



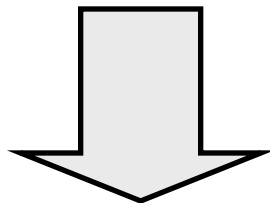


# Single-SNP Hypothesis / Exact QQ Plots

## Wide-locus Hypothesis / Approx. QR Plots

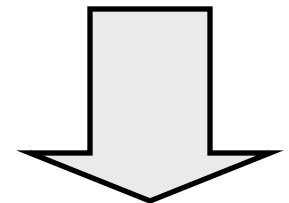


ssGWAS



μGWAS

confirmatory testing



selection procedure

WITTKOWSKI KM (2013) *Pharmacogenomics* 14: 391

