Why Gene Studies Use Such Tiny P-Values

Why do gene studies demand a far tinier p-value than 0.05?

CARRY FORWARD

Effect size and confidence interval show magnitude and precision; a p-value measures model compatibility, not truth.

10-YEAR TAKEAWAY

Many tests inflate false positives, so thresholds and independent replication must be planned before results are seen.

1. Observe the analogy before naming the biology

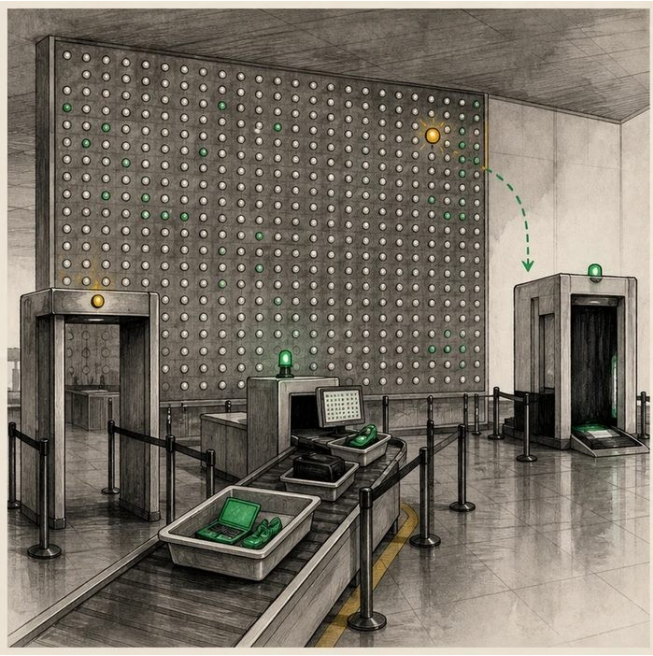


Illustration: Thousands of airport alarms require a stricter screen and a second check

1. What happens when thousands of sensors each can false-alarm?

2. Why set the alert rule before scanning?

3. What does the second scanner add?

2. Turn observations into rules

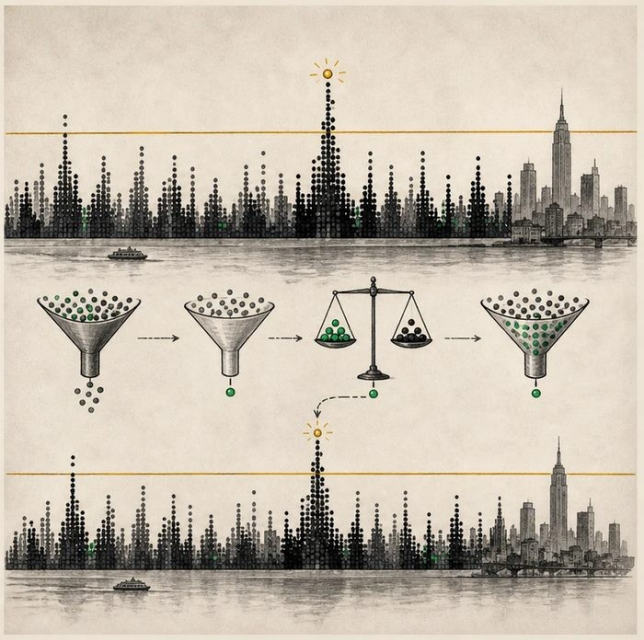
1. Rule 1: Count how many tests are being run. Explain it in your own words.

2. Rule 2: Predefine a correction or threshold suited to the analysis. Explain it in your own words.

3. Rule 3: Replicate the signal in independent data. Explain it in your own words.

Where the analogy stops: Genome signals are not threats, and statistical correction does not determine biological importance.

3. Map the rules onto the biomedical example



Biomedical teaching illustration: Control false positives in a common-variant GWAS

Analogy step	Biology step	How the relationship stays the same
Many sensors	Millions of variant tests	_____
Strict alert rule	Multiple-testing threshold	_____
Second scanner	Independent replication	_____

4. Use the case evidence

ID	Finding supplied in the case	Why it matters
EXP08-E1	Testing one million variants at p below 0.05 would produce many chance hits under the null.	A usual single-test threshold is not adequate for a genome-wide scan.
EXP08-E2	For many common-variant GWAS analyses, p below 5 x 10^-8 is a conventional threshold.	The convention is not universal for every genetic analysis.
EXP08-E3	The candidate peak repeats in an independent cohort with the same direction of effect.	Replication lowers concern that the first signal was sample-specific chance or error.

5. Make the clinical decision

You are the GWAS quality-control lead. A peak has p equal to 2 x 10^-6 in the discovery sample and no replication result.

A. Label it suggestive and require the planned threshold, quality checks, and replication.

B. Declare it significant because it is below 0.05.

C. Use 5 x 10^-8 as a magic rule for every genetic test.

Decision. Choose the report language and explain how multiplicity and replication change the claim.

Claim ceiling: You may apply the supplied common-variant GWAS convention. You may not treat that number as universal or proof of mechanism.

Claim. State the choice you can defend.

Evidence. Use these labeled IDs: EXP08-E1 + EXP08-E2.

Reasoning. Explain why the evidence supports the claim and stays inside the claim ceiling.

6. Reduce the lesson to one lasting idea

Take-home. Many tests inflate false positives, so thresholds and independent replication must be planned before results are seen.

1. Why is p below 0.05 too loose for one million tests?

2. What does replication add beyond a smaller p-value?

Glossary

Term	Plain-English definition
multiple testing	Running many statistical tests at once, which raises the chance of a false positive and requires a stricter cutoff to stay reliable.
false positive	A test result that signals a condition is present when it actually is not, a kind of error that can lead to needless worry or treatment.
Bonferroni correction	A math adjustment that makes the cutoff for significance stricter when you run many tests, so chance results are not mistaken for real ones.
genome-wide significance	A very strict p-value cutoff (about 5 in 100 million) used in genome-wide studies so that testing millions of spots does not produce false hits.
false discovery rate	The expected share of results called significant that are actually false alarms, a key check when many tests are run at once.
replication	Repeating a study to see if the result holds; a key test of whether a finding is real.

Next lesson unlock

Gene studies use a tiny p-value because a million tests breed a million chances to be fooled, and replication seals the deal. But even a rock-solid statistical link never proves IRF6 does anything to tissue. How would we leave the spreadsheet, go to the bench, and test what IRF6 actually does to a developing palate?

The journal links on the lesson page are optional. Everything required for this guided-notes set is already supplied here and in the case file.