

GWAS 4

Matti Pirinen
University of Helsinki
23.1.2019

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

THIS NEUTRINO DETECTOR MEASURES
WHETHER THE SUN HAS GONE NOVA.

THEN, IT ROLLS TWO DICE. IF THEY
BOTH COME UP SIX, IT LIES TO US.
OTHERWISE, IT TELLS THE TRUTH.

LET'S TRY.

DETECTOR! HAS THE
SUN GONE NOVA?

(ROLL)
YES.



FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT
HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$.
SINCE $p < 0.05$, I CONCLUDE
THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50
IT HASN'T.



BAYES RULE COMBINES
PRIOR & OBSERVATION

DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

THIS NEUTRINO DETECTOR MEASURES
WHETHER THE SUN HAS GONE NOVA.

THEN, IT ROLLS TWO DICE. IF THEY
BOTH COME UP SIX, IT LIES TO US.
OTHERWISE, IT TELLS THE TRUTH.

LET'S TRY.

DETECTOR! HAS THE
SUN GONE NOVA?

(ROLL)
YES.



FREQUENTIST STATISTICIAN:

THE PROBABILITY OF THIS RESULT
HAPPENING BY CHANCE IS $\frac{1}{36} = 0.027$.
SINCE $p < 0.05$, I CONCLUDE
THAT THE SUN HAS EXPLODED.



BAYESIAN STATISTICIAN:

BET YOU \$50
IT HASN'T.



BAYES RULE COMBINES PRIOR & OBSERVATION

X = Sun exploded

Y = Detector says "Yes"

We know $P(Y | X) = 0.973$ and $P(Y | \text{not-}X) = 0.027$

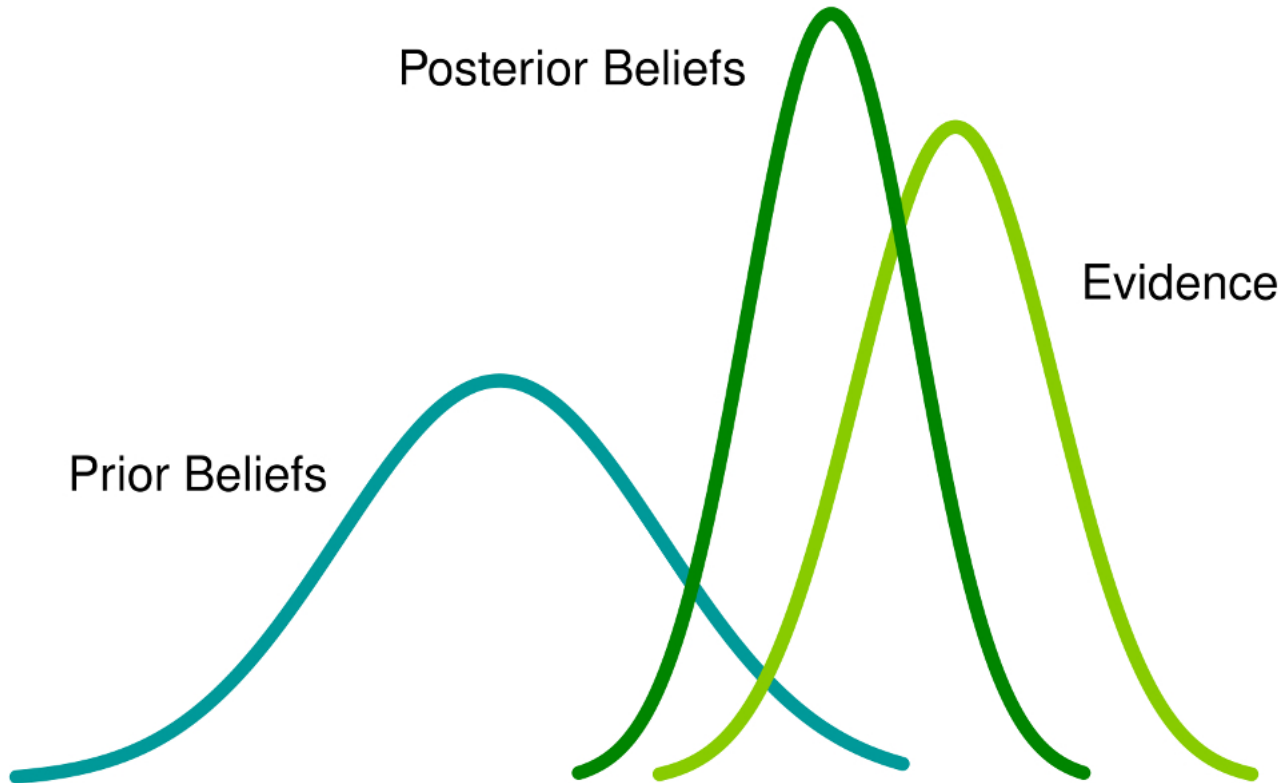
$$P(Y) = P(X) P(Y | X) + P(\text{not-}X) P(Y | \text{not-}X) \\ P(X) * 0.973 + (1 - P(X)) * 0.027$$

Bayes rule:

$$P(X | Y) = P(X) P(Y | X) / P(Y) \\ = P(X) * 0.973 / (P(X) * 0.973 + (1 - P(X)) * 0.027) \\ = P(X) / (0.0277 + 0.972 * P(X)) \\ \leq P(X) / 0.0277 \leq 40 * P(X)$$

So the observation increases probability of X at most 40 fold compared to prior probability, that is likely very very very small.

BAYESIAN INFERENCE



- We are estimating a parameter (like an effect size in GWAS)
- We have some prior beliefs where the parameter value is, but we don't know very accurately
- We gather data to learn about the parameter—this gives the evidence based on the gathered data alone
- Bayes rule tells how to consistently combine the prior beliefs and the evidence from the data into a combined posterior belief
- If prior is flat across a range of values (relative to the amount of evidence in data), then posterior will look like evidence in the data
- If prior of some region is extremely small, then we need a lot of evidence before posterior will support strongly that region

Figure:
<https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/>

BF VS P-VALUES

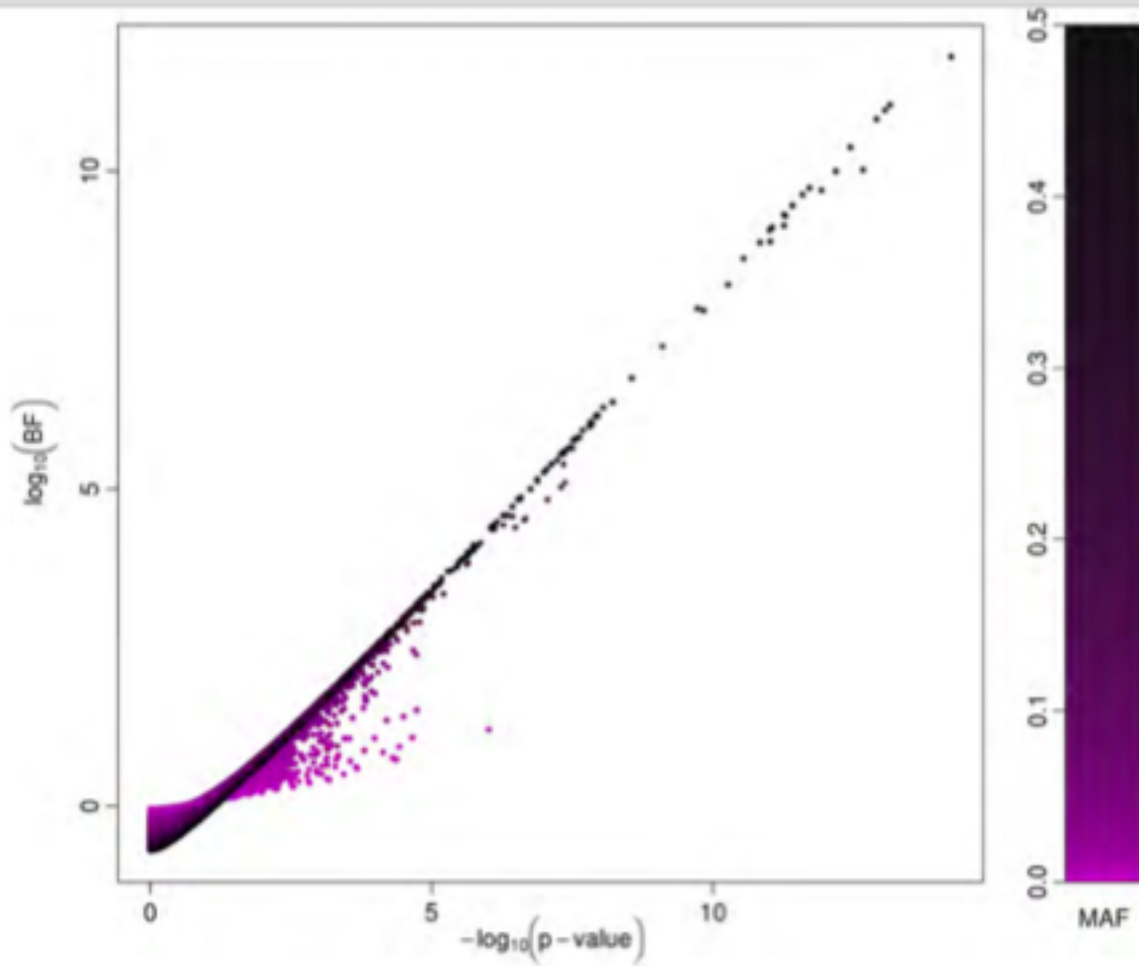
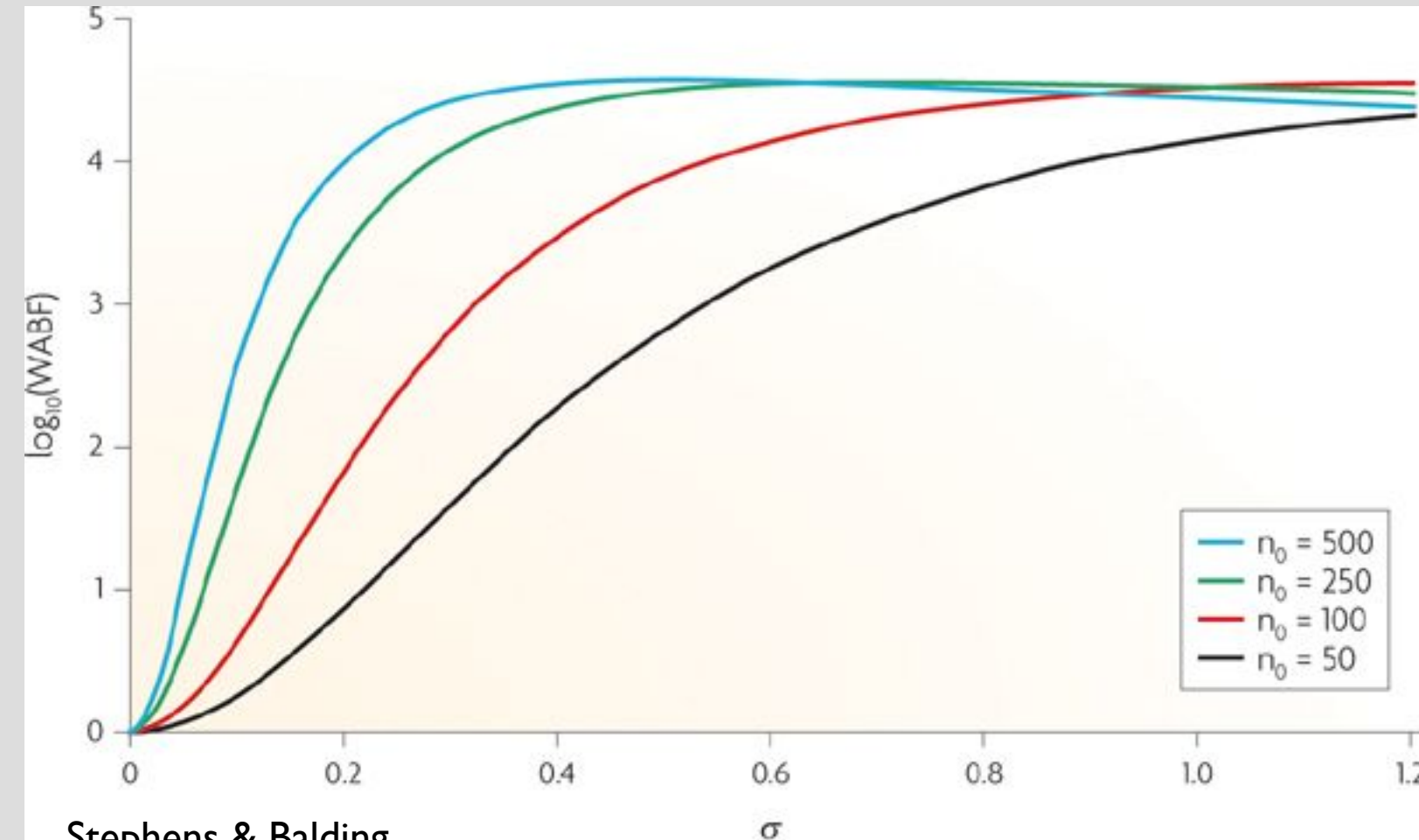


Figure 6.7: **BF versus p-value for Crohn's disease.** Each point represents a SNP from the WTCCC data. BFs are calculated under the conservative prior ($\sigma = 0.2$). Points are coloured according to the MAF, as shown in the legend on the right.

For common variants there is linear relationship Between P-value and BF.

Differences come for rare variants since the standard prior distribution does not allow large effect sizes.

BF FOR A P-VALUE 5E-7



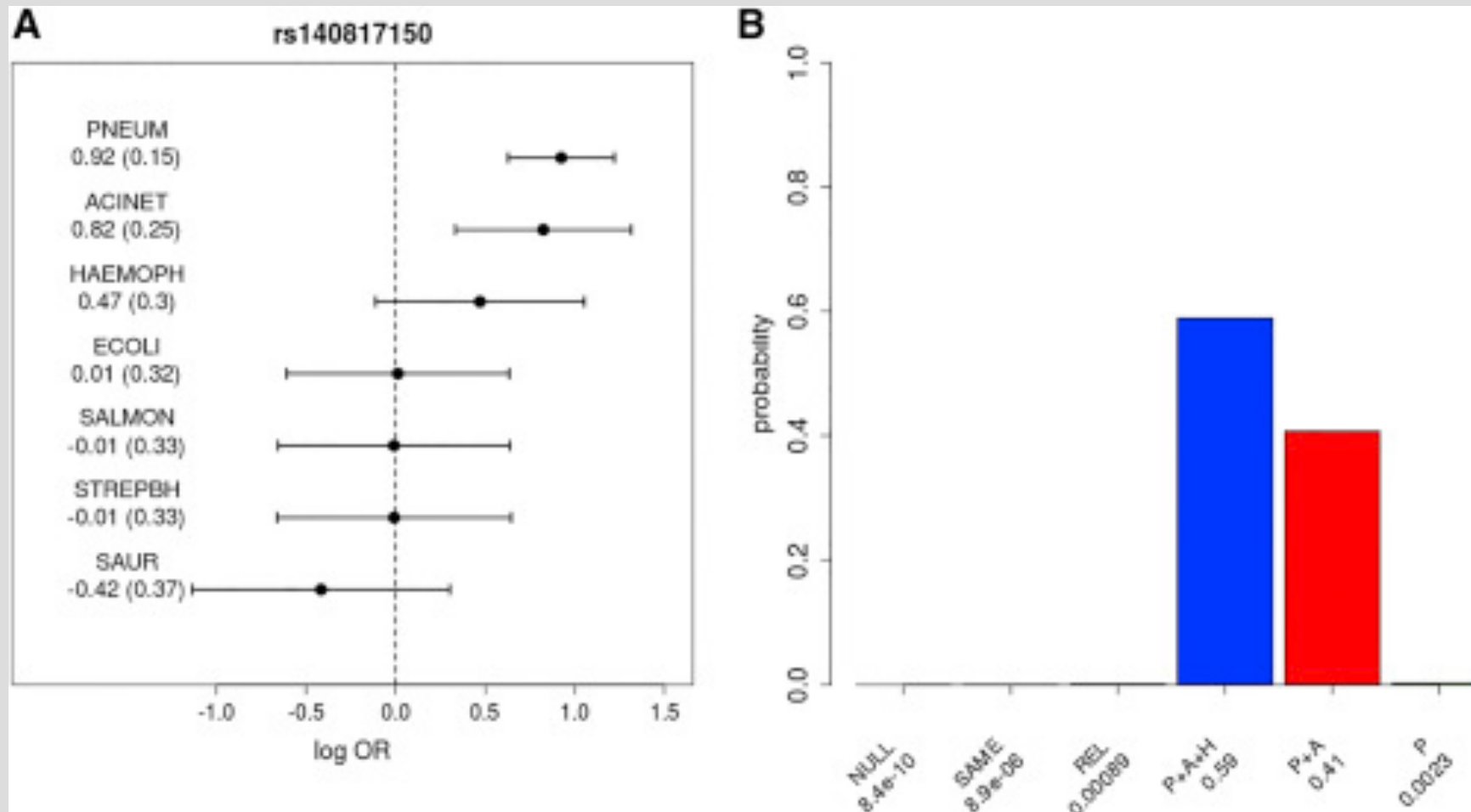
Stephens & Balding

Nature Reviews Genetics **10**, 681–690 (2009)

Nature Reviews | Genetics

The curves show the Wakefield approximate Bayes factor (ABF) for a SNP with a p -value $\approx 5 \times 10^{-7}$ using 4 values of n_0 , which is the minor allele count among cases and controls combined. There are n_0 cases and n_0 controls, so the minor allele fraction remains constant at 0.25. As σ (the standard deviation of the effect size) increases from 0, the $\log_{10}(\text{WABF})$ for each SNP rises from 0 to a maximum value of 4.57 before gradually decreasing as σ continues to increase. If $n_0 \geq 250$, the Bayes factors (BFs) vary by roughly one order of magnitude for $0.2 < \sigma < 1$, but when $n_0 = 50$, the BF varies more markedly, by several orders of magnitude for σ in this range. If $\pi = 10^{-4}$, then $\log_{10}(\text{ABF}) < 4.57$ implies $\text{PPA} < 0.79$. Therefore, under our assumptions, a SNP just reaching the p -value threshold of 5×10^{-7} still has a substantial chance of being a false discovery.

BAYESIAN MODEL COMPARISON



A SNP that associates with Bacetraemia in Kenyan children

The association seems present with several bacetria, but not all.

Bayesian model comparison using ABF framework let us to directly compare different models of association.

BAYESIAN MODEL COMPARISON

4 SNPs
Associated with
Ischemic stroke

Two SNPs
Particularly in
LVD
and 2 in
CE

Subtypes:
LVD large vessel
SVD small vessel
CE cardioembolic

