

GENETICS

# 50% OF ALL DISEASE GENES FOUND, REALLY?

9.03.2011

I have been reading now many times in nature genetics that a few newly found SNPs explain about half of the attributable risk by genes while I fear that this probably mixes up different epidemiological concepts.

The population attributable risk is usually defined as the reduction in incidence that would be observed if the population were entirely unexposed. This cannot be meant as I don't know of any genetic study examining incidence so far.

Maybe it's the etiologic fraction? Given as the proportion of the cases that the exposure had played a causal role in its development? Also this cannot be meant as usually everybody has his own combination of SNPs; any other combination may not be causal at all. Finally, it is probably something like explained variation (the proportion to which a regression model accounts for the variation of a given data set). Sorry, and here comes my discomfort, this measure CANNOT BE EQUALED to 50% of the genes found responsible for a disease. There are two reasons, the first is obvious and already in Wikipedia

Thus  $R^2$  gives the 'percentage of variance explained' by the regression, an expression that, for most social scientists, is of doubtful meaning but great rhetorical value. If this number is large, the regression gives a good fit, and there is little point in searching for additional variables.

My second argument is rather obvious too: Nobody knows what these associated SNPs are tagging. Really a gene in physical proximity? Or just a particular binding site operated by one or several transcription factors? Or is it spurious stratification? Or just a chance finding?

We will see, what this 'explained variance' really means in terms of biology. At least, definitely not what all these nature genetics papers want you to tell every month.

