Rare, not common variants hold the strongest asthma risk

Wjst ,Munich, 28-10-2010

The recent consortium-based analysis of asthma associated SNPs represents the largest effort so far to understand the complex architecture of a heterogenous disease.

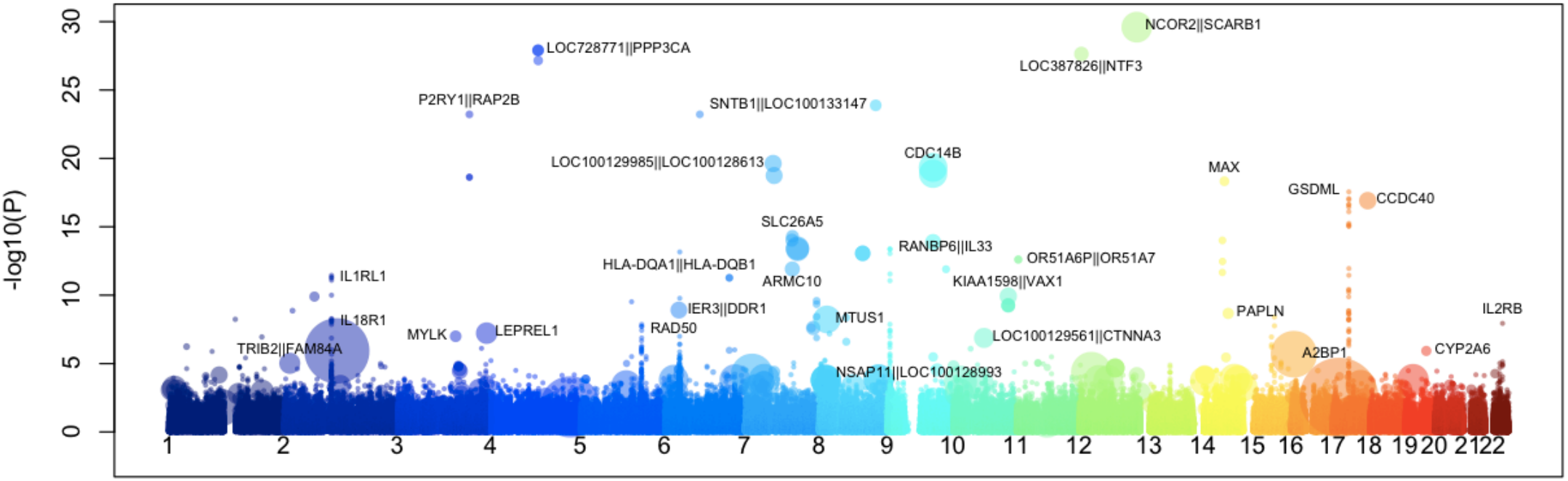
The paper published in this Journal (1) replicates mostly known associations (2) while the supplemental data file deposited at www.cng.fr/gabriel provides some further insights as it includes also those SNPs with a minor allele frequency <1% in controls that had been removed from the print version (1). Such an exclusion of rare variants (RVs) may have been a necessary quality control measure in the early days of genotyping but is now considered as being inadequate with an error rate of the Illumina chip genotyping system believed to be <0,01%.

As RVs are receiving an increased interest now in many complex diseases (3), (4), the results of the combined study sample were plotted without any restrictions (FIG.1). RVs are more than 7-fold enriched in the highly significant group, making up 45% of the 109 SNPs with $P<7.2x10^{-8}$. As there are virtually no LD "hitchhiking" SNPs (in contrast to CVs), RVs make up the majority of unique gene associations. Effect sizes found by RVs are also much larger than the reported CVs in (1) with strong odds ratios up to >50.

All RVs need to be confirmed in another assay, however, it is already interesting to note that many of the RVs are situated close to or in genes involved in lung pathology. Examples include DDR1, a tyrosine kinase receptor that is phosphorylated by collagen (OR 17.2), PERP, a component of intercellular desmosome junctions (OR 3.7), FOXP2, a regulator of lung development (OR 0.04) and CYP2A6, the primary enzyme responsible for the oxidation of nicotine (OR 0.1). Given the excessive risk by some RVs this raises even the possibility that mutations in these tagged genes could be leading to monogenic forms.

Unfortunately this makes the situation far less clear as "asthma genes" may sit everywhere in the genome and not just five regions (1). An imputation using the 1000 genome haplotypes is one possibility for identification of further RVs (4) while probably only full genome sequencing will lead to a better understanding of the complex genetic architecture of asthma.

FIG.1: Manhattan dot plot of the Gabriel GWA file deposited at www.cng.fr/gabriel *showing significance of association in the fixed effects model versus chromosomal position. Symbol size correlates to magnitude of the odds ratios if OR>1 or 1/OR if OR<1 and may be a better parameter for the relevance of results than the arbitrary threshold of a genomewide significance level (5). All annotation data were obtained by snpinfo.niehs.nih.gov*



References

1. *Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. Sep 23;363(13):1211-21.*
2. *Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet. 2009 Mar;41(3): 342-7.*
3. *McClellan J, King MC. Genetic heterogeneity in human disease. Cell. Apr 16;141(2):210-7.*
4. *Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. Oct 28;467(7319):1061-73.*
5. *Ziliak ST, McCloskey DN. The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. University of Michigan Press. 2007.*