

Check but verify

Data sharing provides research with an essential opportunity for error correction by collaborators and disinterested parties alike. Public deposition ensures the useful formatting and recording of essential metadata.

Esophageal squamous cell carcinoma (ESCC) is a major public health problem that has attracted considerable epidemiological interest in China because of its high prevalence and large disparities in regional incidence. Now, an unprecedented level of data sharing and joint analysis of individual-level genome-wide association study (GWAS) data among three groups (see page 1001) shows that data reanalysis can provide robust replication for new findings and an opportunity for error correction, as well as analytical insights into considerable genetic and environmental heterogeneity.

This study confirmed four previously published loci and found two new ESCC susceptibility loci at genome-wide significance, as well as an HLA class II susceptibility locus significantly associated in two high-risk populations. The reanalysis also found no evidence of association for four previously published loci, and we are publishing corrigenda for the corresponding reports (doi:10.1038/ng.648 and doi:10.1038/ng.2411; corrected as of 27 August 2014). **Supplementary Table 8** of the new study is particularly notable, as it details the differences among the three studies and the joint analysis by the three collaborating groups.

So long as collaborating groups interact at a distance via meta-analysis, there will be fewer opportunities of this kind to catch mistakes and replicate discoveries. There will be more misunderstandings about the details of the quality control, analytical workflow and interpretation of data that go into such a study. It is therefore best practice to make all individual-level genotype and phenotype data accessible in a controlled-access repository such as the database of Genotypes and Phenotypes (dbGaP; <http://www.ncbi.nlm.nih.gov/gap>; doi:10.1038/ng.1007–1181) or the European Genome-phenome Archive (EGA; <https://www.ebi.ac.uk/ega/>) or to provide details via a data descriptor (<http://www.nature.com/sdata/>), explaining any special provisions for controlled data access or formal collaboration.

On page 934, the US National Institutes of Health (NIH) data-sharing governance committee summarizes the experience gained in the first seven years of the dbGaP archive and the ways in which it has evolved to provide more rapid access for data requestors and provides some guidance to the complexities of application for and use of data sets, as well as describing some of the success stories of data reuse in research. Prompted by the positive experience of data deposition from GWAS research, the NIH is extending the mandate for data access to other types of data that the Institutes fund, a move we wholeheartedly support.

The 66 articles in this journal published between 2008 and 2013 that cite dbGaP accession codes are highly cited, with a mean of 155 total **Scopus** citations as of 1 August 2014. However, a dbGaP accession does not by itself alter the citation of articles in this journal, according to a preliminary look at 13 pairs of GWAS articles with and without a dbGaP accession published on the same trait on the same day in the same journal (in the case of more than two simultaneous articles, non-overlapping pairs were assigned by sequential DOI number). Citations in the first full year after publication were very similar for paired papers with (mean of 34.62) and without (mean of 34.69) a dbGaP accession. Total Scopus citations as of 1 August 2014 were not significantly greater for paired articles with a dbGaP accession (dbGaP accession: median of 134, quartiles of 103–191; no dbGaP accession: median of 111, quartiles of 89–185; one-tailed Wilcoxon signed-rank test $P = 0.125$). We are aware that the impact of data deposition might take more time to measure; we are interested in further analyzing the contribution of data sharing to citation and in comparing our results with those for other journals that have encouraged data deposition in public archives. To this end, we will provide a list of articles depositing data and citation statistics for interested readers on our blog (<http://blogs.nature.com/freeassociation/>). ■