Readme file for the International Stroke Genetics Consortium (ISGC) distribution files

These are the results files from study:
"Genetic variation in PLEKHG1 is associated with white matter hyperintensities (N=11,226)"

Disclaimer
These data are provided "as is", and without warranty, for scientific and educational use only. The use of these data for commercial purposes is NOT allowed. The inclusion of these data in commercial databases is NOT allowed. If you download these data, you acknowledge that these data will be used only for non-commercial research purposes; that the investigator is in compliance with all applicable state, local, and federal laws or regulations and institutional policies regarding human subjects and genetics research; that secondary distribution of the data without approval by secondary parties is prohibited; and that the investigator will cite the appropriate ISGC publication in any communications or publications arising directly or indirectly from these data. Put more bluntly: there are a multitude of ways to erroneously use these data. The onus is on you to use these data correctly. If you intentionally or unintentionally violate the data use instructions, the responsibility falls to you only.

Caution
Some cases and controls were used by multiple studies across the ISGC. This is a non-trivial, non-ignorable issue. Information on the specific cohorts used in each study can be found in the appropriate paper.

Identifiability
There have been extensive discussions in the human genetics community about SNP level data release. Under some circumstances, it is possible to determine if a person was a case or a control in a GWAS. This is not generally the case, and requires near-ideal circumstances and an independent DNA source and genotyping results for a large number of markers. The data included in this public ISGC distribution differ importantly from the idealized scenario. No individual data are being released. No case/control allele frequencies are included (in aggregate across samples or within each sample). Summary data per SNP were generated by imputing individual data onto a common backbone, analysis with PCA and study covariates, and then results from many different studies were combined to yield summary results. The risk of identifiability using this summary level data has been extensively discussed with expert statistical geneticists. The consensus is that the risk of identifiability, even in the strange case in which a DNA sample from a member of an ISGC study was obtained and analyzed, is extremely small. The Principal Investigators of the ISGC approved this plan for release of results.

Files
See the primary papers (above) for full technical details that produced the enclosed results. Again, appropriate use of these files is entirely your responsibility.

Citation
If you use these data, you must (1) acknowledge that "data were accessed through the ISGC Cerebrovascular Disease Knowledge Portal" and (2) cite the appropriate ISGC publication (see above).