Supplementary Information

Methods

Stastical analyses – determination of adequate sample size for the NAFLD fibrosis and KLF6 genotype association studies. The minimum sample size required to analyse wild type KLF6 being a risk factor for fibrosis in NAFLD was calculated by QUANTO (v.1.2.3, May 2007) Software as detailed in supplementary information. (freeware licence),[1-3] assuming that the prevalence of the NAFLD in the population was 20%. We also assumed that of the two possible alleles at the KLF6 locus (G and A), only individuals with the GG (WT) genotype would be at increased risk of fibrotic liver disease. The proportion of subjects in the population with the WT genotype was estimated to be 85%. The prevalence of the 'G' allele in a population in HW equilibrium was 0.9219544 and the relative risk for GG carriers compared to GA or AA was 5.7. Based on these assumptions, for a desired power of 80%, with a significance level of 0.05 for a 2-sided alternate hypothesis in an unmatched case-control (case: control = 1:1) design, we calculated that at least 68 subjects would be necessary in each of the case and control arms. We considered the presence of stage 2 fibrosis as indicative of progressive disease as stage 1 fibrosis in adults often does not, or can take many years, to progress. For our UK Group 1 controls (fibrosis 0-1) n = 178, while cases (fibrosis 2-4) n = 128. For Italian Group 2, controls n = 53, while cases n = 56. For the combined groups, controls n = 231 versus cases n = 184.

References

1. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 2002;155:478–84.
2. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med 2002;21:35–50.
3. Gauderman WJ, Morrison JM. QUANTO 1.1: A computor program for power and sample size calculations for genetic-epidemiology studies. http://hydra.usc.edu/gxe 2006.