

## SNP Selection and genotyping for HABC.

Publically available databases were then interrogated for SNP variation in the region surrounding the candidate gene. For the first phase of genotyping (the discovery sample) two SNP selection strategies were utilized. In the first strategy, genetic variation in the region spanning 30kb upstream and 10kb downstream of each candidate gene was captured by creating a reference SNP panel of variants with a minor allele frequency (MAF) of at least 5% in Phase I of the International HapMap Project ([www.hapmap.org](http://www.hapmap.org))(31). Tag SNPs were then selected using a pair-wise correlation method ( $r^2 \geq 0.80$ )(32). Candidate genes that were clustered near each other on the chromosome were tagged as a unit spanning all loci of interest. For example, IGFBP2 and IGFBP5 are located only 7.6kb from each other on chromosome 2. Since the region of interest for these two candidates overlapped they were tagged as a unit. In the second strategy, potentially functional SNPs that were either non-synonymous coding variants, predicted to alter a putative transcription factor binding site in the promoter region, or a putative exon splice enhancer with MAF  $\geq 1\%$  were selected for genotyping using the PupaSNP ([pupasuite.bioinfo.cipf.es/](http://pupasuite.bioinfo.cipf.es/)) and Promolign ([polly.wustl.edu/promolign/main.html](http://polly.wustl.edu/promolign/main.html)) databases(33,34).

Genotyping was performed using the Illumina Golden Gate custom assay. Loci with a call rate  $< 85\%$  were excluded.

31. 2003 The International HapMap Project. *Nature* 426(6968):789-96.
32. Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B 2005 Analysis of single-locus tests to detect gene/disease associations. *Genet Epidemiol* 28(3):207-19.
33. Zhao T, Chang LW, McLeod HL, Stormo GD 2004 PromoLign: a database for upstream region analysis and SNPs. *Hum Mutat* 23(6):534-9.
34. Conde L, Vaquerizas JM, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, Robledo M, Dopazo J 2004 PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res* 32(Web Server issue):W242-8.