Open-access database of candidate associations from a genome-wide SNP scan of the Framingham Heart Study

To the Editor:

Knowledge from the HapMap and the development of massively parallel genotyping technologies has enabled high-density genotyping of large populations at low cost using a common set of SNPs. We have identified 251 candidate associations with quantitative traits from an analysis of the Framingham Heart Study Offspring Cohort, which we genotyped at a density of 116,000 SNPs per individual using the Affymetrix 100K Human Mapping arrays. These and similar arrays are also being used in studies of other population samples, making it possible to validate associations *in silico* by comparison of results from the independent analyses. To facilitate this exchange of information, we have created a website to rapidly disseminate the large number of candidate associations uncovered in our study. The associations and additional gene information are available through an open-access website named GMED (for 'genomic medicine database') accessible at http://gmed.bu.edu.

The Framingham Heart Study began in 1948 with the enrollment of two-thirds of the adult population of Framingham, Massachusetts, a small town 15 miles west of Boston[1]. Since that time, two additional generations have been enrolled in the study, one in 1972 and another in 2002, making the study family based. A data set consisting of repeated and highly accurate measurements of many different phenotypes collected over multiple exam cycles is available for each participant, as are the genotypic data used in our analysis (http://www.nhlbi.nih.gov/about/framingham/policies/index.htm). Heritability estimates for different traits range from 0.30–0.56 (http://www.nhlbi.nih.gov/about/framingham/policies/pagetwelve.htm).

We have used this data set to identify a common obesity-predisposing variant near the *INSIG2* gene[2]. We have identified an additional 251 putative associations for different phenotypes, including body mass index (50), diastolic (24) and systolic blood pressure (24), LDL cholesterol (18), triglycerides (15), VLDL cholesterol (32), plasma cholesterol (36), HDL cholesterol (28) and blood glucose (24), using the analysis methods explained below. Some associations are consistent with those reported previously—for example, ARTS-1 and hypertension and lipoprotein lipase are associated with VLDL—indicating that the database is likely to include other *bona fide* associations. The results are viewable via GMED.

Our analysis used a strategy implemented in the software package PBAT[3] in which the genetic effect size for a given SNP is estimated in one step using parental genotypes, and then the ten best SNPs predictive of offspring phenotype are analyzed in a second, transmission-based step that is based on offspring genotypes. Only those SNPs that achieve genome-wide significance after a Bonferroni correction based on the number of SNPs tested in the second step are presented in GMED. We also used a less conservative strategy in which the screen is performed using linear mixed-effects models based on offspring genotypes (P value cutoff = 0.001) and validated within the data set using the family-based test of association (P value threshold = 0.01). False-positive rates for both these approaches, derived by permutation of individuals with phenotypes while maintaining family structure, are given on the website.

GMED presents each association with links to additional gene, protein and literature information. The main association details page for each association is reached by hyperlinks from either the genome browser or a phenotype and analysis–specific map. The association details provided include information about the SNP, including physical position, alleles (for the reported strand), Affymetrix SNP ID, dbSNP ID, allele frequencies observed in the HapMap genotypes[4] and a representation of Entrez Gene–curated genes along with RefSeq-curated transcripts within a 200-kb region centered on the SNP. The association details page also provides details about the analysis in which the association was uncovered, a graph showing the mean phenotypic value and its standard error per age range stratified by SNP genotype and sex, as well as the number of data points included in each mean. To protect the confidentiality of the Framingham Heart Study participants, no individual phenotypic or genotypic data are accessible through GMED. Further, means and genotype counts derived from fewer than 20 individuals are censored to prevent identification of individuals by either uncommon genotype or uncommon phenotype.

To facilitate further exploration of the possible biological implications of the associations, we provide detailed gene information, reached through hyperlinks from the representations of the Entrez Gene loci. These pages provide information about gene names, symbols, synonyms, protein products, short description of the gene's function, gene ontology information, protein complexes that contain protein products of the locus and a sampling of manuscripts related to the gene or its products. This information is collected from public sources such as Entrez Gene[5] and RefSeq[6] as well as BIND[7] and is pulled into GMED on demand from the Cognia Molecular Web Service. Literature abstracts and links to full-text articles are pulled into GMED on demand from NCBI's PubMed.

The database will be updated as the results from other association studies performed in Framingham are completed. Information from

studies in other population samples, including those based on the Affymetrix SNP genotyping platforms, will be added as they become available. To facilitate comparison of results from different studies, associations in GFF format (http://www.sanger.ac.uk/Software/formats/GFF/GFF_Spec.shtml) can be uploaded for visualization within the GMED genome browser as temporary files for individual use. Users are encouraged to deposit their data, either for or against association, along with annotation information describing study design, method of analysis and significance threshold as well as relevant references for viewing by other members of the community.

Alan Herbert[1,3], Marc E Lenburg[1,3], David Ulrich[2], Norman P Gerry[1], Karen Schlauch[1] & Michael F Christman[1]

[1]Department of Genetics and Genomics, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [2]GeneHome, Inc., Brookline, Massachusetts 02445, USA. [3]These authors contributed equally to this work.
e-mail: aherbert@bu.edu or mlenburg@bu.edu

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

1. Kannel, W.B. *J. Atheroscler. Thromb.* **6**, 60–66 (2000).
2. Herbert, A. *et al. Science* **312**, 279–283 (2006).
3. Van Steen, K. *et al. Nat. Genet.* **37**, 683–691 (2005).
4. The International HapMap Consortium. *Nature* **437**, 1299–1320 (2005).
5. Maglott, D., Ostell, J., Pruitt, K.D. & Tatusova, T. *Nucleic Acids Res.* **33**, D54–D58 (2005).
6. Wheeler, D.L. *et al. Nucleic Acids Res.* **34**, D173–D180 (2006).
7. Alfarano, C. *et al. Nucleic Acids Res.* **33**, D418–D424 (2005).

On the cloning of animals from terminally differentiated cells

To the Editor:

In a recent *Nature Genetics* Letter entitled 'Differentiated cells are more efficient than adult stem cells for cloning by somatic cell nuclear transfer', Sung *et al.*[1] draw two major conclusions that are not convincingly supported by the data presented. The authors claim to provide the first unequivocal evidence for the generation of cloned animals from terminally differentiated cells and to prove that differentiated cells are more efficient donor cells for nuclear transfer than adult stem cells.

The question of whether the nuclei of differentiated cells remain equivalent to the nuclei of embryonic cells was raised over 50 years ago and led to the seminal cloning experiments in amphibians and later to the generation of Dolly from an adult mammary gland cell[2]. However, due to the heterogeneity of the donor cell populations and because of a lack of genetic markers, it was impossible to unambiguously determine the identities and differentiation states of the donor cells that gave rise to cloned mammals. In fact, the low efficiency of cloning—usually, only 1%–3% of cloned embryos develop into live animals—did not exclude the possibility that only rare adult stem cells rather than differentiated cells could serve as donors for cloned animals. A line of solid genetic experiments laid this idea to rest. The generation of cloned mice from terminally differentiated lymphocytes[3] and from postmitotic olfactory neurons[4,5] provided unequivocal evidence for the nuclear totipotency of terminally differentiated cells. In these initial studies, cloned mice were produced using a two-step approach that involved the derivation of embryonic stem (ES) cells from cloned blastocysts followed by the generation of entirely ES cell–derived mice by injecting ES cells into tetraploid host blastocysts. Importantly, ES cells derived by nuclear transfer from donor neurons supported the development of cloned mice in a second round of nuclear transfer, hence unambiguously proving the genomic totipotency of postmitotic terminally differentiated cells[4]. This finding was later corroborated by the generation of cloned mice from terminally differentiated natural killer T cells by direct nuclear transfer[6].

Sung *et al.* argue that the cloning of mice from lymphocytes and neurons did not resolve the question of 'nuclear equivalence' because of the two-step cloning procedure. The authors used a single surface marker (Gr-1) and cellular granularity (side scatter) as parameters to isolate a population of terminally differentiated granulocytes that was alleged to be of 99.4% purity, based on morphological criteria. Two viable clones were produced from sorted granulocytes and were claimed to be the first true examples of animals derived from terminally differentiated cells without the recourse of ES cell derivation. These conclusions are not convincing. Sorting of Gr-1–positive cells from bone marrow will enrich not only for mature but also for immature granulocytes as well as for monocytes[7]. Parameters such as granularity and morphology are poor markers to determine the differentiation state of a cell, and expression profiling of a bulk population is of little value to prove the postmitotic state of two randomly picked donor cells. Thus, in the absence of a genetic marker for terminal differentiation, as was used in the previous experiments[3–6], the use of a heterogeneous donor cell population[7] and statistical calculations do not exclude the possibility that less-differentiated cells or contaminating bone marrow cells served as donors for the two surviving clones. The claim that these experiments demonstrate for the first time the development of clones from terminally differentiated donor cells is, therefore, not only based on poor and ambiguous evidence but is also redundant given previous findings.

The second claim of the paper is the generalized statement that differentiated cells are more efficient donor cells than adult stem cells. This claim is based on the lower number of

Table 1 Cloned pups derived from hematopoietic stem cells (HSCs) and granulocytes

Donor cell	Cleaved embryos (two-cell/four-cell)	Embryos transferred	Offspring (percentage cleaved/percentage transferred)	Ref.
HSC	947/453	411	2 (0.2%/0.5%)	10
HSC	379[a]	None	?	1
Granulocytes	394[a]	182	2 (0.5%/1.1%)	1

[a] Combined number of two-cell and four-cell embryos.