

Genetic determinants of HSP70 gene expression following heat shock

Narelle Maugeri, Jayachandran Radhakrishnan and Julian C. Knight*

Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

Received July 15, 2010; Revised September 8, 2010; Accepted September 20, 2010

The regulation of heat shock protein expression is of significant physiological and pathophysiological significance. Here we show that genetic diversity is an important determinant of heat shock protein 70 expression involving local, likely *cis*-acting, polymorphisms. We define DNA sequence variation for the highly homologous *HSPA1A* and *HSPA1B* genes in the major histocompatibility complex on chromosome 6p21 and establish quantitative and specific assays for determining transcript abundance. We show for lymphoblastoid cell lines established from individuals of African ancestry that following heat shock, expression of *HSPA1B* is associated with rs400547 ($P 3.88 \times 10^{-8}$) and linked single nucleotide polymorphisms (SNPs) located 62–93 kb telomeric to *HSPA1B*. This association was found to explain 31 and 29% of the variance in *HSPA1B* expression following heat shock or in resting cells, respectively. The associated SNPs show marked variation in minor allele frequency among populations, being more common in individuals of African ancestry, and are located in a region showing population-specific haplotypic block structure. The work illustrates how analysis of a heritable induced expression phenotype can be highly informative in defining functionally important genetic variation.

INTRODUCTION

Induction of heat shock proteins is a critical and highly conserved cellular response protecting cells from a range of stresses including damage caused by normal physiological processes, extreme environmental stress or disease. Protection arises from a variety of mechanisms, notably the molecular chaperoning function of heat shock proteins acting to identify misfolded or partially denatured proteins, leading to their repair or transportation to sites of degradation (1). The heat shock protein 70 (Hsp70) family is the most highly conserved of the many heat shock protein families across a wide range of species from bacteria to plants and animals (2). A number of different genes encode human inducible Hsp70 proteins notably *HSPA1A* (*HSP70-1*) and *HSPA1B* (*HSP70-2*) located in the major histocompatibility complex (MHC) class III region on chromosome 6 (3–5). These highly homologous single exon genes are found in a tandem arrangement within a 14 kb region of chromosome 6p21.3, with an additional homologous heat shock gene *HSPAIL* located in the reverse orientation approximately 4 kb upstream of *HSPA1A* whose expression is constitutive and predominantly confined to the

testis (6). This cluster of three heat shock genes is also found in the mouse and rat and is hypothesized to have arisen by gene conversion with homogenization of the whole coding region of *HSPA1A* and *HSPA1B*, together with parts of *HSPAIL* (7). The coding regions of *HSPA1A* and *HSPA1B* are highly homologous, differing at only six single base substitutions, while some sequence differences are noted in the promoter and 3'-untranslated regions (UTRs).

Expression of heat shock proteins has been shown to be highly heritable (8,9), suggesting a significant genetic component in determining individual responsiveness at the level of gene expression. The occurrence of *cis*- or *trans*-acting modulators of heat shock protein gene expression would be of significant biological interest, notably in the context of susceptibility to disease. To date, however, no studies have been reported analysing heat shock-induced gene expression as a quantitative trait for human subjects, a powerful approach which has been applied almost exclusively to unstimulated cells (10). There is growing evidence, however, that the role of regulatory genetic variants is likely to be highly context specific (11). For *HSPA1A* and *HSPA1B*, the question is of particular relevance given that these genes are heat shock

*To whom correspondence should be addressed at: Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK. Tel: +44 1865287651; Fax: +44 1865287533; Email: julian@well.ox.ac.uk

inducible, and that polymorphism involving these genes or extended haplotypes over the MHC has been implicated in susceptibility to a number of infectious, inflammatory and autoimmune diseases (12–16) as well as malignancy (17–19) and drug hypersensitivity (20,21).

Extensive linkage disequilibrium (LD) has made fine mapping of disease associations difficult, while the identification of specific functional variants involving *HSPA1A* and *HSPA1B* based on reporter gene assays or association of candidate single nucleotide polymorphisms (SNPs) with levels of gene expression has often been controversial (12–14,22). Moreover, the remarkable homology of these genes has caused some difficulties in the unambiguous assignment of SNP locations to specific sequences and in clearly distinguishing between *HSPA1A* and *HSPA1B* transcript when assaying gene expression (23–25). In this study, we sought to define the extent and nature of DNA sequence variation involving *HSPA1A* and *HSPA1B* among individuals of European and African ancestry, and to determine by quantitative trait mapping how genetic variation may modulate expression of these important stress response genes in a physiologically relevant context, namely response to heat shock.

RESULTS

Sequence variation involving *HSPA1A* and *HSPA1B*

DNA sequence variants involving *HSPA1A* and *HSPA1B* genic sequences and a 2 kb interval 5' or 3' to each of these genes were identified by resequencing 48 unrelated individuals from the International HapMap Project (26), 24 individuals of European ancestry (CEU panel) and 24 of African ancestry [Yoruba from Ibadan Nigeria (YRI) panel]. In order to unambiguously assign variants for these highly homologous gene regions, genomic DNA was first PCR amplified using long range specific primers and high fidelity Taq polymerase, and then cloned into pCR4-TOPO before being subject to Sanger sequencing. We obtained overlapping gene-specific sequence amplicons spanning chr6: 31889526–31894556 (*HSPA1A*) and chr6: 31902517–31906207 (*HSPA1B*) with 2× coverage. This identified 20 SNPs in the *HSPA1A* gene region and 16 SNPs in the *HSPA1B* region (Fig. 1), including five novel SNPs involving *HSPA1A* and two for *HSPA1B*. We noted that of the SNPs identified in the *HSPA1A* region, five lie upstream of the transcriptional start site (TSS), five within the 5'-UTR, seven in the coding region and three downstream of the gene. In the *HSPA1B* region, five SNPs were identified upstream of the TSS, three in the 5'-UTR and eight in the coding region of the gene. No insertions or deletions were identified. Of the seven novel SNPs identified, three were missense mutations (Fig. 1). No SNPs were found in the 3'-UTR of either gene.

We proceeded to genotype 60 unrelated founder individuals in the CEU and YRI panels in order to establish allele frequencies for the variants identified and determine their relationship to underlying allelic structure. The location, nature and minor allele frequency (MAF) of SNPs in the two populations are summarized in Figure 1. Overall, there was greater SNP diversity among individuals of African ancestry (YRI panel). Allele frequencies were found to vary between populations such that of the 36 SNPs identified in the two populations, 12 were

monomorphic in the CEU panel and 7 in the YRI panel. Among the SNPs polymorphic in both populations, MAF varied significantly. These included SNPs such as rs1043618, rs6457452 and rs1061581 previously implicated in disease susceptibility and longevity (12,13,27–29).

HSPA1A and *HSPA1B* are thought to have arisen by duplication and gene conversion. Our data showed that two pairs of SNPs were found in the same positions in the coding sequence of *HSPA1A* and *HSPA1B*. The first pair, rs562047 (*HSPA1A*) and rs17856061 (*HSPA1B*), is non-synonymous SNPs located 330 bp from the ATG of each gene which results in a missense mutation in the Hsp70 protein, p.Glu110Asp, as a result of a G to C nucleotide substitution. The second pair of SNPs, rs541340 (*HSPA1A*) and rs35682610 (*HSPA1B*), are synonymous and are located 1710 bp downstream of the coding start site.

HSPA1A and *HSPA1B* expression following heat shock

In order to determine how local or distant genetic variants may modulate expression of *HSPA1A* and *HSPA1B*, we established EBV-transformed lymphoblastoid cell lines (LCLs) as a model system in which we could reproducibly induce heat shock for a physiologically relevant cell type. This would allow expression quantitative trait (eQTL) mapping across a panel of LCLs established from unrelated individuals for whom dense genotyping data are available as part of the International HapMap Project (30). This approach has previously been highly informative for genome-wide mapping of expression-associated SNPs using resting HapMap LCLs (31–33), notably for individuals of African ancestry in whom high levels of haplotype diversity facilitates fine mapping of SNP associations (34). Inducible Hsp70 expression following heat shock has been reported in LCLs (35) as well as the B cell lymphoma cell line Raji (36) and primary peripheral blood lymphocytes (25,37).

We proceeded to design a quantitative real-time PCR assay that would discriminate between transcript arising from *HSPA1A* or *HSPA1B*. The high degree of sequence homology between the genes has led to difficulties in previous studies with the specificity of transcript quantification (23–25). A minor groove binder (MGB) TaqMan probe with a 5' reporter dye and 3' non-fluorescent quencher was used to minimize background while stabilizing probe hybridization and allowing use of shorter primers (38). Assays were designed to the 3'-UTR of *HSPA1A* and *HSPA1B* as some sequence variation is seen here between the two genes (Supplementary Material, Fig. S1). In order to investigate the specificity of the resulting primer-probe sets, the full-length cDNA sequence for either *HSPA1A* or *HSPA1B* was cloned into pCR4-TOPO and used as template for the gene-specific TaqMan assays. No signal above background was detected when the plasmid not matching the primer-probe set was used as template for the TaqMan probe assay. To quantify expression of *HSPA1A* and *HSPA1B* in the biological samples, a standard curve was included on all plates with assays performed in technical duplicates for each of the three biological replicate experiments. This demonstrated that, for LCL GM18517, both transcripts were highly inducible within 30 min of heat shock, rising to a maximum following heat shock for 1 h plus 1 h recovery (Fig. 2). We also observed marked upregulation of Hsp70 protein

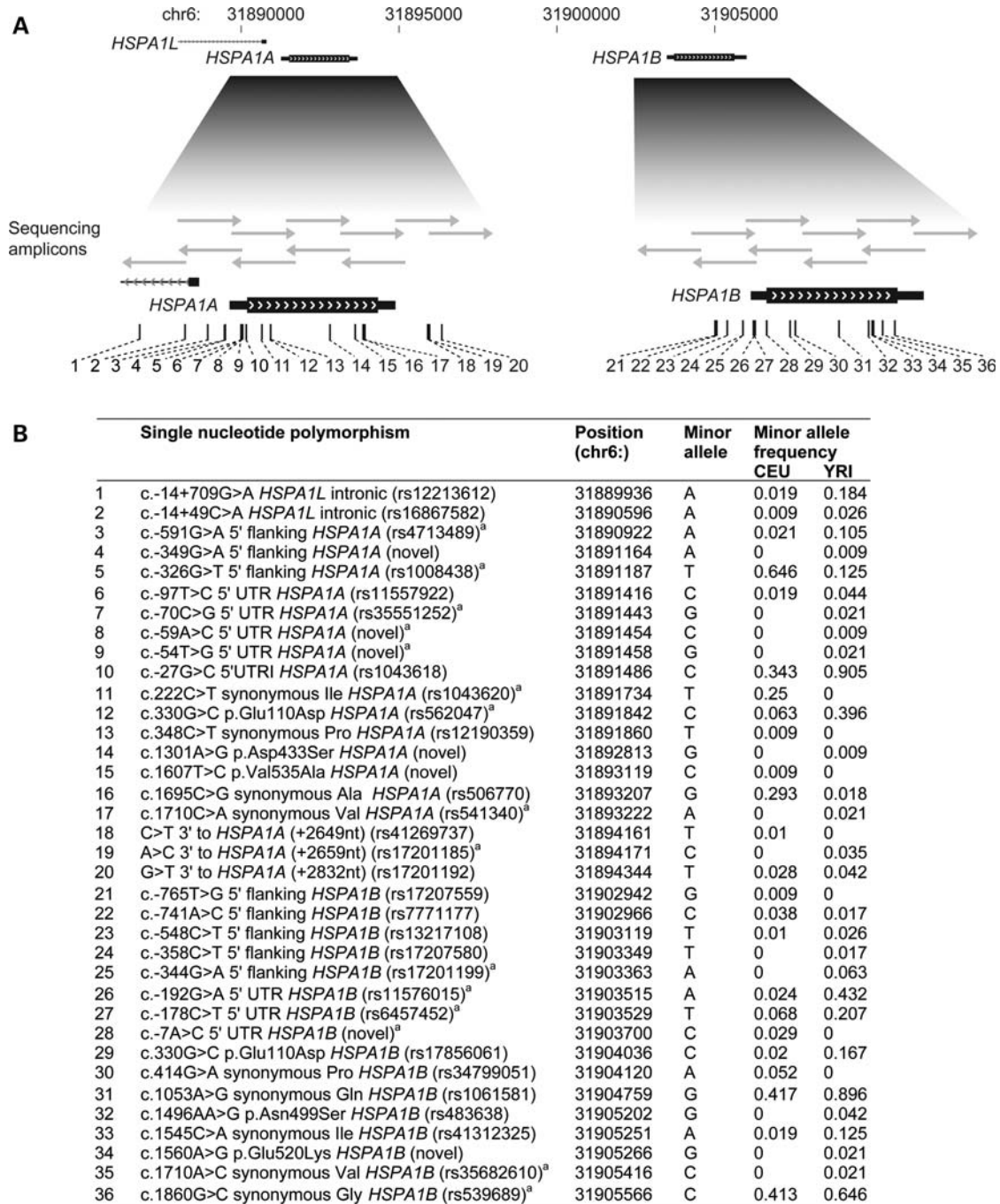


Figure 1. Sequence diversity in *HSPA1A* and *HSPA1B* gene regions. (A) Sanger-based sequencing amplicons shown (grey arrows) together with the location of SNPs. (B) SNPs were identified by resequencing and then genotyped in all 60 founder individuals for each HapMap population panel. ^aSNPs for which MAF data were available from resequencing only (24 individuals each panel).

production following 1 h of heat shock at 42°C, which was maximal after 6 h of recovery at 37°C when a 77-fold induction in Hsp70 levels was found (Fig. 2).

eQTL mapping for *HSPA1A* and *HSPA1B*

We proceeded to map expression of the two inducible Hsp70 genes as a quantitative trait using LCLs established from 60 unrelated individuals from the YRI HapMap panel. We

assayed gene expression at the transcript level for *HSPA1A* and *HSPA1B* using gene-specific quantitative real-time PCR for resting cells, and for cells after heat shock at 42°C for 1 h. Growth rates were determined to ensure uniform cell densities. Two independent biological replicate experiments were performed for each LCL. *HSPA1A* and *HSPA1B* gene-specific expression was normalized against the housekeeping gene *GAPDH* based on the difference in the threshold cycles (ΔC_t).

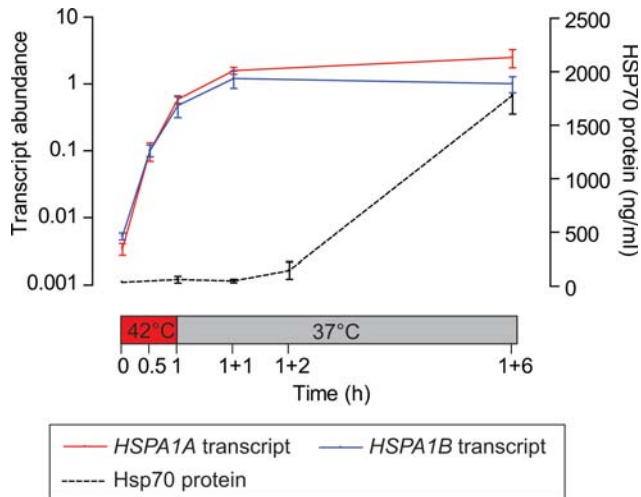


Figure 2. Hsp70 expression following heat shock in GM18517. *HSPA1A* and *HSPA1B* transcript quantification expressed relative to *GAPDH* (left y-axis) and Hsp70 protein (right y-axis) in resting cells or after heat shock at 42°C for the times indicated (x-axis).

eQTL mapping was performed using SNP genotyping data for the *HSPA1A* and *HSPA1B* gene regions we had determined for these individuals together with genome-wide genotyping data for 2 268 955 SNPs with an MAF greater than 5% available through the International HapMap Project (26). We first considered the ΔC_t values following heat shock for Hsp70 gene expression using the mean of two biological replicate experiments carried out for each LCL. The most striking result was seen for *HSPA1B* expression with evidence of a major local, likely *cis*-acting, eQTL located 62–93 kb telomeric to *HSPA1B* (Fig. 3, Supplementary Material, Table S1). This was the strongest genome-wide association observed and showed four SNPs in complete LD were associated with *HSPA1B* expression after heat shock ($P = 3.88 \times 10^{-8}$): rs400547, an intronic SNP located in *CLIC1* (c.40–670G>A); rs707915 and rs1150793, two intronic SNPs in *MSH5* (c.415+21T>A and c.812+2487A>G, respectively) and rs707936, an intronic SNP in *C6orf27* (c.2499+10C>T).

We then investigated whether this result was reproducible. Strikingly, the same SNP markers are robustly associated with expression of *HSPA1B* after heat shock in both biological replicate experiments (Supplementary Material, Fig. S2). Possession of a copy of the A allele of rs400547 is associated with 60% lower expression of *HSPA1B* following heat shock compared with those individuals who do not have a copy (Fig. 4). A clear allelic dose effect was seen with mean levels of *HSPA1B* expression (relative to *GAPDH*): 0.69 (95% confidence intervals 0.49–0.89) for those homozygous GG, 0.34 (0.23–0.44) for individuals heterozygous GA and 0.08 (0.04–0.12) for those homozygous AA.

When we analysed basal *HSPA1B* expression, the same SNPs showed the strongest association with mean *HSPA1B* expression (Fig. 4, Supplementary Material, Fig. S3). Basal expression of *HSPA1B* is considerably lower than that seen following heat shock (mean levels 0.010 versus 0.453) but is detectable in LCLs. We found that the same four linked SNPs were associated with basal *HSPA1B* expression ($P = 6.45 \times 10^{-8}$) (Supplementary Material, Table S1). Possession of

rs400547 was associated with 85% lower basal expression of *HSPA1B*. The same association was also demonstrated when we considered each biological replicate experiment separately and found that the strongest association for each replicate was with rs400547 and linked SNP markers (Supplementary Material, Fig. S2).

In order to fine map this association further, we augmented the HapMap 2 SNP data set with data recently released from the 1000 Genomes Project (www.1000genomes.org) for the YRI cohort. For a 220 kb window which included the most strongly associated HapMap SNPs and the Hsp70 gene cluster, a total of 1538 SNPs were tested for the association (Supplementary Material, Fig. S4). This confirmed the strong peak of association but did not resolve the association further. The analysis revealed that one additional SNP was associated with *HSPA1B* expression (rs378538) in complete LD with rs400547. rs378538 is a C to T SNP located in the *CLIC1* promoter region, 683 bp upstream of the transcriptional start site.

The association with *HSPA1B* expression falls within a 158 kb region flanked by peaks of recombination (chr6: 31784000–31942000) with the strongest association involving five SNPs in complete LD (rs400547, rs1150793, rs707936, rs707915, rs378538) located in *CLIC1*, *MSH5* and *C6orf27*. The extent of LD across the region of association and flanking regions was defined for the YRI HapMap population of African ancestry and contrasted with that seen among individuals of European ancestry (CEU HapMap population) (Supplementary Material, Fig. S5). There was no evidence of block structure involving the Hsp70 genes. However, there was some evidence of haplotype block structure elsewhere which was more extensive among African individuals.

This was notable for expression-associated SNPs in the *LY6G* gene cluster, *DDAH2*, *CLIC1* and *MSH5* where a 31.2 kb haplotypic block (chr6: 31794476–31825675) was defined using Haploview and the confidence intervals method (39). This resolved nine common haplotypes with a frequency greater than 0.01 based on 21 SNPs with a MAF >0.05 (Supplementary Material, Fig. S6). We then tested for evidence of haplotypic association using a linear regression model and found significant association. Haplotype 4, which includes the minor allele of rs400547, was significantly associated with *HSPA1B* expression following heat shock or in resting cells ($P = 4.6 \times 10^{-6}$ and 9.0×10^{-6}), explaining 31 and 29% of the variance in gene expression, respectively (Supplementary Material, Fig. S6). The other haplotype bearing the minor allele of rs400547, haplotype 5, also shows association but with lower statistical significance as it is rare. A further haplotype, denoted haplotype 8, showed an apparent independent haplotypic association with increased *HSPA1B* expression (Supplementary Material, Fig. S6).

The functional variant responsible for the observed association remains unresolved but a *cis*-acting effect is likely given the proximity of the region to *HSPA1B*. Four of the linked SNP markers are intronic (rs400547, rs1150793, rs707936, rs707915) and one is located in the promoter region of *CLIC1* (rs378538), but none correspond to known regulatory elements. No association was found for the haplotype tagged by rs400547 with *CLIC1* expression using publicly available data sets of gene expression for the YRI panel (31,40). *In silico* analysis of transcription factor

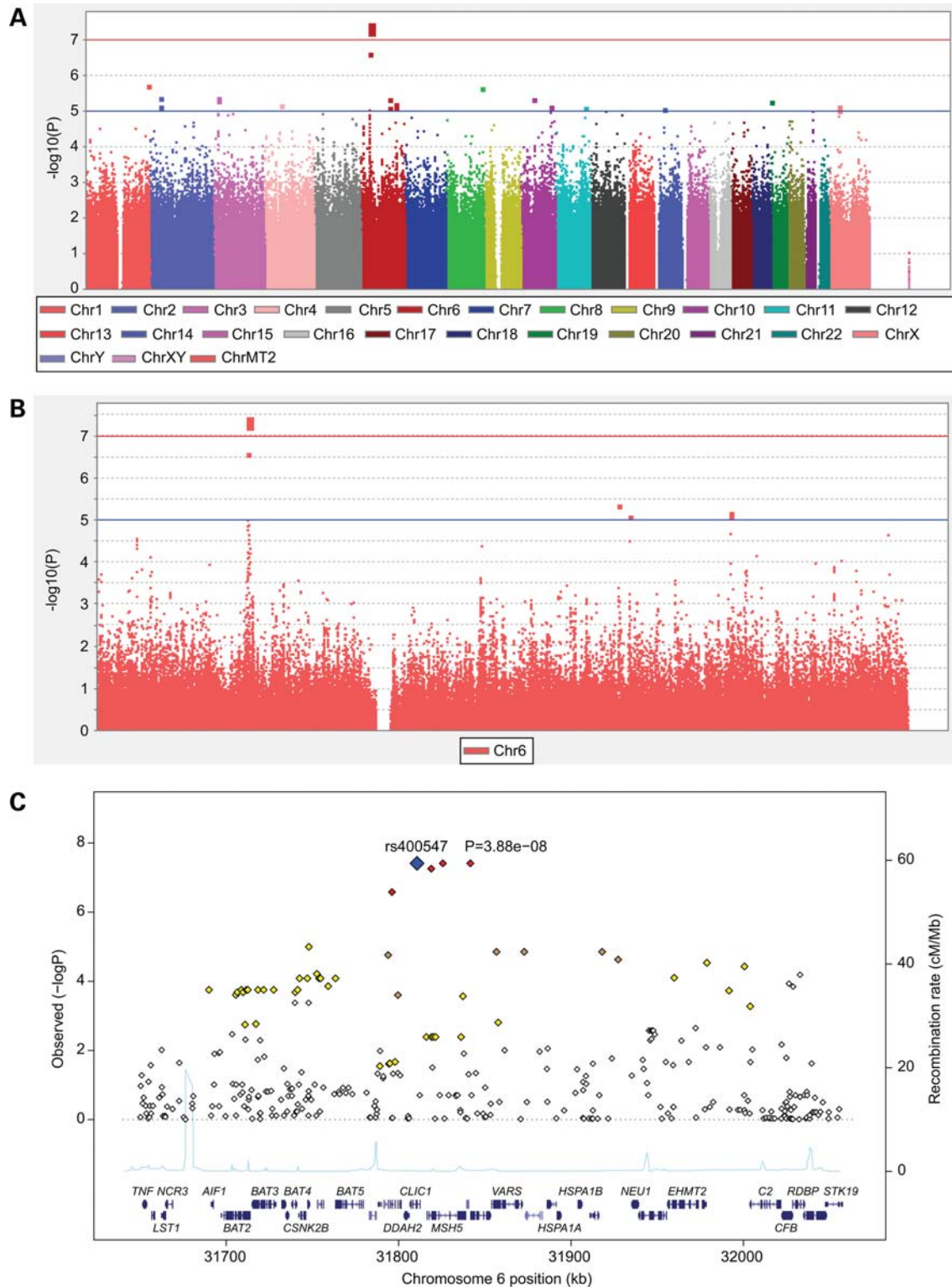


Figure 3. Quantitative trait mapping for *HSPA1B* expression following heat shock. Expression-associated SNPs based on mean ΔC_t values for *HSPA1B* expression for two biological replicate experiments in YRI HapMap founder LCLs following heat shock at 42°C for 1 h. (A) Manhattan plot showing strength of association from PLINK analysis plotted as $-\log_{10}(P)$ values by chromosome for *HSPA1B* expression following heat shock. For rs400547 and linked SNPs, the observed association at $P = 3.88 \times 10^{-8}$ was significant at a conservative genome-wide significance threshold of $< 5 \times 10^{-8}$ (59) and on permutation analysis (empirical P -value = 2×10^{-4}). Other signals with moderate evidence of association which are candidates for further study are highlighted between 10^{-5} and 10^{-7} by the blue and red lines, respectively (60). (B) Association data shown at higher resolution for chromosome 6. (C) Single marker allelic association results for a 400 kb region spanning *HSPA1B* (chr6: 31650000–32050000) plotted as $-\log_{10}(P)$ values (left y-axis) by genomic coordinate (x-axis). With reference to rs400547, SNPs with MAF greater than 5% and $r^2 < 0.2$ are shown as white squares, 0.2–0.5 (yellow), 0.5–0.8 (brown) and > 0.8 (red). Recombination rate is also plotted (right y-axis). Gene structure for the region is shown below.

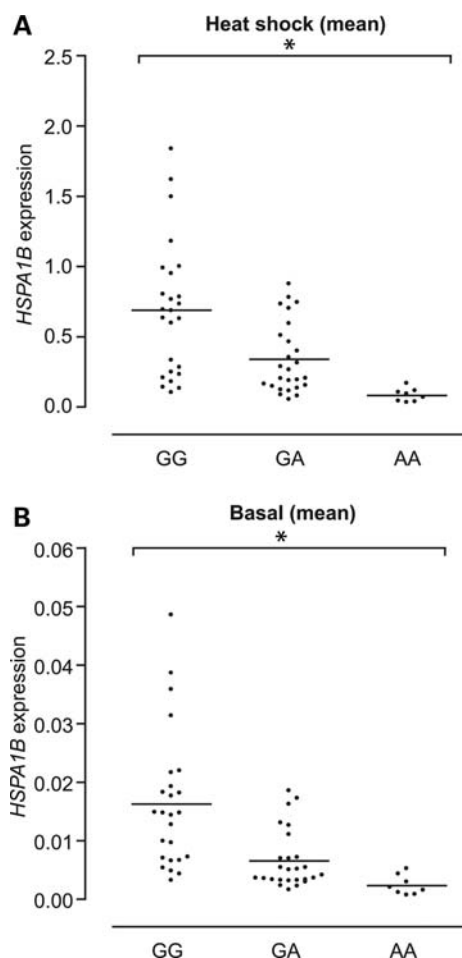


Figure 4. Expression of *HSPA1B* clustered by genotype for rs400547. Expression values are shown for *HSPA1B* (expressed relative to *GAPDH*) for LCLs either (A) following heat shock at 42°C for 1 h or (B) resting (basal expression). Data are shown for the mean of the two replicate experiments. Uncorrected *P*-values derived from one-way analysis of variance for each condition are shown. **P* < 0.0001. Lines indicate mean values for each genotype.

binding was performed using the JASPAR (<http://jaspar.genereg.net>) open-access database of transcription factor binding profiles (41). This showed evidence of allele-specific binding for only one SNP, predicting that BRCA1 binds with greater affinity to the associated G allele of rs1150793 (ACAAGAC) (score 6.4) (relative profile score 90.4%). The five SNP markers are in complete LD and show marked differences in allele frequency between populations based on data from the HapMap Project (30), being common in populations of African ancestry (MAF for rs1150793 is 0.344 in YRI panel of individuals from the Yoruba people of Ibadan, Nigeria) but rarer in non-African populations (MAF 0.031 in CEU panel of European ancestry) (pairwise F_{ST} YRI versus CEU 0.18). Within Africa, the observed MAF is lower in East Africa, being 0.177 in the Luhya people in Webuye, Kenya and 0.094 in the Maasai people in Kinyawa, Kenya.

For *HSPA1A* expression, we found a number of distant expression-associated SNP markers (Supplementary Material, Fig. S7 and Table S1) which are candidates for further investigation. However, the observed expression-associated SNPs

did not include the major local association on chromosome 6 seen for *HSPA1B*.

Previous studies looking to define SNP association with Hsp70 gene expression have been restricted to specific candidate SNPs. Association was reported for rs6457452 and rs10615181 (also known as ‘HSPA1B–179’ and ‘HSPA1B+1267’, respectively) with lower expression associated with the C allele of rs6457452 and no apparent effect of rs10615181, although these data have been controversial (12–14,22). In our data set, we find a weak association with rs6457452, but this is not consistent across data sets (Supplementary Material, Fig. S8). This is unlikely to have arisen due to LD with rs400547 ($r^2 = 0.148$).

DISCUSSION

Heat shock proteins play a critical role in normal physiological processes and disease pathogenesis. The role of genetic variation in modulating expression of the different members of heat shock protein families remains unresolved but is of significant interest given evidence of heritability and disease association. Here we define a major local eQTL locus for *HSPA1B*, mapping to a region between 60 and 90 kb telomeric to the gene. The relatively large magnitude of the effect observed combined with the use of a carefully defined and robust phenotype, heat shock and a quantitative and specific assay of transcript abundance has allowed resolution of this eQTL using a relatively modest sample size. The association was seen to be reproducible between biological replicate experiments and also be present in the basal levels of gene expression observed in this cell type. The study was designed to detect local, likely *cis*-acting associations and is not adequately powered to establish distant, likely *trans*-acting associations unless these were of large magnitude.

The association with *HSPA1B* expression is located in a region involving *CLIC1/MSH5* previously proposed to have been subject to selection (42). The five SNP markers tagging the associated allele show marked variation in allele frequency between populations, being high among those of African ancestry but rare in other populations. Allele frequencies were highest in Yoruban individuals from West Africa compared with the Luhya or Maasai populations in East Africa. We note the haplotypic block structure present in Yoruban individuals for the expression-associated region and that haplotypic association suggests there may, in addition, be a high producer allele. However, the defined block does not include rs707936 (in complete LD with rs400547), which is located in *C6orf27*. Sequencing and genotyping of additional individuals, including individuals from different African populations, would allow the haplotypic structure to be resolved with greater precision and enable fine mapping of the observed association with gene expression. Additional alleles are likely to be identified which will direct efforts to establish the functional basis of the observed association. The very low allele frequency in Caucasians means that currently available eQTL data sets are not informative for this haplotype and that further analysis in populations of European ancestry is unlikely to be productive unless adequately powered. For the Hsp70 gene locus itself, our analysis showed no evidence of

haplotypic block structure in either the African or Caucasian populations studied with relatively few coding variants and no variants in the 3'-UTR of either *HSPA1A* or *HSPA1B*.

The local nature of the association and the fact it is not seen with *HSPA1A* suggest that spatial conformation or promoter sequence differences between *HSPA1A* and *HSPA1B* may be important in determining specificity for a *cis*-acting event, perhaps involving recruitment of a transcriptional complex to a promoter or enhancer element involving *CLIC1/MSH5* which comes into proximity with *HSPA1B* through DNA looping. One of the five SNPs, rs378538, is located in the *CLIC1* promoter region, but the regulatory significance of this region has not been characterized. The biological role of *CLIC1* as a sensor and effector during cell stress (43) may be relevant in terms of co-regulation with a heat shock response; however, the eQTL analysis suggests that the associated SNPs do not modulate *CLIC1* expression in resting cells. A further linked SNP, rs1150793, *in silico* has the potential to recruit BRCA1 which is known to be rapidly responsive to heat shock at the protein level and participate in heat shock pathways (44), including modulation of heat shock gene expression using reporter gene systems. The functional basis of the observed association will require further detailed analysis.

This work has defined expression-associated SNPs for *HSPA1B* which will be important as candidate SNPs for disease association studies and to facilitate the fine mapping of reported associations involving this gene region. This may be particularly relevant for infectious diseases including severe sepsis that has been associated with genetic diversity involving the Hsp70 locus (12,13,28) and may have exerted significant selective pressures. Particular variants conferring a selective advantage would be consistent with reported associations with longevity and survival (25,27,29), and our findings of population differences in MAF. We have shown that a previously reported SNP rs6457452 implicated in septic shock (12) is weakly associated with *HSPA1B* expression but that the observed association with rs400547 and linked SNPs is likely to be much more informative. It is noteworthy that one of these SNPs, rs707915, which is in complete LD with rs400547, has been reported to be associated with type 1 diabetes after accounting for MHC class II alleles (45). A further linked SNP rs707936 found on the expression-associated haplotype showed some evidence of disease association with multiple sclerosis in African Americans, but further work is required to define this (46).

MATERIALS AND METHODS

Cell culture and heat shock

LCLs were maintained in RPMI 1640 supplemented with 10% fetal calf serum (FCS) and 2 mM L-glutamine at 37°C in a humidified 5% CO₂ environment. Growth rates were determined for each cell line to ensure uniformity in cell numbers on harvesting and Trypan blue staining used to define cell viability. Two biological replicate heat shock experiments were performed for each cell line. Cells were subject to heat shock by incubation in a water bath at 42°C followed by recovery at 37°C for the times indicated. 2 × 10⁷ cells were harvested for RNA quantification.

Quantitative TaqMan PCR and RNA analysis

Total RNA was purified using QIAGEN RNeasy mini kit including on-column DNase digestion (Qiagen). cDNA was synthesized using oligo dT primers and Superscript III reverse transcriptase (Invitrogen). Real-time PCR assays were performed using the ABI7900 (Applied Biosystems) using TaqMan minor groove binder (MGB) probes. Primer and probe sequences are given in Supplementary Material, Table S2. Three technical replicates were run for each sample. PCRs comprised 1×TaqMan Universal PCR Master-Mix (without UNG), 250 nM probe, 400 nM forward primer, 400 nM reverse primer and 10–100 ng of template. PCR conditions were 50°C for 2 min, then 40 cycles of 95°C for 10 min, 95°C for 15 s and 60°C for 1 min. A standard curve comprising dilutions of homologous standards from a known starting concentration of mRNA were included on each plate. In order to confirm the specificity of the primer probe sets, plasmids containing *HSPA1A* or *HSPA1B* were prepared. To do this, a 2.5 kb fragment spanning *HSPA1A* (chr6: 31891253–31893749) was cloned into pCR4-TOPO (Invitrogen) using the primer pair 5'-GGTCTCCGTGACGACTTATAA-3' and 5'-CAACCTATGCAGACCCTACTGA-3'. Similarly, a 2.6 kb fragment spanning *HSPA1B* (chr6: 31903457–31906065) was cloned into pCR4-TOPO (Invitrogen) using the primer pair 5'-CCACCGACGACTTATAAAAGCCGA-3' and 5'-TCAGACACTATCCCTCCGCAA-3'.

DNA sequencing and genotyping for *HSPA1A* and *HSPA1B*

The SequelPrep Long PCR Kit (Invitrogen) was used to generate specific amplicons spanning *HSPA1A* (chr6: 31889526–31894556) and *HSPA1B* (chr6: 31902517–31906207) from 200 ng genomic DNA according to the manufacturer's instruction. PCR amplification was performed using Tetrad thermo cyclers (MJ Scientific) cycling at 94°C for 2 min, then 94°C for 10 s, 62–54°C for 30 s, 68°C for 5 min for 9 cycles decreasing by 1°C per cycle, then 94°C for 10 s, 54°C for 30 s, 68°C for 5 min for 26 cycles (adding 20 s per cycle to extension time), and a final extension of 68°C for 5 min. PCR amplification was confirmed by the agarose gel electrophoresis and purified using PCR Cleanup Filter Plates (Millipore) before sequencing using BigDye terminator V3.1 cycle sequencing kit (Applied Biosystems) following the manufacturer's instructions. Primer sequences are shown in Supplementary Material, Table S2. Sequences were aligned using PhredPhrap (47,48) and visualized using Consed (49,50) to identify sequence variants. Genotyping was performed using the Sequenom iPLEX Mass-Array assay system in a 384-well format according to the manufacturer's instructions. Primer sequences are shown in Supplementary Material, Table S2.

HSP70 ELISA

A sandwich ELISA was designed and optimized for cell lysates based on previously published methods (51–55). After heat shock, the cells were lysed in 200 µl lysis solution [PBS with 1 mM EDTA, 0.5% Triton X-100, and 1X Protease Inhibitors (Roche)] for 10⁷ cells by incubating on ice for 30 min, then

three rounds of freeze–thaw cycles in a dry-ice ethanol bath. Cells were then centrifuged at 16 000g in a microcentrifuge at 4°C for 10 min, and total protein concentration of the supernatant estimated using the Bio-Rad Protein Assay solution (BioRad). Certified High Bind 96-well plates (Corning) were coated with 0.1 µg/ml of mouse anti-HSP70 monoclonal antibody (Stressgen) diluted in 0.1 M carbonate buffer, pH 9.6, and plates were incubated at 4°C overnight. They were then washed three times with PBS/0.05% Tween 20, blocked for 1 h at RT with blocking solution (1% BSA in PBS) and washed three times. One hundred microlitres of lysed cell solutions diluted to 1 µg/ml total protein in PBS and a range of HSP70 protein standards (Stressgen, NSP-555) ranging from 400 to 6.25 ng/ml (in duplicate) were used to create a standard curve. Plates were then incubated for 2 h at 37°C and then washed three times with PBST. Rabbit Anti-Hsp70 (hsp72) polyclonal antibody (Stressgen, SPA-812) was then added at a dilution of 1:500 [diluted in PBST and 2% mouse serum (Sigma)] and incubated for 1 h at RT. The plates were then washed three times with PBST. Goat Anti-Rabbit IgG:HRP [Horseradish Peroxidase Conjugate Absorbed with Human IgG (Stressgen, SAB-300)] was diluted 1:10 000 in a PBST/1% BSA solution and incubated for 1 h. The plates were then washed three times with PBST. One hundred microlitres of TMB (3,3',5,5'-TetraMethylBenzidine; Invitrogen) were added to each well and incubated approximately 10 min before addition of a stopping solution of 1 M HCl. Optical density was read at 450 nm using SoftMax microplate reader and SoftMax Pro 4 software.

Statistical genetic analysis

For the YRI panel, HapMap release 24 was used which included ~2.46 million autosomal and 72 mitochondrial SNPs of >0.05 MAF. These represent ~1.6 million independent SNPs genome-wide, after the exclusion of SNPs in complete LD ($r^2 = 1.0$). Quantitative trait analysis to define SNP association with gene expression was performed using PLINK (56). Samples and SNPs with genotyping failure rates of 0.1% were excluded from the analysis. Phenotypic means were compared across the genotypic states using the Wald test, to generate an asymptotic *P*-value with no normality assumptions. Empirical *P*-values controlling for family-wise error rates were calculated using a permutation procedure ($n = 5000$) involving phenotype label swapping to randomly allocate sample identifiers and break the phenotype–genotype relationship but retain the same patterns of LD between SNPs in the observed and permuted samples. For fine mapping a 900 kb region including the *HSPA1A* and *HSPA1B* genes, 1000 genome data (www.1000genomes.org) available for 38 founder YRI individuals were used with genotypes for the remaining YRI individuals imputed with impute version 2.0 (57). Haplotype analysis was performed using HaploView version 4.1 (58), and the inferred haplotypes were analysed for the association with gene expression.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We would like to thank all members of the Knight lab for helpful discussion and suggestions relating to this study.

Conflicts of Interest statement. None declared.

FUNDING

This work was supported by the Wellcome Trust (074318 to J.C.K., 075491/Z/04 core facilities WTCHG). Funding to pay the Open Access Charge was provided by the Wellcome Trust.

REFERENCES

- Morimoto, R.I. (1998) Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes Dev.*, **12**, 3788–3796.
- Beere, H.M. and Green, D.R. (2001) Stress management—heat shock protein-70 and the regulation of apoptosis. *Trends Cell Biol.*, **11**, 6–10.
- Milner, C.M. and Campbell, R.D. (1990) Structure and expression of the three MHC-linked HSP70 genes. *Immunogenetics*, **32**, 242–251.
- Sargent, C.A., Dunham, I., Trowsdale, J. and Campbell, R.D. (1989) Human major histocompatibility complex contains genes for the major heat shock protein HSP70. *Proc. Natl Acad. Sci. USA*, **86**, 1968–1972.
- Goate, A.M., Cooper, D.N., Hall, C., Leung, T.K., Solomon, E. and Lim, L. (1987) Localization of a human heat-shock HSP 70 gene sequence to chromosome 6 and detection of two other loci by somatic-cell hybrid and restriction fragment length polymorphism analysis. *Hum. Genet.*, **75**, 123–128.
- Ito, Y., Ando, A., Ando, H., Ando, J., Saijoh, Y., Inoko, H. and Fujimoto, H. (1998) Genomic structure of the spermatid-specific hsp70 homolog gene located in the class III region of the major histocompatibility complex of mouse and man. *J. Biochem.*, **124**, 347–353.
- Kudla, G., Helwak, A. and Lipinski, L. (2004) Gene conversion and GC-content evolution in mammalian Hsp70. *Mol. Biol. Evol.*, **21**, 1438–1444.
- Dumas, P., Sun, Y., Corbeil, G., Tremblay, S., Pausova, Z., Kren, V., Krenova, D., Pravenec, M., Hamet, P. and Tremblay, J. (2000) Mapping of quantitative trait loci (QTL) of differential stress gene expression in rat recombinant inbred strains. *J. Hypertens.*, **18**, 545–551.
- Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
- Cookson, W., Liang, L., Abecasis, G., Moffatt, M. and Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, **10**, 184–194.
- Knight, J.C. (2009) *Human Genetic Diversity. Functional Consequences for Health and Disease*. Oxford University Press, Oxford.
- Kee, C., Cheong, K.Y., Pham, K., Waterer, G.W. and Temple, S.E. (2008) Genetic variation in heat shock protein 70 is associated with septic shock: narrowing the association to a specific haplotype. *Int. J. Immunogenet.*, **35**, 465–473.
- Temple, S.E., Cheong, K.Y., Ardlie, K.G., Sayer, D. and Waterer, G.W. (2004) The septic shock associated HSPA1B1267 polymorphism influences production of HSPA1A and HSPA1B. *Intensive Care Med.*, **30**, 1761–1767.
- Pociot, F., Ronningen, K.S. and Nerup, J. (1993) Polymorphic analysis of the human MHC-linked heat shock protein 70 (HSP70-2) and HSP70-Hom genes in insulin-dependent diabetes mellitus (IDDM). *Scand. J. Immunol.*, **38**, 491–495.
- Bogunia-Kubik, K., Koscińska, K., Suchnicki, K. and Lange, A. (2006) HSP70-hom gene single nucleotide (+2763 G/A and +2437 C/T) polymorphisms in sarcoidosis. *Int. J. Immunogenet.*, **33**, 135–140.
- Spagnolo, P., Sato, H., Marshall, S.E., Antoniou, K.M., Ahmad, T., Wells, A.U., Ahad, M.A., Lightman, S., du Bois, R.M. and Welsh, K.I. (2007) Association between heat shock protein 70/Hom genetic polymorphisms

- and uveitis in patients with sarcoidosis. *Invest Ophthalmol. Vis. Sci.*, **48**, 3019–3025.
17. Sfar, S., Saad, H., Mosbah, F. and Chouchane, L. (2010) Synergistic effect and VEGF/HSP70-hom haplotype analysis: relationship to prostate cancer risk and clinical outcome. *Hum. Immunol.*, **71**, 377–382.
 18. Shibata, T., Arisawa, T., Tahara, T., Yoshioka, D., Maruyama, N., Fujita, H., Kamiya, Y., Nakamura, M., Nagasaka, M., Iwata, M. *et al.* (2009) Protective role of genetic polymorphism of heat shock protein 70-2 for gastric cancer risk. *Dig. Dis. Sci.*, **54**, 70–74.
 19. Ucisik-Akkaya, E., Davis, C.F., Gorodezky, C., Alaez, C. and Dorak, M.T. (2009) HLA complex-linked heat shock protein genes and childhood acute lymphoblastic leukemia susceptibility. *Cell Stress Chaperones*, **15**, 475–485.
 20. Alfirevic, A., Mills, T., Harrington, P., Pinel, T., Sherwood, J., Jawaaid, A., Smith, J.C., March, R.E., Barratt, B.J., Chadwick, D.W. *et al.* (2006) Serious carbamazepine-induced hypersensitivity reactions associated with the HSP70 gene cluster. *Pharmacogenet. Genomics*, **16**, 287–296.
 21. Martin, A.M., Nolan, D., Gaudieri, S., Almeida, C.A., Nolan, R., James, I., Carvalho, F., Phillips, E., Christiansen, F.T., Purcell, A.W. *et al.* (2004) Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. *Proc. Natl Acad. Sci. USA*, **101**, 4180–4185.
 22. Schroeder, S., Reck, M., Lehmann, L.E., Book, M., Hoeft, A. and Stuber, F. (2000) The PstI polymorphism of the endotoxin-inducible heat-shock protein 70-2 gene does not affect messenger RNA level in human whole-blood cultures. *Intensive Care Med.*, **26**, 1139–1143.
 23. Fourie, A.M., Peterson, P.A. and Yang, Y. (2001) Characterization and regulation of the major histocompatibility complex-encoded proteins Hsp70-Hom and Hsp70-1/2. *Cell Stress Chaperones*, **6**, 282–295.
 24. Milarski, K.L. and Morimoto, R.I. (1986) Expression of human HSP70 during the synthetic phase of the cell cycle. *Proc. Natl Acad. Sci. USA*, **83**, 9517–9521.
 25. Singh, R., Kolvraa, S., Bross, P., Jensen, U.B., Gregersen, N., Tan, Q., Knudsen, C. and Rattan, S.I. (2006) Reduced heat shock response in human mononuclear cells during aging and its association with polymorphisms in HSP70 genes. *Cell Stress Chaperones*, **11**, 208–215.
 26. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
 27. Li, J., Niu, W., Qi, Y., Mayila, W., Zhu, P., Muhuyati, Cheng, Z. and Qiu, C. (2009) Interactive association of heat shock protein 70 genes variants with natural longevity in Xinjiang Hetian Uygur ethnicity. *Transl. Res.*, **154**, 257–264.
 28. Waterer, G.W., ElBahawan, L., Quasney, M.W., Zhang, Q., Kessler, L.A. and Wunderink, R.G. (2003) Heat shock protein 70-2+1267 AA homozygotes have an increased risk of septic shock in adults with community-acquired pneumonia. *Crit. Care Med.*, **31**, 1367–1372.
 29. Singh, R., Kolvraa, S., Bross, P., Christensen, K., Bathum, L., Gregersen, N., Tan, Q. and Rattan, S.I. (2010) Anti-inflammatory heat shock protein 70 genes are positively associated with human survival. *Curr. Pharm. Des.*, **16**, 796–801.
 30. Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J. and Donnelly, P. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
 31. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D. *et al.* (2007) Population genomics of human gene expression. *Nat. Genet.*, **39**, 1217–1224.
 32. Cheung, V.G. and Spielman, R.S. (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat. Rev. Genet.*, **10**, 595–604.
 33. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
 34. Teo, Y.Y., Small, K.S. and Kwiatkowski, D.P. (2010) Methodological challenges of genome-wide association analysis in Africa. *Nat. Rev. Genet.*, **11**, 149–160.
 35. Clayton, A., Turkes, A., Navabi, H., Mason, M.D. and Tabi, Z. (2005) Induction of heat shock proteins in B-cell exosomes. *J. Cell. Sci.*, **118**, 3631–3638.
 36. Leppa, S., Kajanane, R., Arminen, L. and Sistonen, L. (2001) Differential induction of Hsp70-encoding genes in human hematopoietic cells. *J. Biol. Chem.*, **276**, 31713–31719.
 37. Dressel, R. and Gunther, E. (1999) Heat-induced expression of MHC-linked HSP70 genes in lymphocytes varies at the single-cell level. *J. Cell. Biochem.*, **72**, 558–569.
 38. Yao, Y., Nellaker, C. and Karlsson, H. (2006) Evaluation of minor groove binding probe and Taqman probe PCR assays: influence of mismatches and template complexity on quantification. *Mol. Cell. Probes*, **20**, 311–316.
 39. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
 40. Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y. and Pritchard, J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.
 41. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Nilen, E., Yusuf, D., Lenhard, B., Wasserman, W.W. and Sandelin, A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **38**, D105–D110.
 42. Hanchard, N., Rockett, K., Udalo, I., Wilson, J., Keating, B., Koch, O., Nijnik, A., Diakite, M., Herbert, M. and Kwiatkowski, D. (2006) An investigation of transmission ratio distortion in the central region of the human MHC. *Genes Immun.*, **7**, 51–58.
 43. Averaimo, S., Milton, R.H., Duchon, M.R. and Mazzanti, M. (2010) Chloride intracellular channel 1 (CLIC1): sensor and effector during oxidative stress. *FEBS Lett.*, **584**, 2076–2084.
 44. Xian Ma, Y., Fan, S., Xiong, J., Yuan, R.Q., Meng, Q., Gao, M., Goldberg, I.D., Fuqua, S.A., Pestell, R.G. and Rosen, E.M. (2003) Role of BRCA1 in heat shock response. *Oncogene*, **22**, 10–27.
 45. Valdes, A.M. and Thomson, G. (2009) Several loci in the HLA class III region are associated with T1D risk after adjusting for DRB1-DQB1. *Diabetes Obes. Metab.*, **11**, 46–52.
 46. McElroy, J.P., Cree, B.A., Caillier, S.J., Gregersen, P.K., Herbert, J., Khan, O.A., Freudenberg, J., Lee, A., Bridges, S.L. Jr, Hauser, S.L. *et al.* (2010) Refining the association of MHC with multiple sclerosis in African Americans. *Hum. Mol. Genet.*, **19**, 3080–3088.
 47. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
 48. Ewing, B., Hillier, L., Wendt, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
 49. Gordon, D., Abajian, C. and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.*, **8**, 195–202.
 50. Gordon, D., Desmarais, C. and Green, P. (2001) Automated finishing with autofinish. *Genome Res.*, **11**, 614–625.
 51. Njemini, R., Demanet, C. and Mets, T. (2004) Inflammatory status as an important determinant of heat shock protein 70 serum concentrations during aging. *Biogerontology*, **5**, 31–38.
 52. Njemini, R., Demanet, C. and Mets, T. (2003) Determination of intracellular heat shock protein 70 using a newly developed cell lysate immunometric assay. *J. Immunol. Methods*, **274**, 271–279.
 53. Njemini, R., Demanet, C. and Mets, T. (2005) Comparison of two ELISAs for the determination of Hsp70 in serum. *J. Immunol. Methods*, **306**, 176–182.
 54. Pockley, A.G., Shepherd, J. and Corton, J.M. (1998) Detection of heat shock protein 70 (Hsp70) and anti-Hsp70 antibodies in the serum of normal individuals. *Immunol. Invest.*, **27**, 367–377.
 55. Pockley, A.G., Wu, R., Lemne, C., Kiessling, R., de Faire, U. and Frostegard, J. (2000) Circulating heat shock protein 60 is associated with early cardiovascular disease. *Hypertension*, **36**, 303–307.
 56. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 57. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
 58. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
 59. Manolio, T.A. (2010) Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.*, **363**, 166–176.
 60. WTCCC. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.