and individuals. Most of the UK samples were obtained with written consent from patients or next of kin; however, where this was not available, for example, for archival vCJD tissue obtained at post-mortem examination, we obtained the specific approval of our local ethics committee for the use of these samples in the research.

Procedures

For the samples from patients with vCJD, genomic DNA was mostly extracted from peripheral blood, although 45 samples were extracted from brain tissue. For a few samples, whole-genome amplification, either with a Φ29 protocol called multiple displacement amplification (MDA; Geneservice, Cambridge, UK; ten samples) or GenomePlex Complete Whole Genome Amplification Kit (WGA2; Sigma, UK; two samples) was necessary.

For the samples from patients with sCJD or iCJD, whole-genome amplification with either MDA in 138 samples or WGA2 in 29 samples was needed. Most of the samples from patients with sCJD and iCJD were extracted from blood, although DNA from eight samples in the iCJD group and 70 samples from the sCJD group was derived from brain tissue. 112 samples from patients with sCJD were sent as DNA to the MRC Prion Unit for analysis, most of which were extracted from blood. Genomic DNA was usually extracted from peripheral blood. PAXgene blood-derived RNA samples were also collected (Reanalytix, QIAGEN, UK).

DNA from degraded archival kuru sera was isolated by QIAamp Blood DNA minikit (QIAGEN, UK) followed by whole-genome amplification with WGA2 in all but seven samples. The validation of this process for the degraded kuru samples has been reported elsewhere.[16]

Good-quality genomic DNA extracted from blood was available for 278 of 285 (98%) healthy controls from Papua New Guinea and 122 of 125 (98%) healthy elderly women with many exposures to kuru at mortuary feasts. All control samples from Papua New Guinea were extracted from blood.

All DNA samples were checked for degradation on 1% agarose gel and stored at 50 ng/μL in low-concentration tris-EDTA buffer.

Rocky Mountain Laboratory (RML) prion-infected mouse brain homogenate (0·001%) or mock-infected brain homogenate from wild-type CD-1 mice (0·001%) was used to infect GT-1 hypothalamic neuronal cells. 5000 cells were seeded into 96-well plates and incubated with either homogenate in standard growth medium (Opti-MEM supplemented with 10% fetal calf serum and 1% penicillin/streptomycin [Invitrogen, CA, USA]). The inoculum was removed after 3 days and the cells were split 1:8. Cells were then split 1:8 twice more at intervals of 3 days. High levels of prion infectivity were confirmed with the scrapie cell assay.[17] Prion-infected and mock-infected cells were maintained in standard growth medium at 37°C in 5% CO_2. Total RNA was extracted in triplicate with the RNeasy Midi kits (QIAGEN, UK) according to the manufacturer's instructions, from prion-infected and mock-infected cells that had been grown on 10-cm-diameter plates. RNA was eluted in RNAase-free water and stored at −80°C. RNA samples adjusted to a concentration of 250 ng in 5 μL were incubated at 50°C for 30 mins with an equal volume of Glyoxyl (Ambion, Warrington, UK) loading dye containing ethidium bromide. Samples were run on a 1·5% agarose mini-gel in 1x NorthernMAx glyoxyl-based gel prep and running buffer (Ambion) at 100 mV for 90 min to check sample integrity. RNA was then sent to AROS Applied Biotechnology AS (Denmark) for the following microarray analyses according to Affymetrix standard protocols: first and second strand complementary DNA synthesis was done with the SuperScriptII System (Invitrogen) from 5 μg RNA (a minor modification was made to the protocol by using an oligo-dT primer that contained a T7 RNA polymerase promoter site); labelled antisense RNA (cRNA) was prepared with the BioArray High Yield RNA Transcript Labelling Kit with biotin-labelled CTP and UTP (Enzo Life Sciences, NY, USA) and unlabelled NTPs. Unincorporated nucleotides were removed using RNeasy columns (QIAGEN). 15 μg of cRNA was fragmented, loaded on to the Affymetrix mouse expression array 430_2.0 probe array cartridge, and hybridised for 16 h. Arrays were washed, stained in the Affymetrix fluidics station and scanned with

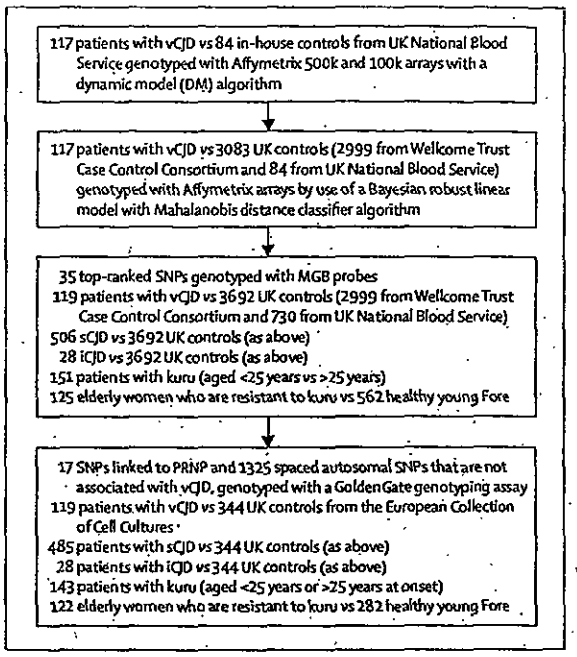

Figure 1: Flowchart of the genotyping in the tiered study
For each tier, the patient and control sample collections used are subsets of those genotypes in the minor groove-binding (MGB) probe study. In the first two tiers, the 117 samples from patients with vCJD are a subset of the 119 used in tiers three and four. In the second tier, the 84 samples from the UK National Blood Service are a subset of the 730 used in the third tier. In the final tier, the 485 samples from patients with sCJD, the 143 samples from patients with kuru, the 122 samples from elderly women, and the 282 samples from healthy young Fore are all subsets of the samples used in the third tier.

a confocal laser-scanning microscope (GeneChip Scanner 3000 System with Workstation and Autoloader).

The following sample comparisons were made in the association studies: vCJD versus UK controls genome-wide with Affymetrix array data; vCJD, sCJD, and iCJD versus UK controls in a validation and replication study with minor groove-binding [MGB] probes; healthy elderly women who were exposed to kuru at mortuary feasts versus geographically matched young individuals from the Eastern Highlands of Papua New Guinea in the replication study; young patients with kuru versus older kuru patients in the replication study. Figure 1 shows the tiered nature of the study. Subsets of each sample group have been used in previous studies of *PRNP* codon 129.[18] The comparison of young versus old patients with kuru was based on a hypothesis derived from mouse models that states that genetic factors control the incubation time of human prion diseases.[12] The incubation time of middle-aged or elderly patients who died of kuru at the peak of the epidemic cannot be calculated with precision and might have been many decades. Incubation times of up to 50 years or longer have been recorded in recently diagnosed patients,[5] whereas children, adolescents, or young adults have a limited incubation time.[5] Because the kuru collection was a mixture of samples from young and old patients, we hypothesised a priori that greater differences would be found between young people with kuru and old people with kuru than between people with kuru versus modern young healthy Fore. This strategy was supported by the precedent of homozygosity at codon 129 of *PRNP*, which was strongly associated with young versus old kuru, but was not significant in a comparison of all kuru with healthy Fore.

For the PLINK toolset see http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml

Genotyping and statistical analysis

We used the Affymetrix 100K and 500K arrays (early access, EA-500K), which use four restriction enzymes in total. Our first case–control study used data generated by the Affymetrix DM (dynamic model) algorithm from 117 samples of patients with vCJD (two samples were not suitable for use with Affymetrix arrays) and 90 UK controls matched for birthplace. The 500K product is comprised of two arrays each of about 250K digested with the restriction endonucleases *Nsp*I or *Sty*I; the 100K product is comprised of two arrays each of about 50K digested with the restriction endonucleases *Xba*I and *Hind*I. Genotypes were called by the dynamic model (DM) and subsequently by BRLMM algorithms. Samples from patients with vCJD were repeated if the DM call rate was less than 85% or the BRLMM call rate was less than 90% and samples were excluded if they underperformed by these criteria (vCJD [n=0], NBS [n=6], WTCCC [n=5]). The median and mean BRLMM call rates (all non-WTCCC samples and all arrays) were 99·0% and 98·5%. No samples were excluded for excess or low heterozygosity. One duplicate sample but no related individuals were identified. With genome-wide SNP data, the PLINK toolset for whole-genome association and population-based linkage analysis enables estimates of the relatedness of individuals. For the purposes of confirming unrelatedness, this can be expressed as a probability for identity by descent (IBD)=0 using complete linkage agglomerative clustering. This probability was greater than 0·75 for all study pairwise comparisons. No samples were identified as ethnic outliers by use of identity by state clustering.

Genotype data quality analysis and filtering was done with PLINK. From 598 676 unfiltered SNPs, the following were excluded from further analysis by standard quality control: monomorphic SNPs or those not genotyped by EA-500K or WTCCC arrays (n=170 334); greater than 10% missing genotypes in vCJD (n=66 659) or WTCCC (n=9828); evidence of Hardy–Weinberg disequilibrium (exact test, $p<0·001$) in our UK samples (n=4873) or (exact test, $p<1·0\times10^{-5}$) in WTCCC samples (n=7673); minor allele frequency less than 0·01 in vCJD and WTCCC samples (n=57 853); allelic test for differences in our in-house UK samples versus WTCCC ($p<0·001$; n=1888). After this trimming, 410 287 SNPs remained for testing in 117 patients with vCJD versus 3083 UK controls (84 in-house UK controls and 2999 WTCCC samples). A more stringent filter applied additional thresholds of less than 3% missing data overall, and minor allele frequencies greater than 3% (n=288 908 SNPs remaining). These stringently filtered data were assessed for whether the skewed quartile–quartile (QQ) plots (figure 2) were caused by cryptic population stratification between the UK control and vCJD groups or alternatively by inaccurate SNP genotyping. The absence of a significantly skewed QQ plot in the stringently filtered data supports the hypothesis that SNPs were inaccurately called, probably on the EA-500K platform. Subsequently this was confirmed by concordance testing with the GoldenGate platform.

Candidate SNPs for further study were identified in stringently filtered dataset or after standard filtering if there was additional evidence of genotype accuracy, by identifying an association signal in nearby SNPs in strong linkage disequilibrium with the candidate SNP. As a further test to identify false-positive associations related to differential genotyping accuracy between cases and controls, we validated (>99% concordance) all genotypes shown in vCJD and in-house UK controls with an independent platform (MGB probe and quantitative PCR) before attempting replication in other categories of prion disease with the same technology (35 SNPs were tested in this way). To maximise coverage, a further 17 SNPs were chosen from the *PRNP* locus (by maximising pairwise r^2 with HapMap build 35 using Haploview) and genotyped with GoldenGate technology. In total 52 SNPs were genotyped for association studies further to the discovery phase.

PLINK was used for association and permutation testing. The primary analysis was an allelic χ^2 test with use of empirical p values if any cell count was less than 15. A secondary analysis implemented genotypic,

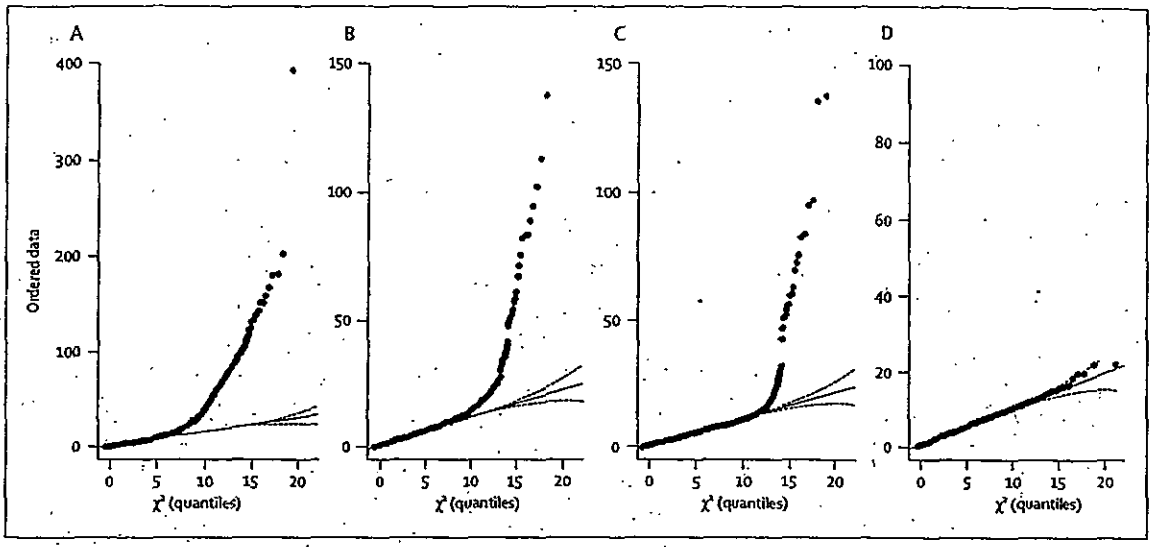

Figure 2: QQ (quartile–quartile) plots of different stages of quality control
(A) Unfiltered allelic χ^2 test of vCJD samples versus all UK samples (internal and Welcome Trust Case Control Consortium [WTCCC]) data. (B) Standard filtering allelic χ^2 test of vCJD versus all UK samples (internal and WTCCC data). (C) Standard filtering with allelic χ^2 test of vCJD versus only internal UK samples. (D) High-stringency filtering. With standard filtering, the inflation factor used for genomic control of confounding factors was estimated as 1·06 (1·01–1·09). Red dots=observed data. Blue lines=expected data. Broken blue lines=95% CI for expected data.

dominant, and recessive models, with empirical significance if necessary, controlling for the four tests done. Imputation of codon 129 genotype was done by the PLINK proxy-impute command (multimarker tagging) with dense SNP data around *PRNP*, including rs1799990, generated in 344 in-house UK controls. A nominal genome-wide significance threshold of $p<5\times10^{-7}$ was used in the primary analysis in concordance with the WTCCC. Owing to the large number of SNPs that were tested, this threshold takes into account multiple-hypothesis testing. In the replication phase of the study, the small number of tested SNPs permits a less stringent threshold of $p<0·001$.

Population structure was analysed with IBS clustering (implemented through PLINK) and principle components analysis (implemented through the Eigenstrat package[19]). Genome-wide data that were filtered to high stringency were used to compare samples from patients with vCJD and UK controls with PLINK and Eigenstrat (no significant eigenvectors were detected with default procedures). A separate low-density study was done with GoldenGate technology for several reasons: to investigate genotype accuracy in the SNPs filtered out by the high-stringency filtering step; to provide evidence with regard to the population structure in the samples from Papua New Guinea, which was previously unknown; to provide evidence that concordant genotypes consistent with the healthy population frequencies could be obtained from the amplified degraded samples from patients with kuru. 17 additional SNPs were genotyped to provide dense coverage of the *PRNP* locus, which is a region that confers susceptibility to prion disease, with a high probability of novel susceptibility being discovered here. These 17 SNPs complemented an existing dataset of 25 SNPs from the earlier genome-wide phase of the study. SNPs were selected from standard stringency filtered data in the genome-wide phase of the study; all autosomes were equally represented, with a median intermarker distance of 1·3 Mb. 1523 individuals were genotyped for 1325 SNPs: 344 randomly selected, non-related, white blood donors from the UK provided by the ECCCs; 119 patients with vCJD; 485 patients with sCJD; 28 patients with iCJD; 143 patients with kuru; 122 elderly women who are resistant to kuru and were born before 1950; and 282 young individuals from the kuru region matched to the elderly women by village of residence (figure 1). These patients were a subset of those included in the replication studies. SNPs were filtered for association with vCJD by comparison with UK controls by best permuted $p<0·001$ from any of four genetic models (allelic, trend, genotypic, or recessive) with the GoldenGate platform at the St Bartholomew's Hospital Genome Centre. Genotyping quality was assessed by Hardy–Weinberg equilibrium (excluding those assessed by exact test $p<0·001$) and visual inspection of all genotype clusters with Beadstudio version 3.1. The overall genotype call rate was 99·7%, and concordance of duplicate samples was excellent (nine WGA2 degraded amplified kuru samples [concordance 99·7%] and 20 healthy control duplicates [concordance >99·9%]). This study confirmed that the skew in the QQ plots was caused by inaccurate genotyping of SNPs in our genome-wide study that were not adequately filtered by the low-stringency criteria. For Eigenstrat, ten eigenvectors were

For more on the Genome Centr
see http://www.
bartsandthelondon.nhs.uk/
research/core_facilities_to_
support_research.asp#genome

	Chromosome	Locus	Minor allele	Major allele	vCJD genotypes	UK control genotypes	Model	p (vCJD)	OR
rs1799990	20	4628251	G	A	119/0/0	294/324/81	A	$2·0\times10^{-27}$	..
rs6107516	20	4625092	A	G	117/2/0	1960/1227/227	A	$2·5\times10^{-17}$	38·5 (9·6–155·2)
rs6116492	20	4646626	T	G	104/12/1	2979/104/0	A	$8·2\times10^{-5}$	3·71 (2·09–6·59)
rs1460163	8	80390003	A	G	7/25/86	31/657/2734	R	$5·6\times10^{-5}$	6·9 (3·0–16·0)
rs6794719	3	24777543	T	A	3/31/84	346/1465/1596	A	$1·9\times10^{-7}$	2·5 (1·7–3·7)

Genetic models: A=allelic; G=genotypic.

Table 1: Discovery tests of rs1799990 and four novel candidate SNPs in patients with vCJD and UK controls

	Resistance to kuru despite exposure				Early-onset and late-onset kuru				Combined p value*
	Elderly women exposed to kuru	Healthy young Fore	Model	p value	Young kuru (age <25 years)	Older kuru (age >25 years)	Model	p value (kuru incubation)	
rs1799990	16/86/23	112/287/163	G	0·001	16/18/25	9/71/12	G	$9·1\times10^{-4}$	$2·2\times10^{-5}$
rs6116492	80/37/2	393/151/18	A	0·848	47/11/0	57/20/0	A	0·44	..
rs1460163	30/77/16	140/144/40	R	$2·5\times10^{-4}$	23/29/5	26/42/23	A	0·017	$5·7\times10^{-5}$
rs6794719	5/47/63	26/118/136	A	0·12	4/25/27	7/33/47	A	0·687	..

Genetic models: A=allelic; G=genotypic; R=recessive. *Fisher's method.

Table 2: rs1799990 and three novel candidate SNPs in samples from Papua New Guinea

	iCJD genotypes	UK control genotypes	Model	p (iCJD)	sCJD genotypes	Model	p (sCJD)
rs1799990	4/13/11	294/324/81	A	$2·7\times10^{-5}$	307/98/101	G	$2·3\times10^{-22}$
rs6107516	9/12/7	1960/1227/227	R	0·002	320/100/55	G	$3·6\times10^{-33}$
rs6116492	28/0/0	2979/104/0	A	0·62	483/22/0	A	0·30
rs1460163	0/7/21	31/657/2734	A	0·66	7/93/396	A	0·83
rs6794719	1/8/19	346/1465/1596	A	0·03	61/207/207	A	0·08

Genetic models: A=allelic; G=genotypic; R=recessive.

Table 3: Replication tests of rs1799990 and four novel candidate SNPs in patients with iCJD, patients with sCJD, and UK controls

generated through default procedures and outlier detection (6 of 826 samples from Papua New Guinea were removed). No significant eigenvectors (p>0·01) were identified between patients with sCJD or iCJD and UK controls, or between patients with kuru, elderly women who are resistant to kuru, and healthy young Fore (five comparisons in total).

The Gene Expression Analysis Software (MAS 5.0) was used to analyse the raw image files from the quantitative scanning, which resulted in files that contained background corrected values for the probes. Significance analyses to compare prion versus mock-infected cells used a two-class unpaired test with a Benjamini–Hochberg (false discovery rate) p-value correction.

Role of the funding source

The sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. Simon Mead and John Collinge had full access to all the data in the study and final responsibility for the decision to submit for publication.

Results

After standard quality control for call rate and minor allele frequency, Hardy–Weinberg disequilibrium, and differences between control datasets, we analysed 410 287 SNPs in the primary analysis. QQ plots showed an excess of large allelic χ^2 test statistics (figure 2) owing partly to the comparison between cases and controls among platforms and laboratories and was completely resolved by stringent filtering by call rate and minor allele frequency, leaving about 300K SNPs for association testing. Comparison of these stringently filtered data between vCJD and our own UK controls, ECACC controls, and WTCCC controls with Eigenstrat and GC methods did not provide evidence of significant population stratification; therefore, association statistics were not corrected. Similarly, we found no evidence of population stratification in comparisons of 1325 SNPs in replication cohorts (sCJD, iCJD, and control groups from the UK) or between patients with kuru, elderly women who were resistant to kuru, and healthy Fore from Papua New Guinea. In the stringently filtered data, two SNPs were significant at the genome-wide level ($p<5\times10^{-7}$) on the basis of allelic tests: rs6107516 in the intron of *PRNP* and rs6794719 in an intergenic region between *RARB* and *THRB*, which encodes thyroid hormone receptor beta (figure 3).

A block of linkage disequilibrium that was larger than ~100 kb and included all of *PRNP* was shown by 25 SNPs (figure 3). We added 17 more SNPs from vCJD and UK controls, including *PRNP* codon 129 (rs1799990). Unsurprisingly, rs6107516, which is located in the intron of *PRNP* and is in moderately strong linkage disequilibrium with codon 129 (rs1799990, $r^2=0·6$), was the top-ranked single SNP in the discovery phase. All patients with vCJD who have been genotyped to date are homozygous for

rs1799990A (MM at *PRNP* codon 129). rs6107516 was most strongly associated in an allelic model ($p=2\cdot5\times10^{-17}$; OR 38·5, 95% CI 9·6–155·2).

From the 25 SNPs in the discovery phase, codon 129 of *PRNP* was best tagged by a two SNP haplotype formed by rs6031692 and rs6107516 ($r^2=0\cdot7$, based on Hapmap build 35 data) with a haplotypic association of $p=1\times10^{-24}$. To test for more association at the locus, we conditioned for the association of codon 129 by imputing this genotype and including only methionine-homozygous UK controls from the WTCCC series. We thus identified evidence of additional genetic risk at this locus. rs6116492, which is downstream of *PRNP* and also in strong linkage disequilibrium with rs1799990, had a frequency of 0·06 in patients with vCJD and 0·017 in UK controls (allelic model $p=8\cdot2\times10^{-5}$; 0·022 in 1544 UK controls with an imputed codon 129 methionine homozygous genotype; allelic model p=0·001, OR 2·63, 95% CI 1·43–4·82). rs6116492 is located in an intergenic region between *PRNP* and *PRND*, which encodes prion-like protein doppel. Genetic risk factors for sCJD have previously been identified upstream and downstream of *PRNP* but not for vCJD.[20,21] Because we cannot guarantee that the rs1799990 genotype has been imputed perfectly, we also compared cases of vCJD (n=119) with in-house UK controls genotyped at rs1799990 and rs6116492 (n=701); we again found a significant, independent association of rs6116492, both by haplotype test conditioned on codon 129 (PLINK, likelihood ratio test with one degree of freedom; p=0·037), or simply by excluding controls with methionine/valine or valine/valine encoded at codon 129 genotypes followed by an allelic test (119 patients with vCJD *vs* 294 in-house UK controls with the genotype that encodes methionine/methionine at codon 129; p=0·02). SNP-1368 (rs1029273C, 24466 base pairs upstream of codon 129), which we and others have confirmed to be associated with sporadic CJD but not vCJD, also showed no evidence of association with vCJD independent of codon 129.[20,21]

Because we tested the entire collection of samples from white British patients with vCJD (the majority of cases of vCJD), we then looked to closely related prion diseases to replicate independently and more broadly candidate SNPs with risk of prion disease. We tested for the association of rs6794719, rs6116492, and 33 other top-ranking SNP associations from the vCJD study in patients with iCJD who were exposed to prion disease through cadaver-derived growth hormone therapy versus UK controls (n=28); 506 patients with sCJD—a worldwide disease of uniform incidence that affects about 1–2 million people per year—versus UK controls. We also tested patients with kuru and healthy elderly women who were exposed to but survived the kuru epidemic from the Eastern Highlands Province of Papua New Guinea. In the groups from Papua New Guinea, we tested whether candidates for genetic risk of vCJD were associated with kuru incubation time (by comparison with a cohort of young-onset kuru [n=59] versus old-onset kuru [n=92]) or resistance to kuru, by comparison of elderly female survivors (n=125) with the young population (n=280–526). Homozygosity at codon 129 of *PRNP* was significantly associated with risk of iCJD ($p=2\cdot7\times10^{-4}$), sCJD ($p=2\cdot3\times10^{-21}$), and tests done in Papua New Guinea (Fisher's method $p=2\cdot2\times10^{-9}$; tables 1 and 2).

rs6794719A was associated with the risk of vCJD at a nominal genome-wide significance of $p=1\cdot9\times10^{-7}$. The more frequent allele, rs6794719A, was also associated with disease risk in the small collection (n=28) of patients with iCJD (p=0·030; table 3) but not those with sCJD, kuru, or resistance to kuru.

From the 33 top-ranked SNPs that failed to achieve genome-wide significance in patients with vCJD, the strongest overall evidence of association in replication cohorts was for rs1460163 (combined $p=6\cdot3\times10^{-8}$ by Fisher's method across orally acquired prion disease categories [combination of vCJD and Papua New Guinea

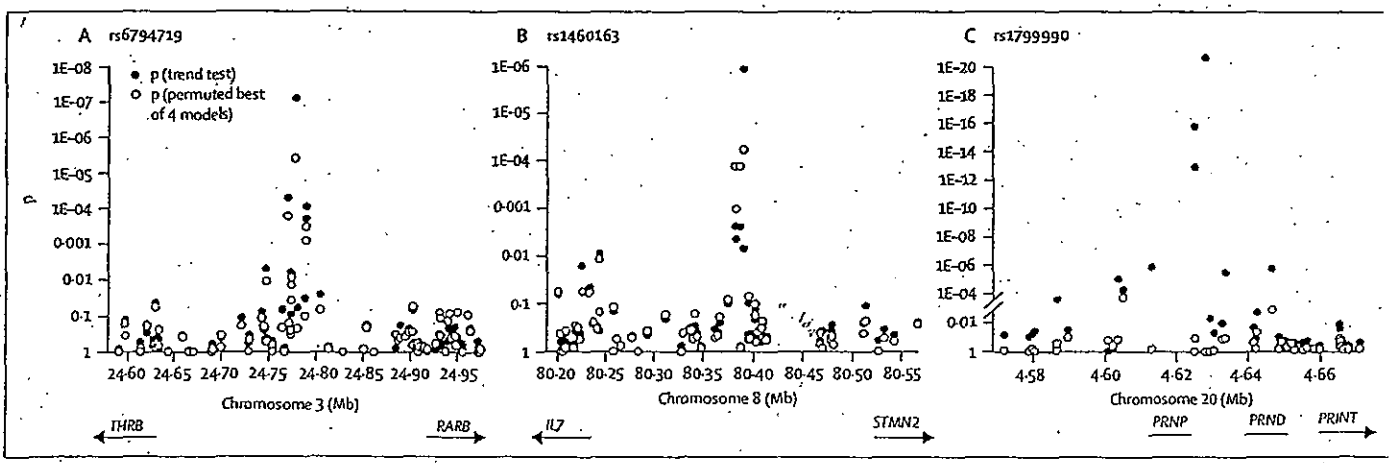

Figure 3: Physical location and p of allelic test and best of four genetic models
(A) SNPs between *THRB* and *RARB*, including rs6794719. (B) SNPs upstream of *STMN2* including rs1460163. (C) SNPs at the *PRNP* locus, including rs1799990, rs6107516, and rs6116492, showing trend test (filled circles) and a test comparing vCJD with UK controls with codon 129 methionine homozygous genotypes (empty circles [imputed for WTCCC controls]).