

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Mead S, Whitfield J, Poulter M, et al. A novel protective prion protein variant that colocalizes with kuru exposure. *N Engl J Med* 2009;361:2056-65.

Supplementary Appendix

Research ethics

The clinical and laboratory studies were approved by the Medical Research Advisory Committee of the Government of Papua New Guinea and by the local research ethics committee of the Institute of Neurology and National Hospital for Neurology and Neurosurgery in London.

Consent arrangements

Full participation of the communities involved was established and maintained through discussions with village leaders, communities, families and individuals. The field studies followed the principles and practice of the Papua New Guinea Institute of Medical Research, which included individual oral consent from all participants before any samples were obtained.

Kuru epidemiology and sampling

Kuru was a well defined condition diagnosed by the Fore people prior to Western medical intervention and has an extensively documented, highly characteristic and invariably fatal clinical course which presents little diagnostic difficulty in the population affected(1). Prior to 1987 kuru surveillance was conducted by many different investigators (Gajdusek, Zigas, Baker, Alpers, Hornabrook, Moir and others) and from 1987 to 1995 solely by the Kuru Surveillance Team of the Papua New Guinea Institute of Medical Research. From 1996 onwards a field base and basic laboratory for sample processing and storage was established in

the village of Waisa in the South Fore. A team of local kuru reporters communicated to the field base the details of any person suspected of suffering from kuru. The field team comprised MRC, PNGIMR and local staff and undertook regular field patrols throughout the kuru-affected area, which includes the North and South Fore, Keiagana, Kanite and Gimi linguistic groups. Less frequent visits were made to neighbouring linguistic groups with no documented kuru patients. Genealogies and venous blood samples were taken for molecular genetic studies from kuru patients and where possible from healthy family and community members, some of whom were exposed to multiple mortuary feasts. Genealogical information was obtained blind to genetic data. Although there is no formal documentation of age in Eastern Highlands populations this could be reliably calculated by comparison of an individual's personal history with well-known community historical events(2).

The exposure index (EI) for each village community was defined as the total number of kuru deaths per village, divided by estimated village population in 1958, times 1000. The EI was calculated for each village in the South and North Fore and the kuru-affected parts of the Gimi and Keiagana linguistic groups. For the smaller number of villages, and deaths, in other linguistic groups the EI was calculated for each linguistic group, based on the estimated kuru-affected population.

The kuru collection comprises young children, adolescents and young adults with short or average incubation times, and elderly recent kuru cases with long incubation times (see Supplementary Appendix figure 2 for geography). It is not possible to determine the incubation time of middle-aged or elderly individuals who died of kuru at the peak of the epidemic. We hypothesized that these subgroups might display different degrees of genetic kuru resistance or susceptibility and they were therefore analyzed separately. DNA from degraded archival kuru sera was isolated by QIAGEN QIAamp Blood DNA minikit followed by whole genome amplification either through using a Φ 29 protocol (Geneservice), or GenomePlex Complete Whole Genome Amplification Kit (WGA2) (Sigma). Elderly exposed women were defined by date of birth <1960 from a kuru-exposed region: South Fore (74), North Fore (36), Gimi (13) and Keiagana (2). The modern day healthy population from the exposed region was obtained by matching each elderly woman to at least two current residents of the same village aged born after 1960. These largely comprised the South Fore, but with a significant number from the North Fore and a small number of individuals from Gimi, Keiagana and Yagaria linguistic groups as indicated. Where identified by either genealogical data or SNP analysis, first degree relatives and cryptically duplicated individuals were excluded from these groups. Unexposed individuals were sampled from collections of Asaro (19), BenaBena (20), Gadsup (22), Gahuku (42), Labogai (47), Siane (28), Tairora (77), and Yabiyufa (26) linguistic groups with no oral history of kuru (see Supplementary Appendix figure 1).

Resequencing and genotyping

Genotyping at *PRNP* codons 127 and 129 was performed using Taqman MGB allelic discrimination probes. Automated fluorescent sequencing of the promoter region and open reading frame was performed on ABI-377, Amersham megaBACE instruments. Testing for *PRNP* deletion was performed by the δC_t method utilizing real-time PCR of genomic DNA using *PRNP* and β -actin Taqman MGB probes. Chi-squared tests, Hardy-Weinberg disequilibrium analysis and association study were performed using SPSS (SPSS inc.) and PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>).

Estimation of relative fitness

The equilibrium frequencies of three *PRNP* haplotypes were determined from genotype data assuming complete linkage disequilibrium between codon 127 and codon 129 (127VV has not been observed). Relative fitness was estimated in two ways (i) for codon 129 in isolation (ii) by modeling genotypes 127GV and 129MV as equally protective but not additive. For codon 129 in isolation, relative fitness was determined by dividing the observed genotype counts with genotype counts calculated from allele frequencies assuming HWE, and standardizing to 129MV. For analysis of joint genotype relative fitness, six possible joint genotypes were allocated to three groups according to their expected susceptibility to prion disease: (a) 127GG-129MM and 127VV-129MM (which was not observed), (b) 127GG-129VV, and (c) all heterozygous combinations 127GV-129MM, 127GG-129MV, and 127GV-129MV. Fitnesses of homozygous

combinations (groups (a) and (b)) were estimated by dividing the observed joint genotype counts by the predicted counts estimated from haplotype frequencies assuming HWE, standardizing to the sum of heterozygous combinations (c).

MRCAs analysis

The age of the G127 polymorphism was estimated using ESTIAGE(3). ESTIAGE estimates the most recent common ancestor of rare mutations by implementing likelihood based analysis of mutation-associated haplotype length in a small sample of individuals. The analysis parameters also include microsatellite allele frequency, mutation and recombination rate. Alternatively we used the formula in Risch(4), as corrected by Colombo(5): for a given marker, g , the age in generations is estimated from $g = \log(\delta)/\log(1 - \theta)$ where, if p_D is the frequency of a specified allele on V carrying chromosomes and p_N the frequency on G carrying chromosomes, $\delta = (p_D - p_N)/(1 - p_N)$, and $\theta =$ recombination fraction. We genotyped 13 microsatellites for these analyses: of these 4 were uninformative and one was excluded because of doubt over the genetic distance between this marker and *PRNP*. We use the median of the results for the remaining 8 markers as our point estimate of age, and provide confidence intervals based on 10000 bootstrap resamples of the data (see(6)). Estimate of 127V most recent common ancestor by ESTIAGE was 10 generations for 32 unrelated individuals (95% CI 7-15) by Risch' method was 12.8 generations (95% CI 0-30) with 51 samples including some known first degree relatives, and 15.3 generations with 32 samples from genealogically unrelated individuals.

Population structure

Population structure was considered by identity by state (IBS) clustering (implemented by PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml>) and principle components analysis (implemented by the EIGENSTRAT package(7)). The same genotype data was used for both analyses. Genotypes were generated for 1536 SNPs in 121 vCJD patients, 344 randomly selected, non-related UK Caucasian blood donors provided by the European Collection of Cell Cultures, 143 kuru patients, 122 women from the kuru region EI>30 born before 1950, 282 young individuals from the kuru region matched to the village of residence of the elderly women, and 275 individuals from multiple EHP populations with no oral history of kuru. We used the Illumina Goldengate platform at the St. Bartholomew's Hospital Genome Centre. SNPs were filtered for association with vCJD by comparison with UK controls by best permuted $P < 0.001$ from any of 4 genetic models (allelic, trend, genotypic or recessive). Genotyping quality was assessed by Hardy-Weinberg equilibrium (excluding those by exact test $P < 0.001$) and visual inspection of all genotype clusters with Beadstudio v3.1. Inspection was performed separately for WGA-genomeplex DNA (kuru 136/143 and individuals from multiple EHP populations with no oral history of kuru 95/275) and genomic DNA for other sample collections. Overall genotype call rate was 99.7% in the PNG samples (n=826). 1039 SNPs remained after filtering, all autosomes were equally represented with a median intermarker distance of 1.3MB. For PLINK, IBS pairwise distances were

calculated between individuals from the PNG collection using the --genome command, and a 4-dimensional scaling plot of IBS pairwise distances was generated from these data using the --mds-plot command. For EIGENSTRAT 10 eigenvectors were generated using default procedures and outlier detection (6/826 were removed).

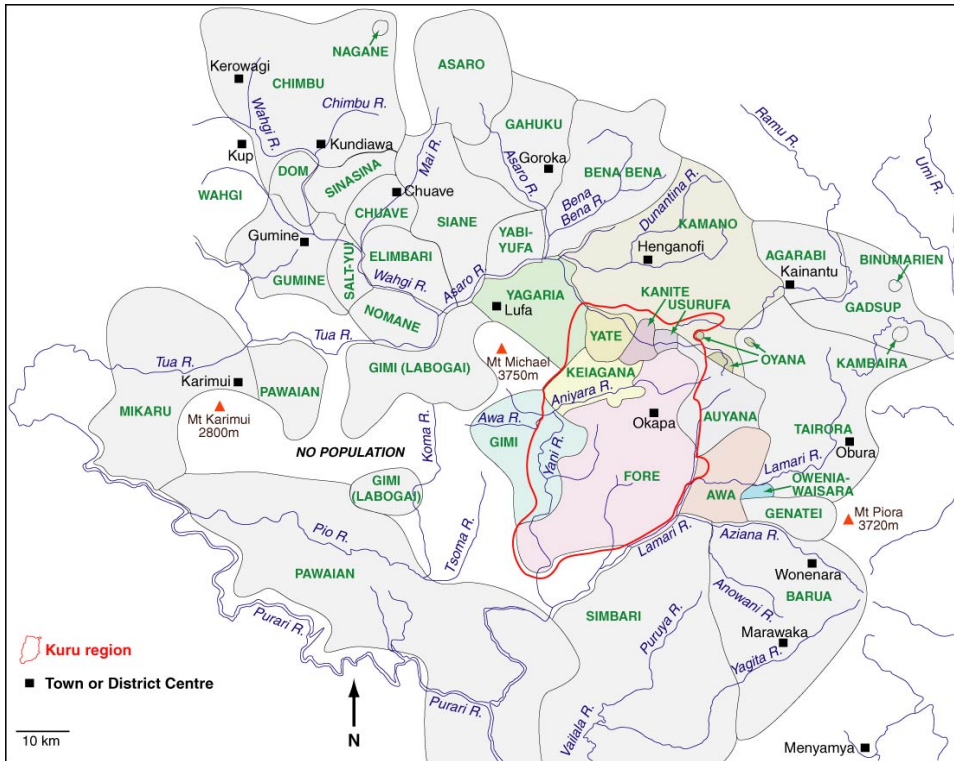
Analysis using PLINK IBS clustering or EIGENSTRAT generated very similar results (Supplementary Appendix figure 3). All sample collections from the kuru region (ie. kuru patients, elderly women and matched young individuals) were not visually distinguishable when plotting the first two (see Supplementary Appendix figure 3a,b) or three (not shown) axes. There was no statistical evidence of genetic differentiation between the three kuru region collections. In contrast, the kuru regions were strongly distinguished from multiple EHP populations with no oral history of kuru (each of three comparisons $P < 10^{-59}$). F_{ST} was between 0.000-0.001 (standard error 0.0002-0.0003) for the three kuru region populations and 0.008 (standard error 0.0005-0.0006) for comparison of any of the kuru region populations with multiple EHP populations with no oral history of kuru. Individuals with a 127GV genotype plotted towards the centre of the kuru region cluster (Supplementary Appendix figure 3b). Furthermore, genomic control on the same set of 1039 SNPs (kuru versus elderly women) yielded no evidence of population stratification ($\lambda=1.09$, 95%CI 0.98-1.20). These data provide evidence against population structure of the kuru region and we have therefore not corrected P values of association between kuru and elderly women. G127V was excluded

from the principle components analysis with EIGENSTRAT. When this was included and corrected for population structure the chisquare statistic (Armitage trend test) of association between kuru and elderly women reduced marginally from 7.1 to 6.6.

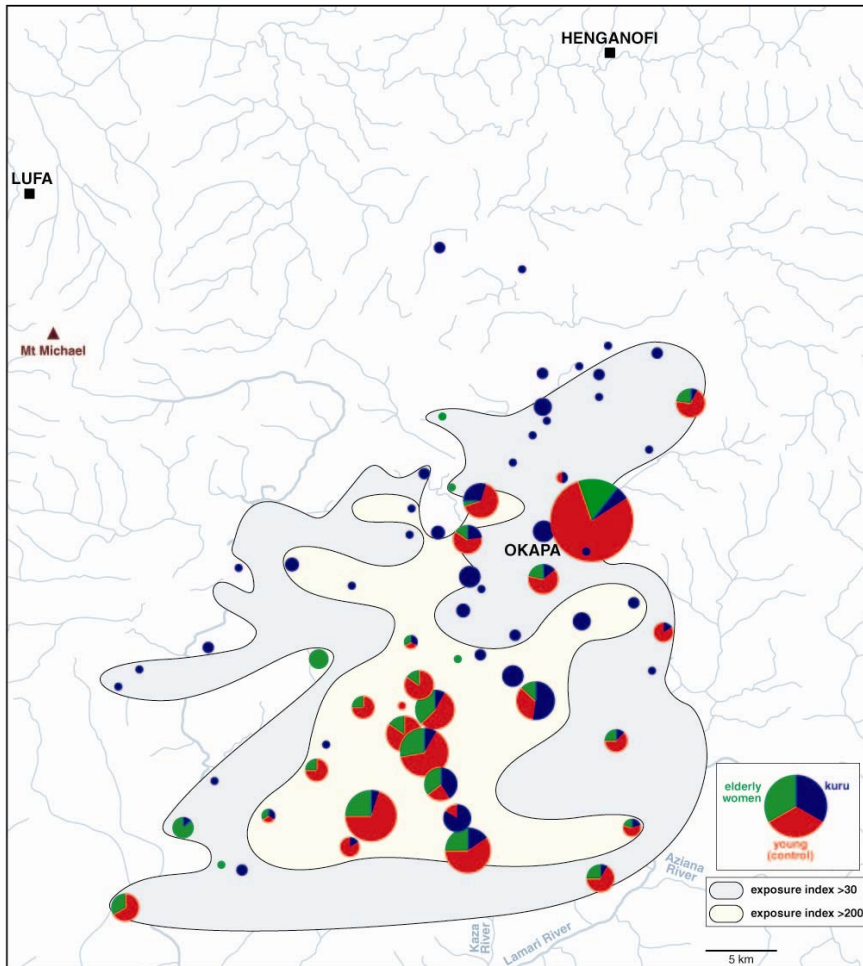
Disequilibrium in the elderly Fore between 127 and 129

Although our failure to observe any 127VV homozygotes in the elderly population is consistent with HWE, 127V alleles were not in equilibrium with codon 129 genotypes. Here we restrict our analysis to the 611 young and old individuals from the Purosa Valley where 127V is found (see Supplementary Appendix table 2). In those born before 1960, 10 127V alleles were associated with 30 129MM individuals whereas only 5 127V alleles were associated with 97 129MV individuals (Fisher's exact test, 10/50 vs. 5/92, $P = 0.02$), in marked contrast to the equilibrium distribution in young individuals from the Purosa valley (in whom 11 127V alleles were associated with 97 129MM individuals, 25 127V alleles were associated with 221 129MV individuals, 11/183 vs. 25/196, Fisher's exact test $P > 0.05$). Of six surviving 129MM women born before 1950 in the EI>200 zone, three were heterozygous for 127V (Fisher's exact test based on haplotype counts, 3/9 vs. 11/183, $P = 0.04$). These findings are consistent with 127V conferring resistance to kuru in otherwise susceptible 129MM individuals and allow us to estimate relative fitness. In the Purosa valley 127V-129M haplotype frequency is 0.167 (+/- 0.02) in 129MM individuals and 0.052 (+/- 0.005) in 129MV individuals. The greater than three-fold ratio in survival provides an

estimate of the relative fitness of the joint genotype 127GV-129MM vs. 127GG-129MM. In keeping with the association analysis, these data are most consistent with 127GV providing strong or complete resistance to kuru.

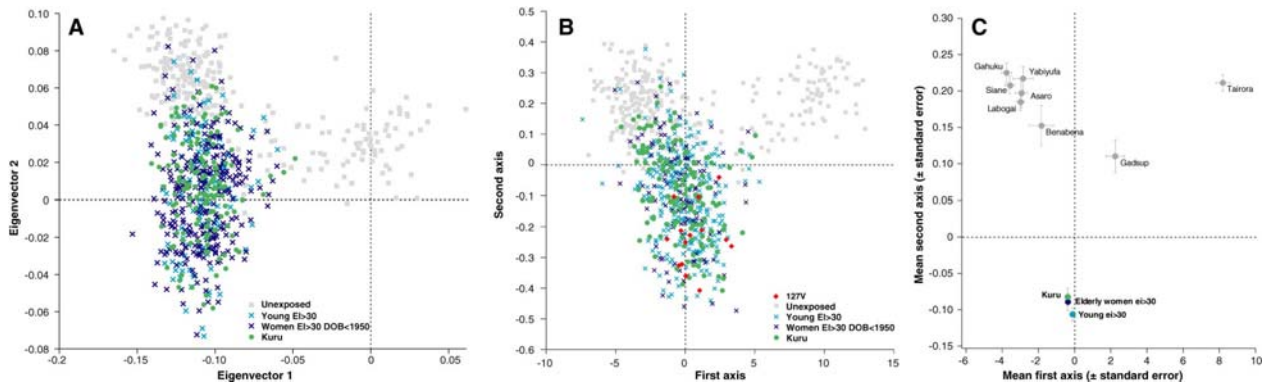


Supplementary Appendix Figure 1. Map of the Eastern Central Highlands Showing linguistic groups and the kuru region (Supplementary Appendix figures 2 and 3 are smaller scale maps of the kuru region)(8)



Supplementary Appendix Figure 2
Geographical matching of kuru, elderly women born before 1950, and young individuals.

Pie charts are shown for the number of persons in each category sampled from each village. The coloured area is proportional to the number of individuals sampled from this location.

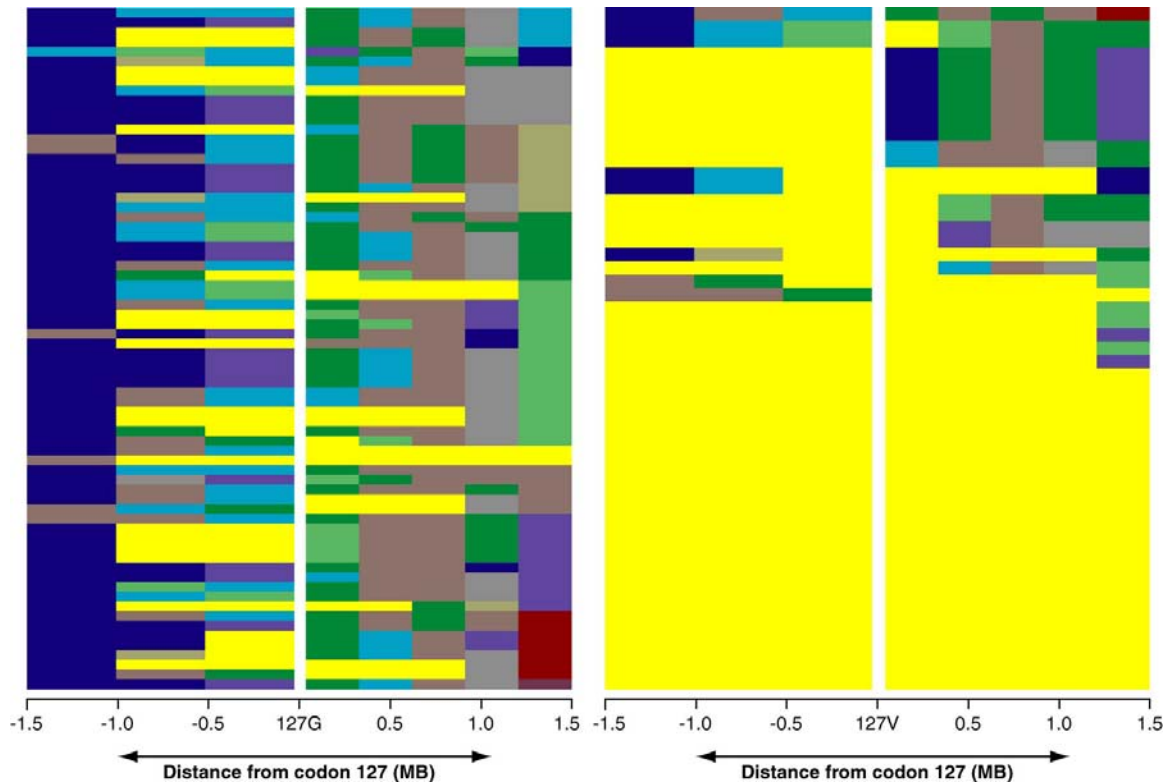


**Supplementary Appendix Figure 3
Population stratification analyses**

(A) Plot of the first two eigenvectors from analysis of 820 kuru patients, elderly women, young individuals from the kuru region and unexposed populations (EIGENSTRAT).

(B) Plot of the first two axes from analysis of 826 kuru patients, elderly women, young individuals from the kuru region and unexposed populations with identity-by-state clustering and multidimensional scaling (PLINK).

(C) Means (+/- standard errors) from (B) subdividing the unexposed populations by linguistic group (refer to Supplementary Appendix figure 1 for locations of the linguistic groups). Overall these data demonstrate population structure in EHP, but that this does not confound the association of 127V with resistance to kuru.



Supplementary Appendix Figure 4. Size of the 127V-linked haplotype (yellow) in 127V and 127G individuals.

8 informative microsatellites over 3MB in 120 chromosomes are illustrated. The distance axis represents the total length of 5' and 3' haplotypes, but is not an accurate description for each individual microsatellite location. Haplotypes were inferred with pedigree information and by PHASE software. A 127V linked haplotype undisturbed by recombination or mutation is shown in yellow. Other microsatellite alleles are shown in different colours (127G on the left is provided for comparison with 127V on the right).

Supplementary Appendix Table 1
PRNP open reading frame sequencing in the screening collection of kuru patients and EHP populations

Population	<i>n</i>	127GV	129MM	129MV	129VV	129V%
Kuru	112	0	29	63	20	46.0
Elderly women EI >30	125	6	16	86	23	52.8
Young EI>30	282	8	52	136	94	57.4
Unexposed EHP	281	0	73	141	67	48.9

Supplementary Appendix Table 2

Genotyping of codon 129 and 127 in healthy groups or populations

P vs. young EI >30 denotes the P value of a chi-squared test of 129V allele count between young individuals of the kuru region with exposure index >30 (marked with *) vs. other linguistic groups and populations. ^aRegarding individuals from the Purosa valley born after 1960, 97 were 129MM, 221 were MV. Of these 318 individuals, 25 had the joint genotype 127GV-129MV and 11 were 127GV-129MM (in marked contrast with the elderly from the same region, see text). 127V was not found as a 129VV genotype.

Group	N	P vs. young EI>30	127GV	127V%	GV-MM	129MM	129MV	129VV	129V%
EXPOSED GROUPS									
Purosa valley old	167	ns	15	0.045	10	30	97	40	0.53
Purosa valley young	444	ns	36	0.042	11	97	221	126	0.53
Purosa valley and neighbouring villages, all ages ^a	611	ns	51	0.042	21	127	318	166	0.53
All Old EI >30	480	ns	15	0.016	10	80	277	123	0.54
Kuru region young EI >30 (sequenced)	282	*	8	0.014	2	52	136	94	0.57
Gimi	87	0.012	0	0		25	43	19	0.47
Jate	157	ns	0	0		35	82	40	0.52
Keiagana	221	ns	0	0		42	106	73	0.57
Kanite	35	ns	0	0		5	18	12	0.60
Awa	46	ns	0	0		5	26	15	0.61
Kuru region	1308	ns	23	0	12	244	688	376	0.55
UNEXPOSED EHP									
Tairora	68	0.003	0	0		21	35	12	0.43
Gahuku	91	0.002	0	0		31	39	21	0.45
Siane	68	0.042	0	0		18	35	15	0.48
Benabena	46	ns	0	0		12	23	11	0.49
Labogai	65	ns	0	0		11	40	14	0.52
Asaro	26	ns	0	0		4	15	7	0.56
Morae	43	ns	0	0		10	19	14	0.55
Yabiyufa	92	ns	0	0		17	42	33	0.59
Gadsup	42	ns	0	0		4	26	12	0.60
Agarabi	90	ns	0	0		17	37	36	0.61
Unexposed EHP (genotyped)	631	0.04	0	0		145	311	175	0.52
Unexposed EHP (sequenced)	281	0.004	0	0		73	141	67	0.49
DISTANT POPULATIONS									
Vanimo & Wewak	5	0.003	0	0		4	1	0	0.10
Island	44	<0.001	0	0		31	12	1	0.16
Morseby	11	<0.001	0	0		7	4	0	0.18
Western Highlands	4	ns	0	0		2	2	0	0.25
Madang	239	<0.001	0	0		127	90	22	0.28
Lae	10	0.046	0	0		4	5	1	0.35
Coastal PNG	313	<0.001	0	0		175	114	24	0.26

Supplementary Appendix Table 3
Matching of 127V and 127G pedigrees

All villages are located within 8km of the centre of the Purosa valley. Each have an exposure index >100.

	127V Pedigrees	127V Parents	127V Documented death from kuru (%)	127G Pedigrees	127G Parents	127G Documented death from kuru (%)
All	18	36	1 (2.7)	112	218	33 (15.1)
Ai	4	8	0	0	0	0
Agakamtasa	2	4	0	0	0	0
Ilesa	1	2	0	0	0	0
Ivaki	1	2	0	23	46	4 (8.7)
Kalu	1	2	0	0	0	0
Kamira	0	0	0	29	57	15 (26.3)
Ketabi	6	12	0	6	12	3 (25)
Mugaimuti	1	2	0	23	45	5 (11.1)
Purosa- Takai	2	4	1	6	10	2 (20)
Takai	0	0	0	25	48	4 (8.3)

Reference List

1. Collinge J, Whitfield J, McKintosh E, Frosh A, Mead S, Hill AF et al. A clinical study of kuru patients with long incubation periods at the end of the epidemic in Papua New Guinea. *Philos Trans R Soc Lond B Biol Sci* 2008; 363(1510):3725-3739.
2. Collinge J, Whitfield J, McKintosh E, Beck J, Mead S, Thomas DJ et al. Kuru in the 21st century--an acquired human prion disease with very long incubation periods. *Lancet* 2006; 367(9528):2068-2074.
3. Genin E, Tullio-Pelet A, Begeot F, Lyonnet S, Abel L. Estimating the age of rare disease mutations: the example of Triple-A syndrome. *Journal of Medical Genetics* 2004; 41(6):445-449.
4. Risch N, Deleon D, Ozelius L, Kramer P, Almasy L, Singer B et al. Genetic-Analysis of Idiopathic Torsion Dystonia in Ashkenazi Jews and Their Recent Descent from A Small Founder Population. *Nature Genetics* 1995; 9(2):152-159.
5. Colombo R. Age and origin of the *PRNP* E200K mutation causing familial Creutzfeldt-Jacob disease in Libyan Jews. *Am J Hum Genet* 2000; 67(2):528-531.
6. Efron B. Bayes and Likelihood Calculations from Confidence-Intervals. *Biometrika* 1993; 80(1):3-26.
7. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 2006; 38(8):904-909.
8. Alpers MP. A History of Kuru. *P N G Med J* 2007; 50:10-19.