nature genetics

A thrifty variant in *CREBRF* strongly influences body mass index in Samoans

Ryan L Minster[1,13], Nicola L Hawley[2,13], Chi-Ting Su[1,12,13], Guangyun Sun[3,13], Erin E Kershaw[4], Hong Cheng[3], Olive D Buhule[5,12], Jerome Lin[1], Muagututi'a Sefuiva Reupena[6], Satupa'itea Viali[7], John Tuitele[8], Take Naseri[9], Zsolt Urban[1,14], Ranjan Deka[3,14], Daniel E Weeks[1,5,14] & Stephen T McGarvey[10,11,14]

Samoans are a unique founder population with a high prevalence of obesity[1–3], making them well suited for identifying new genetic contributors to obesity[4]. We conducted a genome-wide association study (GWAS) in 3,072 Samoans, discovered a variant, rs12513649, strongly associated with body mass index (BMI) ($P = 5.3 \times 10^{-14}$), and replicated the association in 2,102 additional Samoans ($P = 1.2 \times 10^{-9}$). Targeted sequencing identified a strongly associated missense variant, rs373863828 (p.Arg457Gln), in *CREBRF* (meta $P = 1.4 \times 10^{-20}$). Although this variant is extremely rare in other populations, it is common in Samoans (frequency of 0.259), with an effect size much larger than that of any other known common BMI risk variant (1.36–1.45 kg/m^2 per copy of the risk-associated allele). In comparison to wild-type CREBRF, the Arg457Gln variant when overexpressed selectively decreased energy use and increased fat storage in an adipocyte cell model. These data, in combination with evidence of positive selection of the allele encoding p.Arg457Gln, support a 'thrifty' variant hypothesis as a factor in human obesity.

Obesity is essentially a disorder of energy homeostasis and has strong genetic and environmental components. As diets have modernized and physical activity has decreased, the prevalence of overweight and obesity in Samoa has escalated to be among the highest in the world. In 2003, 68% of men and 84% of women in Samoa were overweight or obese by Polynesian cutoffs (BMI >26 kg/m^2)[1]; by 2010, prevalence had increased to 80% and 91%, respectively[3]. Although the contribution of environmental factors to this trend is clear, the estimated 45% heritability of BMI in Samoans remains largely unexplained[1]. Genetic susceptibility to obesity in the contemporary obesogenic environment may have resulted from putative selective advantages from efficient energy metabolism acquired during 3,000 years of Polynesian island discoveries, settlement, and population dynamics[5–8] and/or from genetic drift due to founder effects, small population sizes, and population bottlenecks[9–11].

To discover genes influencing BMI, we genotyped 659,492 markers across the genome in our discovery sample of 3,072 Samoans recruited from 33 villages across Samoa using the Affymetrix 6.0 chip (**Supplementary Fig. 1** and **Supplementary Table 1**). We adjusted for population substructure and inferred relatedness using an empirical kinship matrix and then tested for association with BMI using linear mixed models. Quantile–quantile plots indicated that *P*-value inflation was well controlled ($\lambda_{GC} = 1.07$) (**Supplementary Fig. 2**).

By far, the strongest association with BMI occurred at rs12513649 ($P = 5.3 \times 10^{-14}$) on chromosome 5q35.1 (**Fig. 1a**), and this association was strongly replicated ($P = 1.2 \times 10^{-9}$) in 2,102 adult Samoans from a 1990–1995 longitudinal study and a 2002–2003 family study, with participants of each study drawn from both American Samoa and Samoa (**Table 1** and **Supplementary Table 1**). To fine-map the region encompassing this signal, we used the Affymetrix-based genotypes to select 96 individuals optimal for targeted sequencing of a 1.5-Mb region centered on rs12513649. The haplotypes generated from the sequencing data were used to impute genotypes for the rest of the discovery sample. Analyses of the imputed data highlighted two significantly associated variants in *CREBRF* (encoding CREB3 regulatory factor), rs150207780 and rs373863828 (**Fig. 1b**). Because of high linkage disequilibrium (LD) in the region, conditional analyses were not able to distinguish between the top variants on statistical grounds (**Supplementary Fig. 3**). Annotation indicated that neither rs12513649, located between *ATP6V0E1* and *CREBRF*, nor rs150207780, located in intron 1 of *CREBRF*, had any predicted regulatory function, drawing our attention to rs373863828, which was

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [2]Department of Epidemiology (Chronic Disease), Yale University School of Public Health, New Haven, Connecticut, USA. [3]Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA. [4]Division of Endocrinology, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [5]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [6]Bureau of Statistics, Government of Samoa, Apia, Samoa. [7]Samoa National Health Service, Apia, Samoa. [8]Department of Health, American Samoa Government, Pago Pago, American Samoa, USA. [9]Ministry of Health, Government of Samoa, Apia, Samoa. [10]International Health Institute, Department of Epidemiology, Brown University School of Public Health, Providence, Rhode Island, USA. [11]Department of Anthropology, Brown University, Providence, Rhode Island, USA. [12]Present addresses: Department of Internal Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yun-Lin, Taiwan (C.-T.S.) and Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, US National Institutes of Health, Bethesda, Maryland, USA (O.D.B.). [13]These authors contributed equally to this work. [14]These authors jointly supervised this work. Correspondence should be addressed to S.T.M. (stephen_mcgarvey@brown.edu).

Received 7 December 2015; accepted 15 June 2016; published online 25 July 2016; doi:10.1038/ng.3620

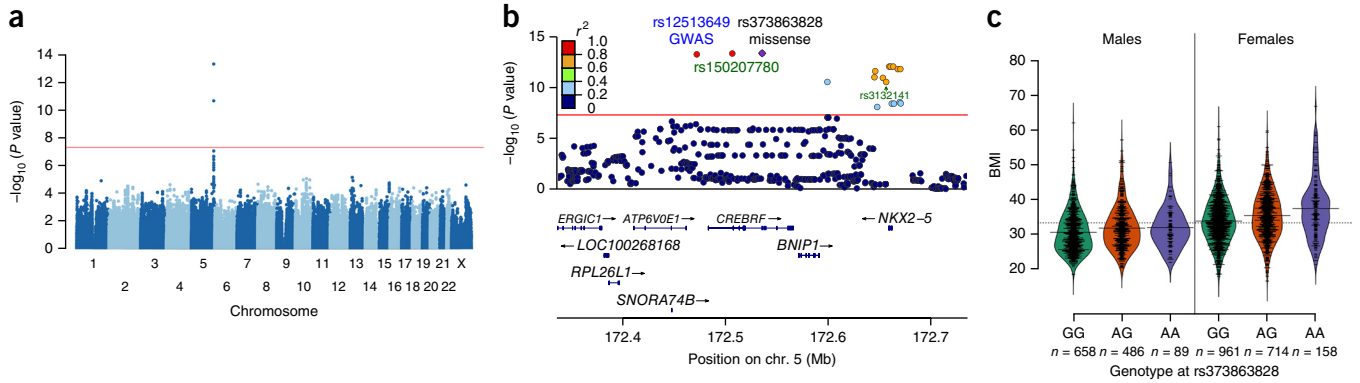

Figure 1 Association results from genome-wide and targeted sequencing and beanplots of BMI versus genotype in men and women from the discovery sample. (**a**) Manhattan plot of the genome-wide association scan for association with BMI. The red line corresponds to a *P* value of 5×10^{-8}. (**b**) Association results using imputed data for the region encompassing *CREBRF*. The strength of LD, as measured by the squared correlation of genotype dosages, between each variant and the missense variant rs373863828 is represented by the color of each point. The red line corresponds to a *P* value of 5×10^{-8}. The plot was generated using LocusZoom[32]. (**c**) Beanplots of BMI versus genotype in men ($n = 1{,}233$) and women ($n = 1{,}833$) from the discovery sample. Each bean consists of a mirrored density curve containing a one-dimensional scatterplot of the individual data. A solid line shows the average for each group, and the dashed line represents the overall average. The plot was generated using the R beanplot package[33].

the only strongly associated missense variant among the 775 variants with $P \leq 1 \times 10^{-5}$ in the targeted sequencing region. The rs373863828 missense variant (c.1370G>A, p.Arg457Gln) is located at a highly conserved position (GERP score 5.49) with a high probability of being damaging (SIFT, 0.03; PolyPhen-2, 0.996). The BMI-increasing A allele of rs373863828 has an overall frequency of 0.259 in Samoans but is unobserved or extremely rare in other populations, with an allele count in the Exome Aggregation Consortium of only 5 among 121,362 measured alleles (**Table 1**)[12]. Bayesian fine-mapping with PAINTOR[13] strongly supported following up the missense variant. The two variants in the region with the highest posterior probability (PP) of being causal were rs373863828 (PP = 0.80) and rs150207780 (PP = 0.22); when Encyclopedia of DNA Elements (ENCODE) functional annotation was included, these probabilities increased to 0.92 and 0.34, respectively.

We then genotyped the missense variant rs373863828 in the discovery and replication samples, obtaining very significant evidence of association with BMI in adults ($P = 7.0 \times 10^{-13}$ and $P = 3.5 \times 10^{-9}$, respectively), with a combined meta-analysis *P* value of 1.4×10^{-20} (**Table 1**). The meta-analysis showed no evidence of heterogeneity ($I^2 = 0\%$; $Q = 1.12$; $P = 0.571$). In our discovery sample, each copy of the A allele increased BMI by 1.36 kg/m^2 (**Fig. 1c**). In our adult replication sample, each copy of the A allele increased BMI by 1.45 kg/m^2. There was a strong effect on BMI at this locus even after stratifying by sex and cohort (**Supplementary Fig. 4**; however, sex–genotype interactions were not significant (discovery $P = 0.060$; replication $P = 0.555$)). There was also suggestive evidence ($P = 1.1 \times 10^{-3}$) that this variant increased BMI in our sample of 409 Samoan children (**Table 1**). The rs373863828 variant (encoding p.Arg457Gln) accounted for 1.93% of the variance in BMI in our discovery sample and 1.08% of the variance in BMI in our replication sample. In comparison, rs1558902, the main risk-associated variant in *FTO*, increases BMI by 0.39 kg/m^2 per copy of the risk-associated allele and accounts for only 0.34% of the variance in BMI in Europeans[14,15]. In searches of the literature and databases (including GRASP[16,17]), we were unable to identify any significant associations with BMI in the *CREBRF* region in other human studies.

In addition to BMI, the A allele of rs373863828 was also positively associated with obesity risk (odds ratio (OR) = 1.305 and 1.441 in the discovery and replication cohorts, respectively) as well as measures of total and regional adiposity, including percent body fat, abdominal circumference, and hip circumference, in both cohorts (**Table 2** and **Supplementary Table 2**). The A allele was also positively associated with serum leptin levels in women (both cohorts) and men (replication cohort) before but not after adjusting for BMI. These data indicate that the association between the missense variant and BMI is indeed due to an association with adiposity.

Higher BMI and adiposity are usually associated with greater insulin resistance (higher fasting insulin levels and homeostatic model assessment–insulin resistance (HOMA-IR)), an atherogenic lipid profile (especially higher serum triglyceride and lower HDL cholesterol levels), and lower adiponectin levels. We therefore expected the BMI-increasing A allele of rs373863828 to also be associated with these metabolic variables. However, even though the A allele was consistently associated with higher BMI and adiposity in both the discovery and replication cohorts, the expected associations with the above obesity-related comorbidities were not observed and, in some cases, were even in the opposite direction to that expected (**Table 2** and **Supplementary Table 2**). Notably, when considering all subjects, the risk of diabetes was actually lower (OR = 0.586 for the discovery cohort, $P = 6.68 \times 10^{-9}$) or trended lower (0.742 for the replication cohorts, $P = 0.029$) in carriers of the A allele. Likewise, even in non-diabetic subjects, the variant was associated with moderately but significantly lower fasting glucose levels in both the discovery and replication cohorts (1.65 mg/dl ($P = 9.5 \times 10^{-5}$) and 1.54 mg/dl ($P = 8.8 \times 10^{-4}$) lower for each copy of the A allele, respectively). These effects became even more significant after adjusting for BMI (2.25 mg/dl, $P = 6.9 \times 10^{-8}$ and 2.09 mg/dl, $P = 7.6 \times 10^{-6}$), suggesting an independent effect of the variant on glucose homeostasis and diabetes risk. Such effects are unlikely to be due to survival bias, as no correlation between age and genotype was observed (linear regression $P = 0.849$). These effects seem to be independent of obesity-associated insulin resistance, as associations with fasting insulin levels and HOMA-IR were not consistently observed across the cohorts (associations were stronger only in the replication cohort before adjusting for BMI). Furthermore, although the variant was associated with lower total cholesterol levels in the discovery cohort, consistent effects on serum lipid or adiponectin levels were likewise not observed. Together, these data suggest that the missense variant does not promote, and may even protect against, obesity-associated comorbidities; however, additional studies will be required to confirm these findings and directly test this hypothesis.

Table 1 Association details for rs12513649 and rs373863828

	Discovery variant	Missense variant
SNP rs ID	rs12513649	rs373863828
Chromosome	5	5
Physical position (GRCh37.p13) (bp)	172,472,052	172,535,774
Effect allele	G	A
Other allele	C	G
Nearest gene upstream of the SNP	*ATP6V0E1*	*CREBRF*
Distance to nearest upstream gene (bp)	10,152	0
Nearest gene downstream of the SNP	*CREBRF*	*CREBRF*
Distance to nearest downstream gene (bp)	11,302	0
Sample sizes (phenotyped and genotyped)		
GWAS Samoans from the 2010s (discovery)	3,072	3,066
Samoans from the 1990s (replication)	1,020	1,020
Samoans from the 2000s (replication)	1,082	1,083
Meta-analysis of the 1990s and 2000s samples	2,102	2,103
Meta-analysis of the 1990s, 2000s, and 2010s samples	5,174	5,169
Samoan children from the 2000s	409	409
P values for log-transformed BMI		
GWAS Samoans from the 2010s (discovery)	5.3×10^{-14}	7.0×10^{-13}
Samoans from the 1990s (replication)	5.8×10^{-4}	8.0×10^{-4}
Samoans from the 2000s (replication)	3.0×10^{-7}	6.5×10^{-7}
Meta-analysis of the 1990s and 2000s samples	1.2×10^{-9}	3.5×10^{-9}
Meta-analysis of the 1990s, 2000s, and 2010s samples	4.0×10^{-22}	1.4×10^{-20}
Samoan children from the 2000s	4.1×10^{-3}	1.1×10^{-3}
Effect sizes (β (s.e.)) for log-transformed BMI		
GWAS Samoans from the 2010s (discovery)	0.041 (0.005)	0.039 (0.005)
Samoans from the 1990s (replication)	0.029 (0.008)	0.028 (0.008)
Samoans from the 2000s (replication)	0.056 (0.011)	0.054 (0.011)
Samoan children from the 2000s	0.031 (0.011)	0.035 (0.011)
Effect allele frequencies		
GWAS Samoans from the 2010s	0.276	0.276
Samoans from the 1990s	0.251	0.251
Samoan adults from the 2000s	0.224	0.225
Samoan children from the 2000s	0.236	0.235
All of the 1990s, 2000s and 2010s samples	0.258	0.259
Individuals of East Asian descent from 1000G	0.063	0.000
Individuals of South Asian descent from 1000G	0.003	0.000
Individuals of European descent from 1000G	0.000	0.000
Individuals of admixed American descent from 1000G	0.059	0.000
Individuals of African descent from 1000G	0.001	0.000
Individuals of East Asian descent from ExAC	NA	<0.001[a]
Individuals of South Asian descent from ExAC	NA	0.000
Individuals of European descent from ExAC	NA	<0.001[b]
Individuals of Latino descent from ExAC	NA	0.000
Individuals of African descent from ExAC	NA	0.000
Individuals of other descent from ExAC	NA	0.001[c]

This table provides detailed results for rs12513649 and rs373863828. 1000G, 1000 Genomes Project; ExAC, Exome Aggregation Consortium[12]; s.e., standard error; NA, not available.
[a]Two A alleles in 8,636 measured alleles. [b]Two A alleles in 73,328 measured alleles. [c]One A allele in 908 measured alleles.

Although the majority of genes contributing to obesity do so by influencing the central regulation of energy balance[18], emerging evidence highlights the contribution of altered cellular metabolism to obesity[19]. Therefore, we examined the impact of rs373863828 on cellular bioenergetics. To do so, we selected the established 3T3-L1 mouse adipocyte model for two reasons: (i) *CREBRF* is widely expressed in virtually all tissues, including adipose tissue (**Supplementary Fig. 5**), suggesting a fundamental cellular function, and (ii) several CREB family proteins have been linked to mitochondrial function and metabolic phenotypes in adipocytes[20–23]. Thus, this model is well suited to assess multiple potentially relevant metabolic phenotypes.

We first characterized the effects of adipogenic differentiation and ectopic overexpression of human wild-type or Arg457Gln CREBRF on endogenous *Crebrf* expression in 3T3-L1 cells. *Crebrf* expression was induced during adipogenesis in conjunction with that of adipogenic markers (*Cebpa*, *Pparg*, and *Adipoq*), suggesting a role for CREBRF in this process (**Supplementary Fig. 6**). Indeed, comparable stable overexpression of the transcripts for human wild-type and Arg457Gln CREBRF (**Fig. 2a**), without changing endogenous *Crebrf* levels (**Fig. 2b**), was sufficient to induce the expression of adipogenic markers (**Fig. 2c–e**) and promote lipid and triglyceride accumulation (**Fig. 2f–h**) in the absence of standard hormonal induction of adipogenesis. Although Arg457Gln CREBRF resulted in slightly weaker induction of adipogenic markers than wild-type protein (**Fig. 2c**,**e**), it promoted significantly ($P < 0.02$) greater lipid and triglyceride accumulation (**Fig. 2f–h**). To determine whether this increased energy storage was associated with decreased energy use, we next assessed glycolysis, mitochondrial respiration, and ATP production. Consistent with published data[24,25], glycolysis was suppressed and mitochondrial respiration and ATP production were enhanced by hormonally induced adipogenic differentiation (**Supplementary Fig. 7**). Stable overexpression of wild-type CREBRF increased whereas Arg457Gln CREBRF decreased multiple measures of cellular energy use, including basal and maximal mitochondrial respiration, mitochondrial ATP production, and basal glycolysis (**Fig. 2i**). These data indicate that the Arg457Gln CREBRF variant promotes more lipid storage while using less energy than wild-type CREBRF.

In addition to having a role in cellular energy storage and use, the *Drosophila melanogaster CREBRF* ortholog *REPTOR* has recently been implicated in both cellular and organismal adaptation to nutritional stress by mediating the downstream transcriptional response to the cellular energy sensor TORC1 (refs. 26,27). In support of this hypothesis, expression of *CREBRF* orthologs is highly induced by starvation in all tissues of *Drosophila*[26,27] as well as in human lymphoblasts[28,29]. Moreover, *REPTOR*-knockout flies[26] and *Crebrf*-knockout mice[30] have lower total energy storage and body weight, respectively. Similarly, we found that nutrient starvation of 3T3-L1 preadipocytes rapidly increased *Crebrf* mRNA levels, which peaked by 4 h at levels 13-fold higher than those seen at 0 h ($P = 1.1 \times 10^{-16}$) and remained elevated by 5-fold at 24 h after the start of starvation ($P = 4.1 \times 10^{-14}$) (**Fig. 3a**). Treatment with rapamycin, a TORC1 inhibitor, also rapidly increased *Crebrf* mRNA levels, but did so to a lesser extent than starvation (**Fig. 3b**), indicating that additional TORC1-independent signals converge on *Crebrf*. Furthermore, overexpression of wild-type and

Table 2 Association of rs373863828 with untransformed adiposity, metabolic, and lipid traits in the discovery sample

Quantitative trait	*n*	β (s.e.)	*P*	Covariates[a]
Adiposity traits				
BMI (kg/m^2)	3,066	1.356 (0.183)	**1.12×10^{-13}**	A, A^2, S, A × S
Body fat (%)	2,893	2.199 (0.345)	**1.78×10^{-10}**	A, A^2, S, A × S
Abdominal circumference (cm)	3,057	2.842 (0.404)	**2.05×10^{-12}**	A, A^2, S, A × S, A^2 × S
Hip circumference (cm)	3,058	2.361 (0.332)	**1.19×10^{-12}**	A, A^2, S, A^2 × S
Abdominal–hip ratio	3,056	0.005 (0.002)	2.23×10^{-3}	A, A^2, S, A × S, A^2 × S
Metabolic traits				
Fasting glucose (mg/dl)[b]	2,393	−1.652 (0.423)	**9.52×10^{-5}**	A, A^2, S
Fasting insulin (μU/ml)[b]	2,392	1.342 (0.449)	2.83×10^{-3}	A, S, A × S
HOMA-IR[b]	2,392	0.241 (0.114)	0.035	A, S, A × S
Adiponectin (μg/ml)	2,858	−0.228 (0.083)	0.006	A, A^2, S, A × S
Leptin in men (ng/ml)[c]	1,151	0.719 (0.326)	0.027	A
Leptin in women (ng/ml)[c]	1,707	1.888 (0.525)	**3.25×10^{-4}**	
Metabolic traits adjusted for BMI				
Fasting glucose (mg/dl)[b]	2,383	−2.248 (0.417)	**6.89×10^{-8}**	A, A^2, S, B
Fasting insulin (μU/ml)[b]	2,382	0.225 (0.420)	0.592	A, A^2, S, B, A × S, A^2 × S
HOMA-IR[b]	2,382	−0.034 (0.107)	0.754	A, B
Adiponectin (μg/ml)	2,844	−0.066 (0.080)	0.412	A, A^2, S, B, A × S
Leptin in men (ng/ml)[c]	1,143	−0.262 (0.210)	0.213	A, A^2, B
Leptin in women (ng/ml)[c]	1,701	−0.516 (0.366)	0.159	A, A^2, B
Serum lipid levels				
Total cholesterol (mg/dl)	2,858	−3.203 (1.029)	**1.84×10^{-3}**	A, A^2, S, A × S, A^2 × S
Triglycerides (mg/dl)	2,858	0.349 (2.769)	0.900	A, S, A × S
HDL cholesterol (mg/dl)	2,858	−0.322 (0.321)	0.317	A, A^2, S
LDL cholesterol (mg/dl)	2,851	−2.347 (0.945)	0.013	A, A^2, S, A^2 × S
Dichotomous traits	***n***	**OR (95% CI)**	***P***	**Covariates[a]**
Obesity (>32 kg/m^2)	3,066	1.305 (1.159–1.470)	**1.12×10^{-5}**	A, A^2, S, A × S
Diabetes	2,876	0.637 (0.536–0.758)	**3.86×10^{-7}**	A
Diabetes adjusted for BMI	2,861	0.586 (0.489–0.702)	**6.68×10^{-9}**	A, B
Hypertension	3,041	1.014 (0.898–1.145)	0.818	A, S

Boldface represents a *P* value <2.17×10^{-3}. s.e., standard error; OR, odds ratio; 95% CI, 95% confidence interval. [a]A, age; A^2, age^2; S, sex; A × S, age × sex interaction; A^2 × S = age^2 × sex interaction, B, log(BMI). [b]Analysis was conducted only in non-diabetics. [c]Leptin was not analyzed in men and women together because the distributions were very different for the sexes.

Arg457Gln human CREBRF equivalently reduced the cell death rate to approximately one-third of that in controls within the first 6 h of nutrient starvation in 3T3-L1 preadipocytes ($P = 5 \times 10^{-6}$ and $P = 4 \times 10^{-5}$, respectively; **Fig. 3c**,**d**). These data indicate that CREBRF is a starvation-responsive factor and that wild-type and Arg457Gln CREBRF when overexpressed confer similar protection against cellular nutritional stress.

Complementing the functional evidence of 'thriftiness', we identified evidence of positive selection at the missense variant in Samoan genomes. The core haplotype carrying the derived BMI-increasing allele exhibited long-range LD (corresponding to the single thick branch in **Fig. 4b** versus **Fig. 4a**) and had elevated extended haplotype homozygosity (EHH) relative to haplotypes carrying the ancestral allele (**Fig. 4c**). Haplotypes carrying the derived allele were longer than haplotypes carrying the ancestral allele (**Fig. 4d**). Evidence of positive selection was provided by an integrated haplotype score (iHS) of 2.94 ($P \approx 0.003$) and a number of segregation sites by length (nS_L) score of 2.63 ($P \approx 0.008$) (**Supplementary Fig. 8**).

In 1962, James Neel posited the existence of a thrifty gene that provides a metabolic advantage in times of famine but promotes metabolic disease in times of nutritional excess[31]. By carrying out a genome-wide association analysis of BMI in Samoans, we discovered and replicated a strong association with a missense variant in *CREBRF* that has a much larger effect size than any other known common risk-associated variant for BMI[18]. Functional evidence from an adipocyte model further demonstrated that CREBRF with this missense variant promotes cellular energy conservation by increasing fat storage and decreasing energy use in comparison to the wild-type protein.

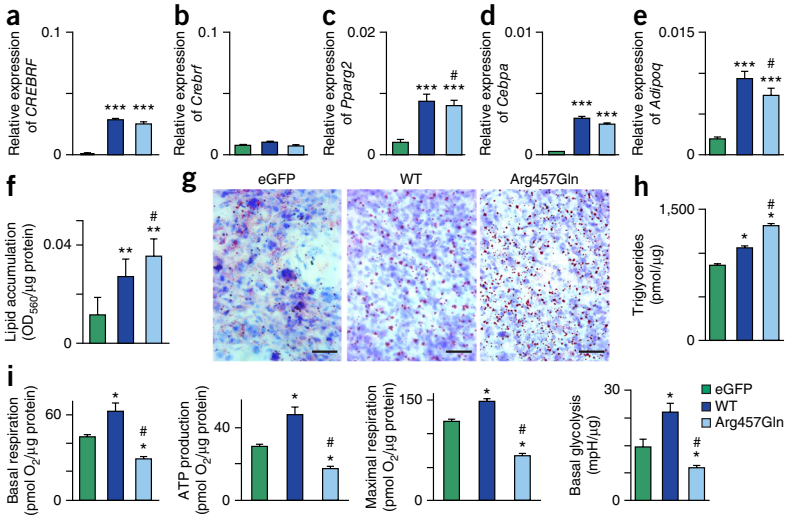

Figure 2 CREBRF variants, adipogenic differentiation, lipid accumulation, and energy homeostasis. 3T3-L1 mouse preadipocytes overexpressing enhanced GFP–only negative control (eGFP), wild-type human CREBRF (WT), or Arg457Gln human CREBRF were collected at 8 d after confluence in the absence of hormonal stimulation of adipogenic differentiation. (**a**–**e**) mRNA levels of human *CREBRF* (**a**) and endogenous mouse *Crebrf* (**b**), *Pparg2* (**c**), *Cebpa* (**d**), and *Adipoq* (**e**) relative to those of the β-actin (*Actb*) reference gene determined using quantitative RT–PCR. Values are given as means ± s.e.m. from three biological replicates with four technical replicates each ($n = 3 \times 4 = 12$). Representative results from one of four experiments are shown. (**f**) Quantification of lipid accumulation with Oil Red O staining normalized to protein content (OD_{560}/μg protein). Data are shown as means ± s.e.m. from three transfection replicates with eight wells for each transfection ($n = 3 \times 8 = 24$). (**g**) Representative photomicrographs of Oil Red O staining to visualize lipid droplets (red) with counterstaining of nuclei with hematoxylin (blue). Scale bars, 50 μm. (**h**) Biochemical assay for triglycerides. Data are shown as means ± s.e.m., $n = 2$ biological replicates. (**i**) Key bioenergenic variables as determined on the basis of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) normalized to protein content. Values are given as means ± s.e.m. ($n = 6$ biological replicates). mpH, 0.01 pH unit. Statistical analysis: one-way analysis of variance (ANOVA), two-sided Games–Howell *post-hoc* test. $^*P < 0.03$, $^{**}P < 1 \times 10^{-3}$, $^{***}P < 1 \times 10^{-4}$ compared to 3T3-L1 cells transfected with eGFP control construct; $^{\#}P < 0.05$ compared to 3T3-L1 cells transfected with construct for wild-type CREBRF.

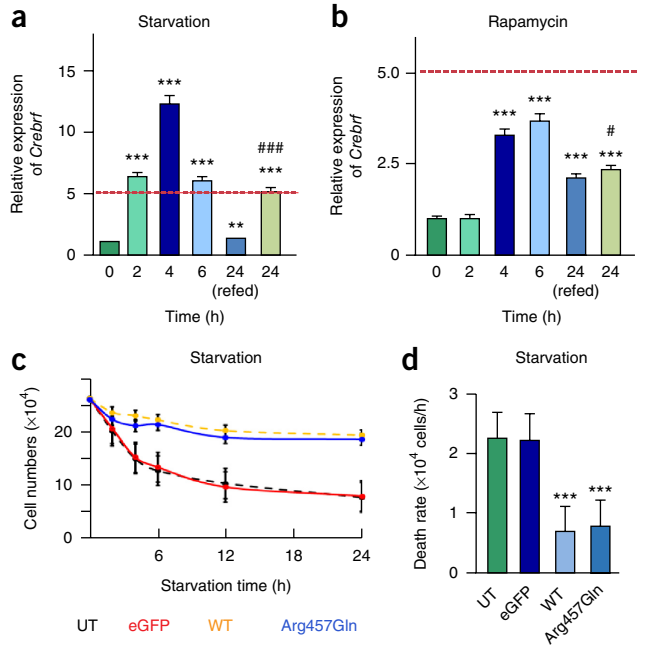

Figure 3 Induction of *Crebrf* expression by nutritional stress and protection against starvation. (**a**,**b**) 3T3-L1 preadipocytes were starved (**a**) or treated with 20 ng/ml rapamycin (**b**) for 0, 2, 4, 12, or 24 h. A set of cells was starved or treated with rapamycin for 12 h and then refed with fresh growth medium for an additional 12 h (24 h (refed)). *Crebrf* mRNA levels were determined relative to *Actb* levels and normalized to baseline levels (0 h). Values are given as means ± s.e.m. from three biological replicates with four technical replicates each ($n = 3 \times 4 = 12$). Statistical analysis: one-way ANOVA and two-sided Bonferroni *post-hoc* tests. $^{**}P = 0.002$, $^{***}P < 1 \times 10^{-11}$ compared to cells at 0 h; $^{\#}P = 0.02$, $^{\#\#\#}P = 8.8 \times 10^{-13}$ compared to cells at 24 h (refed). (**c**,**d**) 3T3-L1 preadipocytes were either untransfected (UT) or transfected with plasmid encoding eGFP-only negative control, wild-type human CREBRF, or Arg457Gln CREBRF and starved. (**c**) Time course of 3T3-L1 cell survival upon starvation up to 24 h. (**d**) Cell death rates after 0–6 h of starvation. Values are given as means ± s.e.m. from two transfection replicates with six wells for each transfection and three technical (cell counting) replicates ($n = 2 \times 6 \times 3 = 36$). This experiment was performed once following a pilot experiment with fewer time points showing similar results. Statistical analysis: one-way ANOVA and two-sided Games–Howell *post-hoc* tests. $^{***}P < 5 \times 10^{-5}$ compared to cells transfected with control eGFP construct.

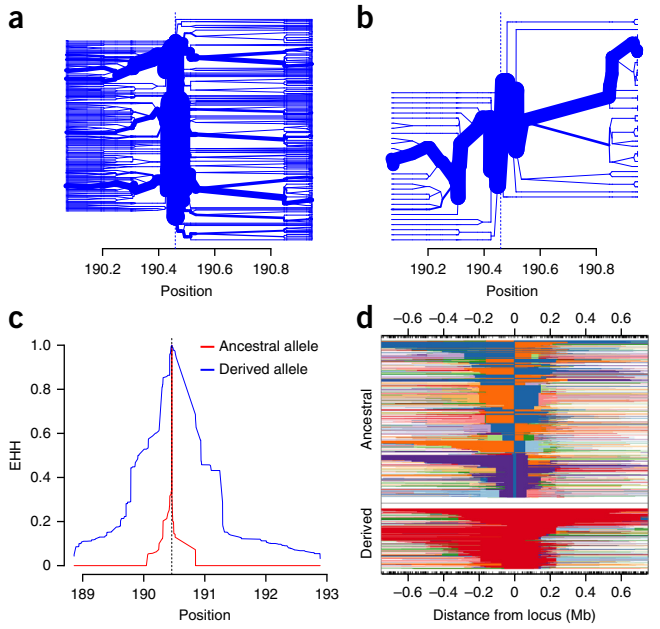

Figure 4 Evidence of positive selection centered on the missense variant rs373863828. Findings are shown for 626 Samoans who are not closely related. (**a**,**b**) Haplotype bifurcation plots for haplotypes carrying the ancestral allele (**a**) and the derived allele (**b**) at rs373863828 show that haplotypes carrying the derived allele have unusual long-range homozygosity. (**c**,**d**) Haplotypes carrying the derived allele have elevated EHH values as one moves away from rs373863828 (vertical dashed line) (**c**) and are longer than those carrying the ancestral allele (**d**).

The potential importance of this variant in organismal energy homeostasis is further supported by the 'lean' phenotype of mice[30] and flies[26] lacking the ortholog for this gene. These data, in combination with evidence of positive selection, support a thrifty variant hypothesis for human obesity and underscore the value of examining unique populations to identify new genetic contributions to complex traits.

However, many questions remain unanswered. More detailed studies in animal models and humans are required to define the systemic and tissue-specific (particularly central) contributions of the missense variant to overall energy balance. Such studies would also help confirm and clarify the mechanism by which this missense variant might protect against obesity-associated metabolic disease, which perhaps involves preferential promotion of more metabolically 'safe' or efficient energy storage and use. Studies that consider potential modifying and mediating environmental influences of this variant as well as gene–gene interactions might illuminate additional new factors contributing to these complex traits. Finally, additional anthropological genetic studies might determine the evolutionary origin of this variant or the potential role of drift in determining its frequency. Such research is urgently needed to inform decisions about how to use knowledge of this obesity risk variant to benefit Samoans at both individual and population health levels and to determine how this discovery might contribute to the understanding and treatment of more common obesity in general.

URLs. BGTEx portal, http://www.gtexportal.org/; BioGPS portal, http://biogps.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The discovery data set is available from the database of Genotypes and Phenotypes (dbGaP) under accession phs000914.v1.p1.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank the Samoan participants of the study, and local village authorities and the many Samoan and other field workers over the years. We acknowledge the Samoan Ministry of Health and the Samoan Bureau of Statistics, and the American Samoan Department of Health for their support of this research. We also acknowledge S.S. Shiva and C.G. Corey at the University of Pittsburgh Center for Metabolism and Mitochondrial Biology for assistance with cellular bioenergetic profiling. This work was funded by NIH grants R01-HL093093 (S.T.M.), R01-AG09375 (S.T.M.), R01-HL52611 (I. Kamboh), R01-DK59642 (S.T.M.), P30 ES006096 (S.M. Ho), R01-DK55406. (R.D.), R01-HL090648 (Z.U.), and R01-DK090166 (E.E.K.) and by Brown University student research funds. Genotyping was performed in the Core Genotyping Laboratory at the University of Cincinnati, funded by NIH grant

P30 ES006096 (S.M. Ho). Illumina sequencing was conducted at the Genetic Resources Core Facility, Johns Hopkins Institute of Genetic Medicine (Baltimore).

AUTHOR CONTRIBUTIONS

R.L.M. performed the genotype quality control and association analyses, with guidance from D.E.W. and assistance from O.D.B. and J.L.; D.E.W. and R.L.M. wrote the relevant sections of the manuscript. N.L.H. led the field work data collection and phenotype analyses with guidance from S.T.M. G.S. led and directed genotyping experiments (using the Affymetrix 6.0 chip) and assay development for validation and replication (using the TaqMan platform) with guidance from R.D. H.C. participated extensively in DNA extraction, genotyping, and quality control of the data under the supervision of G.S. and R.D. Z.U. and C.-T.S. designed and performed the *CREBRF* overexpression, lipid accumulation, and adipocyte differentiation and starvation experiments, analyzed the data, and wrote the relevant sections of the manuscript. E.E.K. contributed mouse and human gene expression profiling data as well as contributed to the design and analysis of the functional studies. M.S.R., S.V., and J.T. facilitated fieldwork in Samoa and American Samoa. T.N. contributed to the discussion of the public health implications of the findings. All authors contributed to this work, discussed the results, and critically reviewed and revised the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Åberg, K. *et al.* Susceptibility loci for adiposity phenotypes on 8p, 9p, and 16q in American Samoa and Samoa. *Obesity (Silver Spring)* **17**, 518–524 (2009).
2. McGarvey, S.T. Obesity in Samoans and a perspective on its etiology in Polynesians. *Am. J. Clin. Nutr.* **53** (Suppl. 6), 1586S–1594S (1991).
3. Hawley, N.L. *et al.* Prevalence of adiposity and associated cardiometabolic risk factors in the Samoan genome-wide association study. *Am. J. Hum. Biol.* **26**, 491–501 (2014).
4. Tishkoff, S. Strength in small numbers. *Science* **349**, 1282–1283 (2015).
5. McGarvey, S.T., Bindon, J.R., Crews, D.E. & Schendel, D.E. in *Human Population Biology: A Transdisciplinary Science* (eds. Little, M.A. & Haas, J.D.) 263–279 (Academic Press, 1989).
6. McGarvey, S.T. The thrifty gene concept and adiposity studies in biological anthropology. *J. Polyn. Soc.* **103**, 29–42 (1994).
7. Zimmet, P., Dowse, G., Finch, C., Serjeantson, S. & King, H. The epidemiology and natural history of NIDDM—lessons from the South Pacific. *Diabetes Metab. Rev.* **6**, 91–124 (1990).
8. Kirch, P.V. & Rallu, J.-L. in *The Growth and Collapse of Pacific Island Societies* (eds. Kirch, P.V. & Rallu, J.-L.) 1–14 (University of Hawaii Press, 2007).
9. Friedlaender, J.S. *et al.* The genetic structure of Pacific Islanders. *PLoS Genet.* **4**, e19 (2008).
10. Tsai, H.-J. *et al.* Distribution of genome-wide linkage disequilibrium based on microsatellite loci in the Samoan population. *Hum. Genomics* **1**, 327–334 (2004).
11. Green, R.C. in *The Growth and Collapse of Pacific Island Societies* (eds. Kirch, P.V. & Rallu, J.-L.) 203–231 (University of Hawaii Press, 2007).
12. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Preprint at *bioRxiv* http://dx.doi.org/10.1101/030338 (2016).
13. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
14. Loos, R.J. & Yeo, G.S. The bigger picture of *FTO*: the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.* **10**, 51–61 (2014).
15. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
16. Eicher, J.D. *et al.* GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **43**, D799–D804 (2015).
17. Leslie, R., O'Donnell, C.J. & Johnson, A.D. GRASP: analysis of genotype–phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* **30**, i185–i194 (2014).
18. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
19. Pearce, L.R. *et al. KSR2* mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765–777 (2013).
20. Vankoningsloo, S. *et al.* CREB activation induced by mitochondrial dysfunction triggers triglyceride accumulation in 3T3-L1 preadipocytes. *J. Cell Sci.* **119**, 1266–1282 (2006).
21. Reusch, J.E., Colton, L.A. & Klemm, D.J. CREB activation induces adipogenesis in 3T3-L1 cells. *Mol. Cell. Biol.* **20**, 1008–1020 (2000).
22. Ma, X. *et al.* CREBL2, interacting with CREB, induces adipogenesis in 3T3-L1 adipocytes. *Biochem. J.* **439**, 27–38 (2011).
23. Kim, T.H. *et al.* Identification of Creb3l4 as an essential negative regulator of adipogenesis. *Cell Death Dis.* **5**, e1527 (2014).
24. Wilson-Fritch, L. *et al.* Mitochondrial biogenesis and remodeling during adipogenesis and in response to the insulin sensitizer rosiglitazone. *Mol. Cell. Biol.* **23**, 1085–1094 (2003).
25. Keuper, M. *et al.* Spare mitochondrial respiratory capacity permits human adipocytes to maintain ATP homeostasis under hypoglycemic conditions. *FASEB J.* **28**, 761–770 (2014).
26. Tiebe, M. *et al.* REPTOR and REPTOR-BP regulate organismal metabolism and transcription downstream of TORC1. *Dev. Cell* **33**, 272–284 (2015).
27. Stocker, H. Stress relief downstream of TOR. *Dev. Cell* **33**, 245–246 (2015).
28. Chen, R., Mallelwar, R., Thosar, A., Venkatasubrahmanyam, S. & Butte, A.J. GeneChaser: identifying all biological and clinical conditions in which genes of interest are differentially expressed. *BMC Bioinformatics* **9**, 548 (2008).
29. Dengjel, J. *et al.* Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc. Natl. Acad. Sci. USA* **102**, 7922–7927 (2005).
30. Martyn, A.C. *et al.* Luman/CREB3 recruitment factor regulates glucocorticoid receptor activity and is essential for prolactin-mediated maternal instinct. *Mol. Cell. Biol.* **32**, 5140–5150 (2012).
31. Neel, J.V. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am. J. Hum. Genet.* **14**, 353–362 (1962).
32. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
33. Kampstra, P. Beanplot: a boxplot alternative for visual comparison of distributions. *J. Stat. Softw.* **28**, 1–9 (2008).

ONLINE METHODS

Participants. The participants in this study are derived from the populations of the Independent State of Samoa and the US territory of American Samoa. We used two samples in this study: a discovery sample of 3,072 phenotyped and genotyped Samoans and a replication sample of 2,103 phenotyped and genotyped Samoans and American Samoans (**Supplementary Table 1**). An additional sample of 409 phenotyped and genotyped Samoan children was not included in the main analyses, but analyses with our associated variants were also conducted in this sample. Details about participant recruitment can be found in the **Supplementary Note**. The parent GWAS, sample selection and data collection methods, and phenotype levels, including those of lipids and lipoproteins, have been reported[3]. This study has been approved by the Health Research Committee of the Samoa Ministry of Health and the institutional review boards of Brown University, the University of Cincinnati, and the University of Pittsburgh. All participants gave informed consent.

In the original GWAS study design, our goal of a discovery sample size of 2,500 (which we exceeded) was chosen so as to have high power to detect risk-associated SNPs with realistic effect sizes. Power was estimated as follows: we used Quanto[34,35] to estimate the power to detect the rs9930506 SNP in *FTO*, which in the Sardinia study[36] explained 1.34% of variance in BMI. If we assume that this SNP has the same allele frequencies and that BMI has the same overall mean values and standard deviation as in Scuteri *et al.*[36], then at a significance level of 1×10^{-5} power is ≥80% when the risk-associated SNP explains at least 1.1% of the variance (and power is 90% when the SNP explains 1.3% of the variance). If we instead test at a threshold of 1×10^{-7}, power is ≥80% if the SNP explains at least 1.5% of the variance.

Anthropometric and biochemical measurements. Height, weight, and BMI were measured as previously described[3,37,38]. Polynesian cutoffs were used to classify adults as normal weight, overweight, or obese on the basis of BMI of <26 kg/m^2, 26–32 kg/m^2, and >32 kg/m^2, respectively[39]. Obesity in children was categorized from BMI using the international age- and sex-specific classifications developed by Cole *et al.*[40].

In the discovery sample, abdominal (at the level of the umbilicus) and hip circumferences were measured in duplicate, and the measures were averaged (**Supplementary Table 1**). Bioelectrical impedance measures of resistance and reactance (RJL BIA-101Q device, RJL Systems) were used to estimate percent body fat on the basis of Polynesian-specific equations[38,39]. Serum separated from whole-blood samples, collected after a 10-h overnight fast, was assayed for cholesterol (total, HDL, and LDL), triglycerides, glucose, and insulin. The assay techniques for these metabolic markers have been described previously[1]. Individuals were classified as having type 2 diabetes on the basis of fasting serum glucose levels ≥126 mg/dl or the current use of diabetes medication[41]. Hypertensives either had systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg, or were currently taking hypertension medication. Additionally, serum levels of leptin and adiponectin were obtained by using commercially available radioimmunoassay kits (EMD Millipore). HOMA-IR was calculated as glucose (mg/dl) × insulin (μU/ml)/405, as recommended[42].

Genotyping. Genotyping of the discovery sample was performed using Genome-Wide Human SNP 6.0 arrays (Affymetrix). Extensive quality control was conducted on the basis of a pipeline developed by Laurie *et al.*[43]. Additional details for sample genotyping and genotype quality control can be found in the **Supplementary Note**.

Statistical analysis. During quality control, significant relatedness was observed among the discovery sample participants, so empirical kinship coefficients were estimated using genotyped markers, in two iterations. In the first iteration, we selected 10,000 independent autosomal markers using PLINK[44] and used them to generate empirical kinship coefficients with GenABEL[45]. Individuals with kinship coefficients less than 0.0625 (corresponding to first cousins) were considered unrelated. A maximal set of 1,891 unrelated individuals was then determined using previously published methods[46]. In the second iteration, the kinship matrix for all participants was estimated using a new set of 10,000 independent autosomal markers that had been selected using the set of unrelated individuals.

We tested for association between autosomal marker genotypes and BMI residuals while using the empirical kinship matrix to adjust for population substructure and subject relatedness. The tests were conducted using a score test as implemented in the mmscore function in GenABEL[47]. The statistics for association of X-chromosome genotypes with BMI residuals were calculated in GenABEL without adjusting for the empirical kinship estimates.

Meta-analysis of the adult samples was performed using METAL[48] to generate two replication *P* values: one for the adult replication samples and one for the adult replication samples and the discovery sample together (**Table 1**). Additional details of the statistical analyses, including ancestry principal components (**Supplementary Fig. 1** and **Supplementary Video 1**), can be found in the **Supplementary Note**.

Targeted sequencing. Before undertaking targeted sequencing, we first used SHAPEIT[49–53] and IMPUTE2 (refs. 54–56) for imputation in our region of interest centered on rs12513649 with the December 2013 1000 Genomes Project Phase I integrated variant set release haplotype reference panel. The approach implicated only one strongly associated variant (with a predicted allele frequency of 0.075), but when we genotyped this variant in a pilot sample it turned out to be monomorphic (as it was in the subsequent targeted sequencing experiment). On the basis of this experience, as well as what we would expect given the unique population history of Samoans, we believe that the best way to perform accurate imputation in Samoans is by using a Samoan-specific reference panel. This idea is in agreement with recent recommendations for optimal fine-mapping in populations with unique ancestry not found in a cosmopolitan reference panel[57]. A panel of 1,295 Samoans from the discovery sample is currently undergoing whole-genome sequencing by the National Heart, Lung, and Blood Institute (NHLBI) TOPMed Consortium. Additional details for targeted sequencing can be found in the **Supplementary Note**.

Imputation. We prephased the targeted sequencing sample using SHAPEIT[49–53] and then imputed into our discovery sample using IMPUTE2 (refs. 54–56). Association testing was carried out using ProbABEL[58], adjusting for relatedness with the empirical kinship matrix generated by GenABEL. Three variants had nearly equivalent *P* values (rs12513649, rs150207780, and rs373863828) because of nearly perfect LD between them (r^2 ≥0.988); imputation was very good for rs150207780 and rs373863828 (IMPUTE2 info metric = 0.954 for both variants). To determine which of these variants might be the most likely causal candidate, we tested for association in the targeted sequencing region with conditioning on each of these variants as well as the next most significant variant (rs3095870; info metric = 0.957), using ProbABEL and adjusting for relatedness. As expected for variants in such high LD, the signals in the region were eliminated after conditioning (**Supplementary Fig. 3**).

Bayesian fine mapping. Details can be found in the **Supplementary Note**.

Confirmatory genotyping. Genotyping was attempted for both rs150207780 and rs373863828 using TaqMan technology in all discovery and replication sample participants. The assay for rs150207780 failed; genotyping was not reattempted because this SNP showed no residual association signal in the analyses of the imputed data with conditioning on the missense variant rs373863828 (**Supplementary Fig. 3**). The replication plates included the 96 samples that had been sequenced in the targeted sequencing experiment. Laboratory personnel were blinded to the sequence-derived genotypes of these 96 samples, as well as to the phenotypes for all the samples. Association analysis was performed using the same regression models and meta-analysis as for the GWAS and replication analyses above. Effect size estimates were calculated using untransformed BMI separately for men and women from the discovery sample with age and age^2 as covariates.

Association analyses of additional phenotypes. rs373863828 genotype was examined for association with the additional adiposity-related phenotypes listed in **Table 2**. Association was assessed in both the discovery sample (**Table 2** and **Supplementary Table 2a**) and a mega-analysis of the adults from the replication sample (**Supplementary Table 2b**). Although meta-analysis of properly transformed phenotypes generates more accurate *P* values (as in **Table 1**), we chose instead to carry out mega-analyses here because

we were primarily interested in estimating effect sizes on the natural scale for each trait. Sex-stratified analyses were also conducted in both samples (**Supplementary Table 2**). Diabetics were excluded from analyses of glucose, insulin, and HOMA-IR. Because the distributions of leptin levels varied greatly for women and men, a combined-sex analysis was not conducted for this trait. Residuals for quantitative traits were generated using linear regression. Age, age^2, sex, and the interactions between age and sex and between age^2 and sex were initially included in sex-combined models. For glucose, insulin, HOMA-IR, adiponectin, leptin, and diabetes status, a second set of models was used that included log-transformed BMI as a covariate. Sex and age × sex interactions were not included in the sex-stratified models. In the replication mega-analysis models, polity (Samoa or American Samoa) and cohort (1990s or 2000s) were initially included in the models as well. Stepwise regression was used to reduce the number of covariates for each trait separately. For quantitative traits, residuals were tested for association using the mmscore function of GenABEL[45], adjusted for the empirical kinship matrix as above. Dichotomous traits were analyzed using the palogist function of ProbABEL[58] while adjusting for covariates and empirical kinship. A Bonferroni-corrected *P*-value threshold of 2.17×10^{-3} was used to assess significance; this threshold is conservative, as it adjusts for 23 tests even though some traits are correlated with each other. To assess a possible survivor effect as the cause of the association between the BMI-increasing allele and decreased fasting glucose levels and risk of diabetes, we conducted linear regression of age by genotype. In the discovery sample, in regard to the association of rs373863828 with BMI, fasting glucose, fasting insulin, obesity risk, and diabetes risk, addition of the first ten 'local' principal components from **Supplementary Figure 1b** into the statistical models had a negligible effect on the effect estimates and statistical significance (data not shown).

Expression of *CREBRF* in human and mouse tissues. For human gene expression analysis, a Human Normal cDNA Array was obtained from Origene Technologies (HMRT103 and HBRT101). The human standard curve was prepared from Control Human Total RNA (Thermo Fisher Scientific, 4307281). For mouse gene expression analysis, mouse tissues were collected from 8–10 a.m. from littermate-matched, *ad libitum*–fed male C56BL/6J mice at 10 weeks of age (n = 6 mice/group). The mouse standard curve was prepared from pooled kidney RNA from the above mice. mRNA was prepared using the RNeasy Lipid Tissue Mini kit with on-column DNase treatment (Qiagen) followed by reverse transcription to cDNA using qScript cDNA Supermix (Quanta Biosciences). Gene expression was determined by qPCR (Quanta PerfeCTa SYBR Green FastMix or PerfeCTa qPCR FastMix) using an Eppendorf Realplex System. Human *CREBRF* was amplified using species-specific primers (**Supplementary Table 3**). Mouse *Crebrf* was amplified using a species-specific primer–probe set (Thermo Fisher Scientific, Mm00661538_m1). *CREBRF* expression was normalized to species-specific peptidylprolyl isomerase A or cyclophilin A as the endogenous control gene (Thermo Fisher Scientific, 4333763T and Mm02342430_g1 for human and mouse, respectively). Mouse data are expressed as means plus s.e.m. Data are relative expression values, and so randomization, blinding, and statistical comparisons were not indicated. Gene expression analysis was performed in accordance with Minimum Information for Publication of Quantitative Real–Time PCR Experiments (MIQE) guidelines. Mouse experiments were approved by the University of Pittsburgh Institutional Animal Care and Use Committee and conducted in conformity with the Public Health Service Policy for Care and Use of Laboratory Animals. Human samples from Origene Technologies conform to federal policies for the protection of human subjects (45 CDR 46) and are HIPAA compliant. Additional information and documentation can be obtained by contacting the company.

Plasmid construction and mutagenesis. Expression plasmids with ORFs for eGFP and human *CREBRF* (NM_153607.2) were obtained from GeneCopoeia (EX-EGFP-M10, EX-E3374-M10). The backbone vector was pReceiver-M10, which has a cytomegalovirus promoter and encodes a C-terminal Myc-$(His)_6$ tag. A rare missense variant, c.1447A>G, p.Thr483Ala (rs17854147), affecting a conserved residue was present in the *CREBRF* ORF. To avoid using this potentially function-altering variant, we converted *CREBRF* to the wild-type sequence and introduced the BMI risk-associated mutation c.1370G>A, p.Arg457Gln (rs373863828), using PCR mutagenesis. The segments obtained by PCR in each plasmid were verified by sequencing before large-scale plasmid purification for transfection.

Cell culture and transfection, adipocyte differentiation, Oil Red O plate assays, microscopy, triglyceride assays, and quantitative RT–PCR. These methods are described in detail in the **Supplementary Note**.

Bioenergetic profiling. OCR, a measure of mitochondrial respiration, and ECAR, a measure of glycolysis, were determined using an XF96 extracellular flux analyzer (Seahorse Bioscience). Transfected 3T3-L1 cells were seeded in a 96-well XF96 cell culture microplate (Seahorse Bioscience) at a density of 7,000 cells per well in 200 µl of DMEM (4.5 g/l glucose) supplemented with 10% FBS (Sigma) 36 h before measurement. Six replicates per cell type were included in the experiments, and four wells were chosen evenly in the plate to correct for temperature variation. On the day of the assay, the growth medium was exchanged for assay medium (unbuffered DMEM with 4.5 g/l glucose). Oligomycin at a final concentration of 2.0 µM, FCCP (carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone) at 1.0 µM, 2-deoxyglucose at 100 mM, and rotenone at 15.0 µM were sequentially injected into each well in accordance with the manufacturer's protocol. Basal mitochondrial respiration, maximal respiration, ATP production, and basal glycolysis were determined according to the manufacturer's instructions. At the conclusion of the assay, cells in the analysis plate were lysed using CelLytic M (Sigma). Protein concentration was measured using the Bradford assay[59] and used to normalize the bioenergetic profile data.

Starvation and rapamycin treatment. 3T3-L1 preadipocytes were subjected to starvation for 0, 2, 4, 12, and 24 h by culturing cells in Hank's balanced salt solution (HBSS). To investigate the response to refeeding starving cells, a set of cells undergoing 12 h of starvation was fed with fresh growth medium for an additional 12 h (**Fig. 3a**). For rapamycin stimulation, preadipocytes were treated with 20 ng/ml rapamycin (Sigma), for 2, 4, 12, and 24 h. A set of cells kept in rapamycin for 12 h was cultured in fresh growth medium for the following 12 h (**Fig. 3b**). To quantify cell survival, 3T3-L1 cells and transfected cells were seeded in six-well plates at 86,000 cells per well. Two days later, the cells were starved in HBSS. At 0, 2, 4, 6, 12, and 24 h, the cells were collected and 100 µl of the cell suspension samples was added to an equal volume of trypan blue (Life Technologies). The mixture was loaded into an automated cell counter (Cellometer Mini, Nexcelom Bioscience), and viable cell numbers were measured. Cell death rates were calculated by subtracting the number of viable cells at 6 h from cell numbers at 0 h and dividing the result by the cell numbers at 6 h.

Cell studies statistical analysis. For the cell studies, adequate sample sizes were determined on the basis of publications using similar methods and pilot experiments. No blinding was used. Each experiment was performed twice with similar results unless otherwise stated in the corresponding figure legend. The data were initially evaluated by one-way ANOVA implemented in SPSS (IBM). The homogeneity of variances was examined using Levene's test. Two-sided Bonferroni and Games–Howell *post-hoc* tests were used to compare data with equal and unequal variance, respectively. Alternatively, pairwise two-sided *t* tests for unequal variance were used. $P < 0.05$ was considered to be statistically significant. SPSS analyses were verified using the same tests as implemented in R (ref. 60).

Selection analyses. On the basis of the genome-wide Affymetrix 6.0 SNP genotype data, we used Primus[61,62] to select 626 individuals from the discovery sample using a kinship threshold (0.039) halfway between the values expected for first and second cousins, so that first cousins and more closely related relatives were excluded. These 'unrelated' individuals were then haplotyped using SHAPEIT[49–53] and were annotated with ancestral allele information using the selectionTools pipeline[63]. Haplotype bifurcation diagrams and EHH plots were drawn using the rehh R package[64]. The haplotype bifurcation diagram[65] visualizes the breakdown of LD as one moves away from the core allele at the focal SNP; each branch reflects the creation of new haplotypes, and the thickness of the line reflects the number of samples with the haplotype. EHH represents the

doi:10.1038/ng.3620

probability that two randomly chosen chromosomes are identical by descent from the focal SNP to the current position of interest[65]. Selection at the core allele is expected to result in EHH values close to 1 in an extended region centered on the focal SNP. To measure the deviation, we used selscan[66] to compute the iHS[67], which is defined as the log of the ratio of the integrated EHH for the derived allele over the integrated EHH for the ancestral allele. These values are then normalized in frequency bins across the whole genome (we used 25 bins). Note that selscan's definition of iHS differs from earlier definitions where the ancestral allele was in the numerator of the ratio[66,67]. In our case, a large positive iHS indicates that a derived allele has had its frequency increase owing to selection. We computed an approximate two-sided *P* value under the assumption that after normalization the iHS is approximately distributed as a standard normal. We also used selscan to compute nS_L scores (the number of segregation sites by length)[68]. The nS_L is similar to the iHS, but instead of integrating over genetic distance the nS_L uses the number of segregating sites as a measure of 'distance'. Thus, the nS_L is more robust to demographic assumptions than the iHS, as it does not depend on a genetic map. As with the iHS, we normalized the nS_L scores in 25 frequency bins across the whole genome and computed approximate two-sided *P* values assuming a standard normal distribution. The selscan program was run using its assumed default values. As we were focused on testing whether there is positive selection at the missense variant, we did not adjust the *P* values for multiple testing.

34. Gauderman, W.J. Sample size requirements for association studies of gene–gene interaction. *Am. J. Epidemiol.* **155**, 478–484 (2002).
35. Gauderman, W.J. Sample size requirements for matched case–control studies of gene–environment interaction. *Stat. Med.* **21**, 35–50 (2002).
36. Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
37. McGarvey, S.T., Levinson, P.D., Bausserman, L., Galanis, D.J. & Hornick, C.A. Population-change in adult obesity and blood-lipids in American-Samoa from 1976–1978 to 1990. *Am. J. Hum. Biol.* **5**, 17–30 (1993).
38. Keighley, E.D., McGarvey, S.T., Turituri, P. & Viali, S. Farming and adiposity in Samoan adults. *Am. J. Hum. Biol.* **18**, 112–122 (2006).
39. Swinburn, B.A., Ley, S.J., Carmichael, H.E. & Plank, L.D. Body size and composition in Polynesians. *Int. J. Obes. Relat. Metab. Disord.* **23**, 1178–1183 (1999).
40. Cole, T.J., Bellizzi, M.C., Flegal, K.M. & Dietz, W.H. Establishing a standard definition for child overweight and obesity worldwide: international survey. *Br. Med. J.* **320**, 1240–1243 (2000).
41. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **35** (Suppl. 1), S64–S71 (2012).
42. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
43. Laurie, C.C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).
44. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
45. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
46. Heath, S.C. *et al.* Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet.* **16**, 1413–1429 (2008).
47. Chen, W.M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* **81**, 913–926 (2007).
48. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
49. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
50. Delaneau, O., Howie, B., Cox, A.J., Zagury, J.F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
51. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
52. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* **10**, e1004234 (2014).
53. Delaneau, O. & Marchini, J.; 1000 Genomes Project Consortium; 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
54. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
55. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
56. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
57. Wang, X. *et al.* Evaluation of transethnic fine mapping with population-specific and cosmopolitan imputation reference panels in diverse Asian populations. *Eur. J. Hum. Genet.* **24**, 592–599 (2016).
58. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
59. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
60. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2004).
61. Staples, J., Nickerson, D.A. & Below, J.E. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet. Epidemiol.* **37**, 136–141 (2013).
62. Staples, J. *et al.* PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet.* **95**, 553–564 (2014).
63. Cadzow, M. *et al.* A bioinformatics workflow for detecting signatures of selection in genomic data. *Front. Genet.* **5**, 293 (2014).
64. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
65. Sabeti, P.C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
66. Szpiech, Z.A. & Hernandez, R.D. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).
67. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
68. Ferrer-Admetlla, A., Liang, M., Korneliussen, T. & Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **31**, 1275–1291 (2014).