RESULTS

Associations of a specific *CDKAL1* splicing variant (*CDKAL1-v1*) with T2D-associated SNPs and insulin secretion

The T2D-associated SNPs of *CDKAL1* are associated with impaired insulin secretion as well as a decrease in CDKAL1-mediated ms^2-modification in $tRNA^{Lys}$(UUU) (8,10). To investigate the correlation of T2D-associated SNPs with the gene function, we purified DNA and RNA from peripheral blood samples from non-diabetic volunteers and compared the expression level of full-length *CDKAL1* (NM_017774.3) in individuals carrying risk SNPs. We examined rs10946398 because the SNP has been extensively studied in T2D and are in high linkage disequilibrium with other SNPs such as rs7756992 and rs7754840 (11). The expression level of full-length *CDKAL1* in carriers of the risk allele C/C of rs10946398 was not significantly different from that in carriers of the non-risk alleles A/A and A/C (Fig. 1A), although we observed a trend showing a decrease in *CDKAL1* mRNA levels in carriers of risk alleles ($P = 0.06$, 87% for C/C carriers compared with A/A carriers, Fig. 1A).

Interestingly, we noticed there are two short splicing variants of human *CDKAL1* in the NCBI database: AK310219 and BC064145, hereinafter referred as *CDKAL1-v1* and *CDKAL1-v2*, respectively. We then examined the expression levels of *CDKAL1-v1* and *CDKAL1-v2* in carriers of risk alleles. *CDKAL1-v1*, but not *CDKAL1-v2*, was markedly decreased in carriers of the risk alleles for rs10946398 (Fig. 1B and C). The significant decrease in *CDKAL1-v1* was also observed in carriers of the risk allele for rs7756992 (Supplementary Material, Fig. S1A and B). In addition to peripheral blood samples, the decrease in the *CDKAL1-v1* level was also confirmed in human embryonic fibroblast cell lines carrying the risk alleles for rs10946398 and rs7756992 (Supplementary Material, Fig. S1C and D). Furthermore, to investigate the specificity of the correlation of *CDKAL1-v1* with *CDKAL1* SNPs, we examined the expression level of *CDKAL1-v1* in carriers of the T2D-associated risk alleles of *KCNQ1*, which is also one of the most reliable T2D-associated genes (12,13). There was no

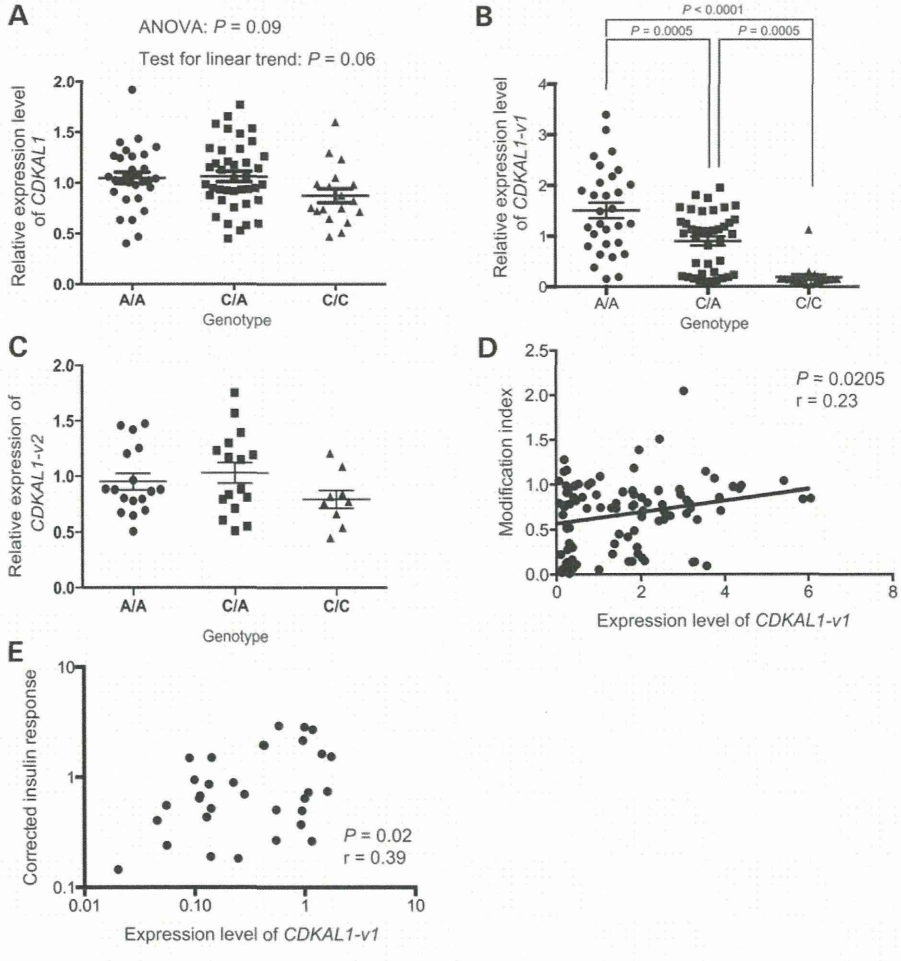

Figure 1. Decreased expression of *CDKAL1-v1* in carriers of *CDKAL1* risk alleles. (**A**–**C**) Comparison of the *CDKAL1* (A), *CDKAL1-v1* (B) or *CDKAL1-v2* (C) levels in groups of homozygous carriers of non-risk alleles (A/A), heterozygous carriers of a non-risk and a risk allele (C/A) and homozygous carriers of risk alleles (C/C) for rs10946398. $n = 30$ for A/A, $n = 40$ for C/A, $n = 18$ for C/C. (**D**) Correlation between the ms^2-modification level of $tRNA^{Lys}$(UUU) and the *CDKAL1-v1* level in individuals with mixed genotypes. $n = 94$. (**E**) Correlation of the corrected insulin response with the CDKAL1-v1 level in individuals with mixed genotypes. $n = 32$. Error bars represent mean ± SEM.

significant change in *CDKAL1-v1* level in carriers of the *KCNQ1* risk allele (C/C) compared with carriers of non-risk alleles (T/T and C/T) (Supplementary Material, Fig. S2), suggesting that a decrease in *CDKAL1-v1* is solely dependent on the risk alleles of *CDKAL1*.

Cdkal1-dependent 2-methylthio (ms^2-) modification in $tRNA^{Lys}$(UUU) is decreased in individuals carrying the risk alleles of the *CDKAL1* SNPs and is correlated with insulin secretion (7,8). Given that *CDKAL1* SNPs were associated with *CDKAL1-v1* levels, it is speculated that *CDKAL1-v1* level might also correlate with the ms^2-modification level and insulin secretion. The ms^2-modification and *CDKAL1-v1* level were then subjected to quantitative analysis in total RNA isolated from peripheral blood specimens. As expected, there was a direct correlation between *CDKAL1-v1* and ms^2-modification level ($r = 0.23$, $P = 0.0205$, Fig. 1D). Next, we performed oral glucose tolerance tests in non-diabetic individuals and investigated the correlation of *CDKAL1-v1* with insulin secretion. There was a direct correlation between *CDKAL1-v1* level and insulin secretion ($r = 0.39$, $P = 0.02$, Fig. 1E). These results suggest that the decrease in *CDKAL1-v1* may contribute to the development of T2D through impairment of CDKAL1 function in individuals carrying the risk alleles in *CDKAL1*.

Regulation of protein and mRNA levels of *CDKAL1* by *CDKAL1-v1*

The *CDKAL1-v1* transcript contains exons 1–4 and a portion of 3′UTR derived from intron 4 (Fig. 2A). Because the open-reading frame (ORF) of *CDKAL1-v1* lacks the entire radical SAM domain and tRNA binding domain, which are absolutely required for methylthiotransferase activity (6), CDKAL1-v1 is not a functional tRNA-modifying enzyme. Thus, the direct correlation of *CDKAL1-v1* with ms^2-modification level could not be explained by a change in the enzyme activity of CDKAL1-v1. We then speculated that *CDKAL1-v1* might modulate the ms^2-modification level by regulation of the cellular CDKAL1 level. To investigate whether *CDKAL1-v1* plays a direct role in the regulation of CDKAL1 levels, *CDKAL1-v1* was acutely silenced by specific small interfering RNAs (siRNAs) targeting the 3′UTR in HEK293 cells. Knockdown of *CDKAL1-v1* markedly reduced the CDKAL1 protein level (Fig. 2B). Accordingly, the ms^2-modification level was significantly decreased in *CDKAL1-v1*-silenced cells (Fig. 2C), suggesting that *CDKAL1*-v1 regulates CDKAL1 activity in a translation-dependent manner. Interestingly, treatment with *CDKAL1-v1*-specific siRNAs resulted in a moderate but significant decrease in *CDKAL1* mRNA (83% decrease by #1 siRNA, and 70% decrease by #2 siRNA compared with control, Fig. 2D). Conversely, treatment of *CDKAL1*-specific siRNA resulted in a moderate but significant decrease in *CDKAL1-v1* (79% decrease by siCdkal1, Fig. 2E). The co-regulation of *CDKAL1* and *CDKAL1-v1* was also observed in other cell lines such as HeLa cells as well as human islet carcinoma QGP-1 cells (Supplementary Material, Fig. S3). These results prompted us to examine the correlation between *CDKAL1* and *CDKAL1-v1* levels in individuals regardless of genotype. Indeed, we observed a significant correlation between *CDKAL1-v1* and *CDKAL1* in individuals ($r = 0.25$, $P = 0.016$, Fig. 2F). The expression levels of *CDKAL1* and *CDKAL1-v1* were also examined in normal human tissues and cell lines. There was a direct correlation between *CDKAL1* and *CDKAL1-v1* ($r = 0.5385$, $P = 0.0038$, Fig. 2G). Taken together, these results suggest that there is a post-transcriptional regulation between *CDKAL1-v1* and *CDKAL1*.

CDKAL1-v1 is a non-coding RNA

The 5′UTR of *CDKAL1-v1* is ~100 bases longer than that of CDKAL1, and contains three upstream ATG codons (Fig. 3A). Given the potential translational suppression by the upstream ORFs and the truncation of tRNA-modifying domains in *CDKAL1-v1*, we suspected that *CDKAL1-v1* may not be translated and may instead regulate protein levels as a non-coding transcript. To confirm this speculation, we conjugated the 5′UTR or putative ORF of *CDKAL1-v1* to *Luciferase* and examined the expression level (Fig. 3B and C). While the 5′UTR of full-length *CDKAL1* did not affect Luciferase activity, the 5′UTR of *CDKAL1-v1* significantly attenuated Luciferase activity (48% compared with Control 1). The putative ORF of *CDKAL1-v1* markedly reduced Luciferase activity (4.8% compared with Control 1). Combining both the 5′UTR and the putative ORF of *CDKAL1* almost eliminated Luciferase activity (2.8% compared with Control 1). In contrast to the luciferase activities, the mRNA levels of the *CDKAL1*- and *CDKAL1-v1*-derived luciferase reporters were not significantly different from the mRNA level of the control reporter. These results suggest that *CDKAL1-v1* is not actively translated into functional protein. Furthermore, we examined the expression of CDKAL1 using a specific antibody targeting the *N*-terminus region of CDKAL1, which is encoded by both *CDKAL1* and *CDKAL1-v1*. The antibody recognized the 61 kDa full-length CDKAL1 protein in extracts of HEK cells, which disappeared following the knockdown of both *CDKAL1* and *CDKAL1-v1* (Fig. 3D and E). In contrast, there was no detectable band of ~11 kDa, which is thought to be the molecular weight of translated *CDKAL1-v1*, in extracts from either the control or *CDKAL1/CDKAL1-v1* knockdown cells. These results suggest that *CDKAL1-v1* regulates CDKAL1 level through a coding-independent mechanism. Because *CDKAL1-v1* is a short splicing variant that contains a premature stop codon, there is a concern that *CDKAL1-v1* may be subjected to nonsense-mediated mRNA decay (NMD) (14). *CDKAL1-v1* levels were examined in cells treated with siRNA targeting *UFP1*, which is a key component of the cellular NMD system (Fig. 3F). However, the *CDKAL1-v1* level in UFP1-silenced cells was not different from that in control-siRNA treated cells. Furthermore, we performed *in situ* hybridization in HEK cells to examine the cellular location of *CDKAL1-v1* transcripts. While the probe targeting both *CDKAL1* and *CDKAL1-v1* showed the most intense signals in the cell bodies, the *CDKAL1-v1*-specific probe also resulted in moderate signals at the same locations (Fig. 3G), suggesting that *CDKAL1-v1* is a functional non-coding transcript in human cells.

Regulation of cellular Cdkal1 level by a *CDKAL1-v1*-targeting miRNA

Because *CDKAL1-v1* was positively correlated with the CDKAL1 protein level as well as the *CDKAL1* mRNA level, it is unlikely that *CDKAL1-v1* acts as an anti-sense non-coding

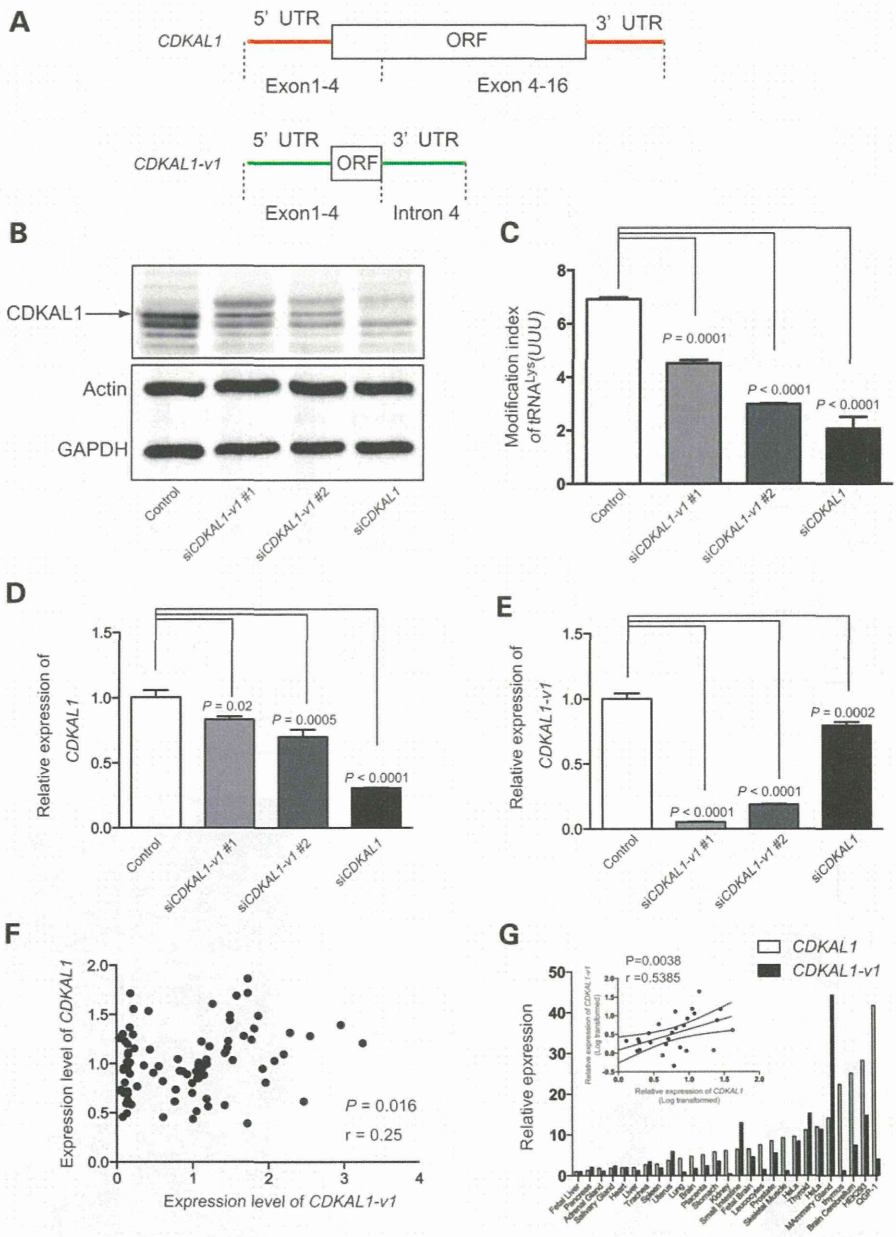

Figure 2. Regulation of CDKAL1 level by *CDKAL1-v1*. (**A**) Structures of *CDKAL1* and *CDKAL1-v1*. The 5′UTR and 3′UTR of *CDKAL1-v1* are different from those of *CDKAL1*. The difference is highlighted by a red line in *CDKAL1* and a green line in *CDKAL1-v1*. (**B**) Decrease in the CDKAL1 level in HEK293 cells transfected with *CDKAL1-v1*-targeting siRNAs (si*CDKAL1-v1* #1 and si*CDKAL1-v1* #2). The *CDKAL1*-targeting siRNA (si*CDKAL1*) was used as positive control. (**C**) Decrease in ms2-modification in HEK cells transfected with *CDKAL1-v1*-targeting siRNAs. $n = 4$. (**D**) Downregulation of *CDKAL1* by *CDKAL1-v1*-targeting siRNAs. $n = 4$. (**E**) Downregulation of *CDKAL1-v1* by *CDKAL1*-targeting siRNAs. $n = 4$. (**F**) Correlation of *CDKAL1* and *CDKAL1-v1* level in individuals with mixed genotypes. (**G**) Correlation of *CDKAL1* and *CDKAL1-v1* level in human tissues and cell lines. Inserted graph in the upper left corner shows linear regression of *CDKAL1* versus *CDKAL1-v1*. Error bars represent mean ± SEM.

RNA for *CDKAL1*. Interestingly, recent studies have revealed a unique regulatory mechanism by which some transcripts could regulate the cognate genes through a coding-independent mechanism (15–18). In this model, non-coding RNA indirectly regulates its cognate gene by the competitive binding of common miRNAs at the 3′UTR region. We speculated that *CDKAL1-v1* might regulate CDKAL1 levels through the same mechanism. To investigate whether the 3′UTR of *CDKAL1-v1* is capable of regulating CDKAL1 levels, the 3′UTR region of *CDKAL1-v1* or the entire *CDKAL1-v1* was transfected into primary human fibroblast cells that carried the risk alleles and had low CDKAL1 protein levels as well as *CDKAL1-v1* level (HE37 in Supplementary Material, Fig. S4). Expression of the 3′UTR of *CDKAL1-v1* or the entire *CDKAL1-v1* transcript induced

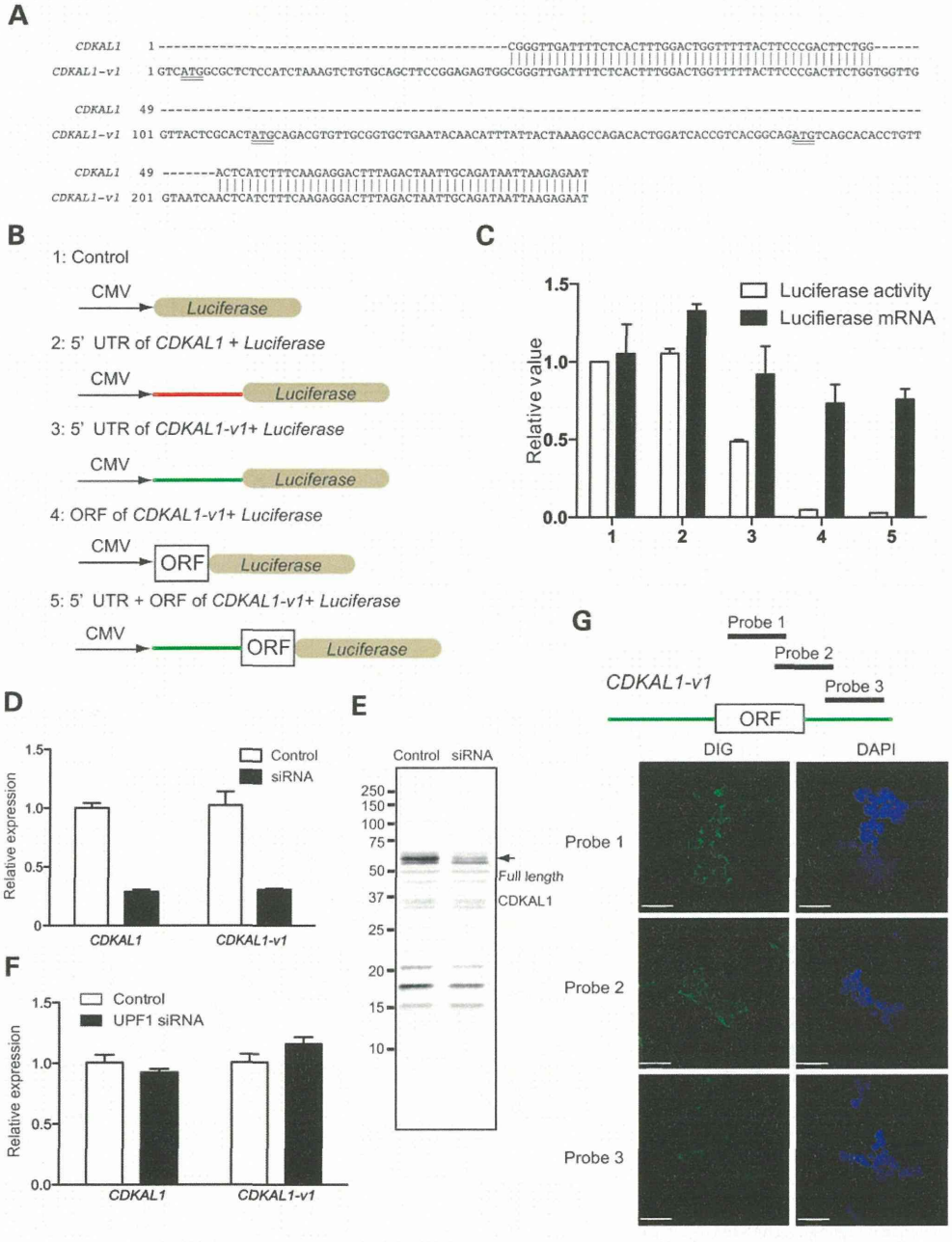

Figure 3. *CDKAL1-v1* is a non-coding transcript. (**A**) Alignment of 5′UTR of *CDKAL1* and *CDKAL1-v1*. ATG codons in upstream ORF are underlined. (**B**) Constructs for the luciferase-based reporter assay with which to examine the 5′UTR and ORF of *CDKAL1-v1*. (**C**) Suppression of luciferase activity by the 5′UTR and ORF of *CDKAL1-v1* is translation-dependent but not transcription-dependent. (**D**) Knockdown of both *CDKAL1* and *CDKAL1-v1* in HEK293 cells by a siRNA targeting both transcripts. $n = 4$. (**E**) Western blotting to examine the translated product of *CDKAL1-v1*. The arrow indicates 61 kDa, full length, CDKAL1. (**F**) Knockdown of *UPF1* in fibroblast cells had no effect on *CDKAL1* or *CDKAL1*-v1 level. $n = 4$. (**G**) Detection of cellular localization of *CDKAL1-v1* by *in situ* hybridization. The upper illustration shows the regions to which the probes hybridize. The lower images show the cellular localization of probes (Digoxigenin). The nucleus is detected by DAPI staining. Bars = 100 μm. Error bars represent mean ± SEM.

expression of the CDKAL1 protein (Fig. 4A). The 3′UTR of *CDKAL1-v1* also increased the *CDKAL1* mRNA level (121% compared with control, Fig. 4B). These results suggest that the 3′UTR of *CDKAL1*-v1 is capable of controlling CDKAL1 levels.

Next, we searched potential *CDKAL1-v1*- and *CDKAL1*-targeting miRNAs that are also expressed in human pancreatic islets (19,20). We found miR-494 as a potential candidate from the miRNA database (Fig. 4C). To validate our finding, we transfected miR-494 and 3′UTR of *CDKAL1-v1* or *CDKAL1* fused

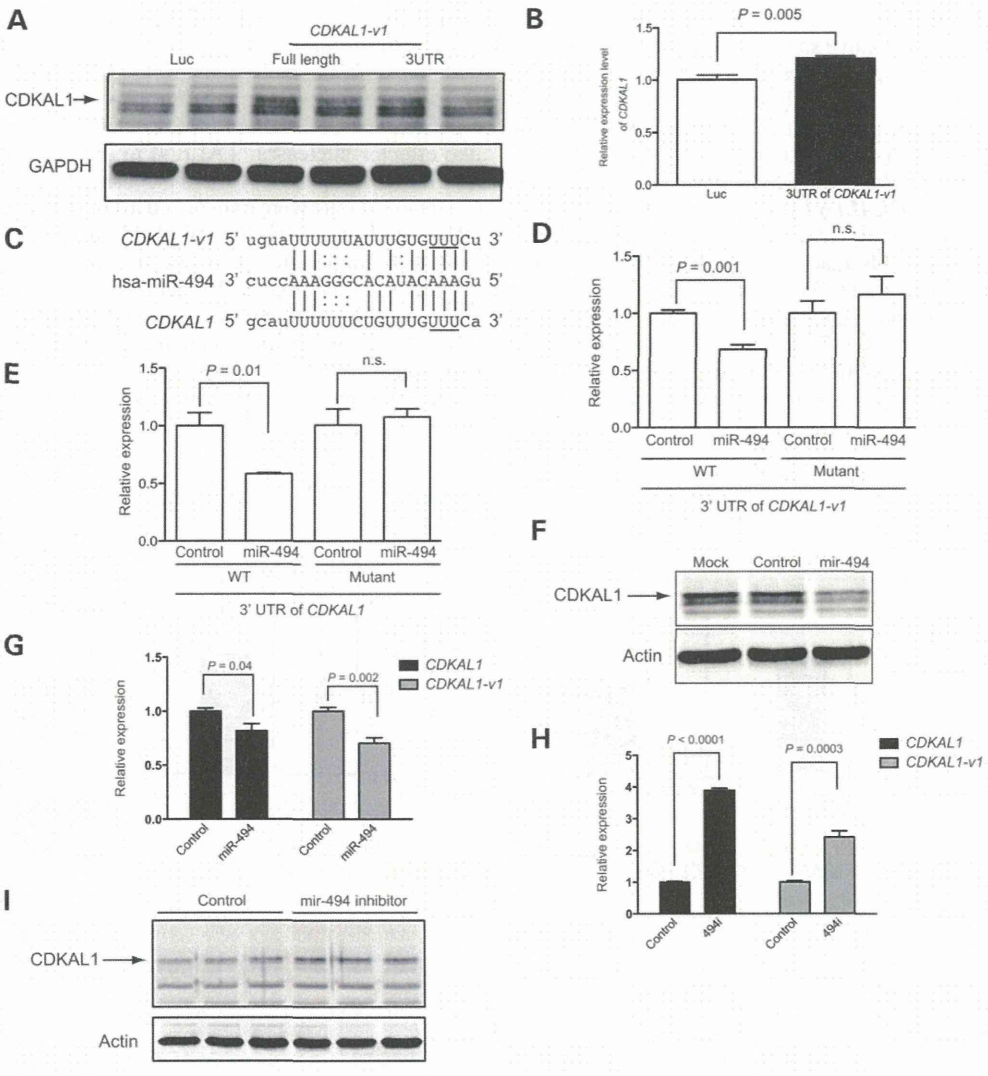

Figure 4. Control of CDKAL1 level by *CDKAL1-v1*-targeting miRNA. (**A**) Increase in the CDKAL1 level following overexpression of full-length *CDKAL1-v1* and 3′UTR of *CDKAL1-v1*. The arrow indicates bands corresponding to CDKAL1. (**B**) Increase in the *CDKAL1* level following overexpression of the 3′UTR of *CDKAL1-v1*. $n = 4$. (**C**) Binding of has-miR-494 to both *CDKAL1* and *CDKAL1-v1*. The solid lines represent canonical pairing. The dotted lines represent non-canonical pairing. The underlined UUU were mutated to AAA to make miR-494-insensitive constructs. (**D** and **E**) miR-494 suppressed the luciferase gene-carrying wild-type (WT) 3′UTR of *CDKAL1-v1* (D) or *CDKAL1* (E) but not the luciferase gene-carrying miR-494-insensitive (Mutant) 3′UTR of *CDKAL1-v1* (D) or *CDKAL1* (E). $n = 4$. (**F**) Reduced Cdkal1 protein level in response to miR-494. (**G**) Reduced *CDKAL1-v1* and *CDKAL1* levels in response to miR494. (**H**) Derepression of *CDKAL1-v1* and *CDKAL1* by a miR-494 inhibitor (494i). $n = 4$. (**I**) Increase in the CDKAL1 protein level in response to a miR-494 inhibitor (494i). The arrow indicates bands corresponding to CDKAL1. Error bars represent mean ± SEM.

with *Luciferase* and monitored Luciferase activity as a readout for miR-494-mediated suppression (Fig. 4D and E). Expression of miR-494 significantly suppressed Luciferase activity. Furthermore, miR-494 no longer suppressed Luciferase activity when the seed sequences of the miR-494-targeting site in the 3′UTR of *CDKAL1-v1* or *CDKAL1* were mutated (Fig. 4D and E). We also examined the direct effect of miR-494 on endogenous *CDKAL1-v1* or *CDKAL1* levels. In HEK293 cells, overexpression of miR-494 markedly reduced CDKAL1 protein levels as well as *CDKAL1-v1* and *CDKAL1* mRNA levels (Fig. 4F and G), suggesting that *CDKAL1-v1* and *CDKAL1* are common targets for miR-494. Furthermore, introducing miR-494 inhibitor induced both *CDKAL1-v1* and *CDKAL1* levels (Fig. 4H). Accordingly, the inhibitor of miR-494 enhanced CDKAL1 protein levels (Fig. 4I). These results indicate that *CDKAL1-v1* and *CDKAL1* are subject to miR-494-mediated post-transcriptional regulation. Thus, a decrease in *CDKAL1-v1* would conversely result in an increase in miR-494-mediated suppression of *CDKAL1*, as observed in carriers of risk alleles.

A T2D-associated SNP is directly involved in alternative splicing of *CDKAL1-v1*

Finally, we attempted to answer the question of why a single-nucleotide substitution in intron 5 of *CDKAL1* results in a drastic decrease in the expression level of *CDKAL1-v1*.

We hypothesized that the risk SNPs of *CDKAL1* may have an inhibitory effect on the alternative splicing. To test this hypothesis, we utilized an AMO that specifically binds to the 5′ end of the U1 small nuclear ribonucleic acid (snRNA) (U1 AMO) to inhibit splicing (21,22). In cells carrying non-risk SNPs, inhibition of splicing is predicted to result in a decrease of *CDKAL1*. However, in cells carrying risk SNPs, inhibition of splicing should not affect the *CDKAL1-v1* level, because the generation of *CDKAL1-v1* by splicing has already been impaired. As expected, in HEK293 cells that carry non-risk SNPs, the U1 AMO significantly decreased *CDKAL1-v1* (Fig. 5A). In contrast, in HeLa cells that are homozygous for risk SNPs, the U1 AMO had no effect on *CDKAL1-v1* level (Fig. 5B). Moreover, the U1 AMO significantly decreased *CDKAL1* levels in both HEK293 cells and HeLa cells. The same results were also confirmed in primary human fibroblast cell lines (Supplementary Material, Fig. S5). These results suggest that the risk SNPs are specifically associated with impaired alternative splicing to generate *CDKAL1-v1*.

To further investigate the direct involvement of SNPs in the regulation of *CDKAL1-v1*, we utilized transcription activator-like effector nucleases (TALEN) to directly edit the SNP site in *CDKAL1*. Specific TALEN constructs targeting the SNP (rs10946398) site were transfected in combination with template DNA into HeLa cells (Fig. 5C). Although there were no HeLa cells with single-nucleotide substitutions, we established five cell lines containing 3–19 nucleotide deletions around the SNP site (Fig. 5D). Subsequent analysis of *CDKAL1-v1* and *CDKAL1* levels in these cell lines revealed that deletion of nucleotides in the 3′ direction from the SNP site (Cell line #11 and #38) enhanced the expression of both *CDKAL1-v1* and

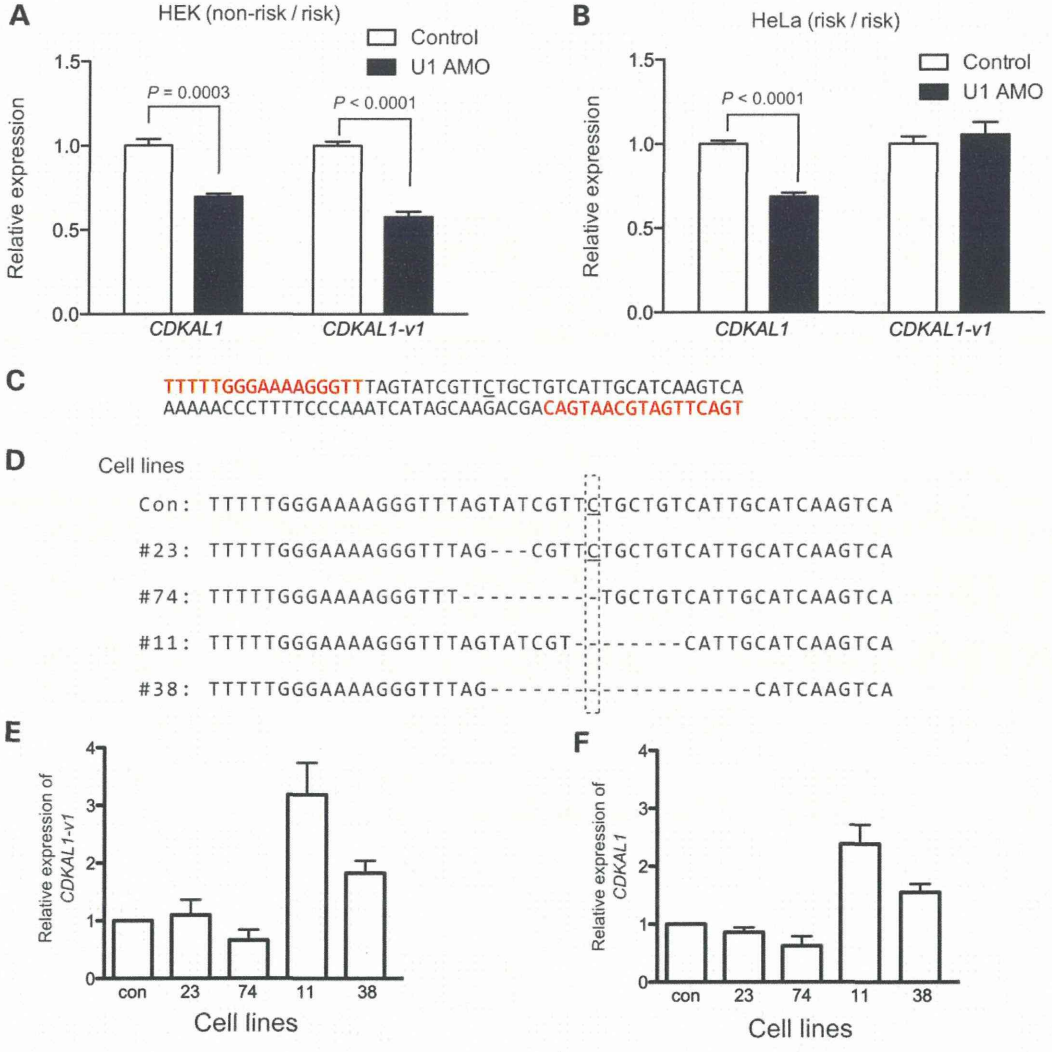

Figure 5. Modulation of splicing by SNPs in *CDKAL1*. (**A**) Inhibition of U1 snRNA by U1 AMO suppressed both the *CDKAL1* and *CDKAL1-v1* level in HEK293 cells carrying A/A alleles (non-risk/risk) for rs10946398. (**B**) U1 AMO suppressed *CDKAL1* levels but not *CDKAL1-v1* levels in HeLa cells carrying C/C alleles (risk/risk). (**C**) Genomic region (nucleotides in red) targeted by TALEN constructs. The underlined C is the SNP site for rs10946398. (**D**) Genomic sequences surround the SNP site in isolated HeLa cell lines transfected with TALEN. Dashed lines represent deleted nucleotides. Dashed rectangles show the position of the SNP site. (**E** and **F**) Expression level of *CDKAL1-v1* (E) and *CDKAL1* (F) in cell lines with an edited SNP site. $n = 4$. Error bars represent mean ± SEM.

CDKAL1, while deletion of nucleotides in the 5′ direction from the SNP site (Cell line #74) slightly decreased the expression (Fig. 5E and F). Furthermore, the expression levels of *CDKAL1* and *CDKAL1-v1* were not altered in cell line containing three nucleotide deletions but without alteration of SNP site (Cell line #23). Taken together, these results suggest that SNPs in *CDKAL1* are critical for alternative splicing to generate *CDKAL1-v1*.

DISCUSSION

In the present study, we discovered a novel molecular mechanism underlying the development of T2D in individuals carrying risk SNPs in *CDKAL1*. A human-specific splicing variant of *CDKAL1*, *CDKAL1-v1*, was significantly correlated with T2D-associated risk alleles of *CDKAL1*. Furthermore, the *CDKAL1-v1* level was directly correlated with CDKAL1-mediated ms^2-modification and insulin secretion. Functional analysis revealed that *CDKAL1-v1* is a non-coding transcript and that it indirectly regulates the CDKAL1 level by acting as a decoy for a *CDKAL1*-targeting miRNA. In addition, we also demonstrated that the nucleotides surrounding risk SNPs were critical for alternative splicing to generate *CDKAL1-v1*. Our results suggest that in individuals carrying the risk SNPs in *CDKAL1*, impaired alternative splicing causes a dramatic decrease in the *CDKAL1-v1* level, which then results in a shift of miRNA targeting from *CDKAL1-v1* toward *CDKAL1*. Subsequently, miRNA-mediated suppression of cellular CDKAL1 levels causes a decrease in ms^2-modification in $tRNA^{Lys}$(UUU), which ultimately contributes to the dysfunction of pancreatic β-cells as revealed in Cdkal1 knockout mice (7).

Twelve SNPs in *CDKAL1*, including the three SNPs examined in the present study, have been associated with T2D [rs4712523 (23); rs4712524 (13); rs10440833 (24); rs6931514 (25); rs9465871 (26); rs9295474 (27); rs9356744, rs7773318, rs9465994 (28)]. To date, there is no general consensus about the causal SNP in these 12 SNPs. In addition, clear evidence to show the correlation between these SNPs and *CDKAL1* level has been insufficient. Interestingly, all SNPs except two SNPs (rs7773318 and rs9465994) locate in the intron 5, suggesting genetic variations in this particular intron might have profound effect on *CDKAL1* level. Previous studies have shown that the nucleotides that surround the SNPs occasionally act as *cis*- or *trans*-regulatory elements and regulate the expression of genes in the vicinity and at a distance (29,30). A conserved element near risk SNPs in *Cdkal1* was shown to affect the expression of the nearby gene *Sox4* in mice and zebrafish (31). Because *Sox4* has been implicated in the development of the pancreas, there is a possibility that the risk SNPs in *CDKAL1* may contribute to T2D through regulation of the expression of *SOX4*. However, the *SOX4* expression level in carriers of risk alleles of *CDKAL1* was not significantly different from that in carriers of non-risk alleles (Supplementary Material, Fig. S6). Given the extent of the reduction of the *CDKAL1-v1* level in the carriers of risk alleles and the significant association of the *CDKAL1-v1* level with insulin secretion, it is most likely that the risk SNPs in intron 5 are the critical elements for controlling expression of *CDKAL1-v1* and ultimately contribute to the development of T2D.

The T2D-associated SNPs in *CDKAL1* are not located at critical sites for splicing such as 3′- and 5′-splicing sites, polypyrimidine tract or branch sites. Nevertheless, several pieces of evidence have shown that such intronic SNPs might be capable of affecting splicing patterns (32–34). The SNPs in *TCF7L2* have been extensively associated with T2D (35–37). These T2D-associated SNPs have also been found to be associated with differential splicing patterns (32,33), which might affect the function of *TCF7L2* and lead to the development of T2D. Similar to these results, we demonstrated that the inhibition of splicing by the U1 AMO specifically inhibits the production of *CDKAL1-v1* in cells carrying non-risk alleles but not in cells carrying homozygous risk alleles. These results not only support our hypothesis that the SNPs in *CDKAL1* are actively involved in splicing, but also provide the molecular basis that these SNPs may affect splicing through interactions with U1 small nuclear RNP (snRNP). Recent studies have found that U1 snRNP is critical for the determination of mRNA length and the regulation of isoform expression in addition to its classic role as a member of a spliceosome, through direct interaction with nucleotides in the intron (38). Interestingly, such interaction does not necessarily require direct base pairing (39,40). We thus propose that the risk SNPs in *CDKAL1* may impair the interaction between the U1 snRNP and the nucleotides near SNPs, which causes aberrant processing of the nascent transcription of *CDKAL1* and results in a decrease in *CDKAL1-v1*. Moreover, given the similar effects of SNPs in *CDKAL1* and *TCF7L2*, our results thus suggest the broad relevance of regulation of splicing pattern in the pathogenesis of T2D-associated SNPs.

In the present study, we utilized the TALEN method for direct editing of the genome and showed that SNP sites play an active role in the regulation of gene expression. We successfully obtained cell lines with deletion around the SNP site and demonstrated that the *CDKAL1-v1* level was altered in these cell lines. Interestingly, regulation of *CDKAL1-v1* seems to depend on the degree and direction of the deletion around the SNP site. It is currently unclear whether the nucleotide in the SNP site or the surrounding nucleotides including SNP site are responsible for the subsequent effects on gene expression. Furthermore, the incomplete recovery of the *CDKAL-v1* expression level after alteration of the risk SNP site (rs10946398) suggest that other risk SNP sites might also contribute the suppression of *CDKAL1-v1*. Future studies utilizing a more efficient genome-editing tool such as CRISPR (41,42) may achieve single-nucleotide substitution in all risk SNP sites and answer these questions. Nevertheless, a previous study has found that the intronic SNPs might cause different structural folds in the nascent transcripts (43). Therefore, it is likely that the risk SNPs in *CDKAL1* may affect the secondary structure of pre-mRNA by changing the chemical interaction with surrounding nucleotides so that splicing is not be properly achieved.

A unique finding in the present study is that *CDKAL1-v1* is a non-coding transcript and acts as a decoy for *CDKAL1*-targeting miRNA to modulate the cellular Cdkal1 level. A similar regulatory mechanism has been revealed in the regulation of cancer-related genes, such as *PTEN* and *KRAS*, by the cognate pseudogenes *PTENP1 and KRAS1P* (15). The so-called competing endogenous RNA (ceRNA) mechanism expands the classic role of non-coding RNA as a suppressor for its anti-sense target gene, and

provides a novel function for non-coding RNA as an active regulator of its cognate gene through its abilities to compete for miRNA binding at the 3′UTR (17). The observation of the miRNA-mediated correlation of *CDKAL1-v1* with *CDKAL1* in human tissues and various cell lines nicely fits *CDKAL1-v1* into the ceRNA theory. In addition to *CDKAL1-v1*, previous studies have also found that some splicing variants of *TCF7L2* are correlated with full-length *TCF7L2* in the same direction (33). These results suggest that the ceRNA theory may apply to a variety of disease including cancer and T2D. Nevertheless, further study using primary human pancreatic islets is needed to elucidate the direct contribution of *CDKAL1-v1* to T2D.

In the present study, we identified miR-494 as a candidate miRNA that targets both *CDKAL1* and *CDKAL1-v1*. Previous studies showed that miR-494 was expressed in various human tissues with the highest expression in pancreatic β-cells (19,20). Interestingly, overexpression of miR-494 impaired mitochondrial function by down-regulating the expression of some genes involved in mitochondrial transcription as well as the respiratory chain (44,45). Since mitochondrial function is crucial for insulin secretion from pancreatic β-cells, it is conceivable that the decrease of *CDKAL1-v1* in risk SNP carriers would enhance miR-494-mediated suppression of mitochondrial functions as well as CDKAL1-dependent tRNA modification, ultimately leading to the dysfunction of β-cells.

A limitation of the functional study of T2D-associated intronic SNPs is that most of the intronic SNPs are not conserved in mouse and other animal models. The risk SNPs in *CDKAL1* are not conserved in the mouse genome, nor is murine *CDKAL1-v1* found in the current databases. However, we found some truncated transcripts in murine mRNA NCBI database (AK169761 and AK158546) that contain an upstream ORF and share common miRNA targeting sequences with *Cdkal1* in 3′UTR. Thus, it is conceivable that the ceRNA-mediated regulatory mechanism might be evolutionarily conserved, while the intermediate miRNA might be different between species.

In conclusion, T2D-associated SNPs in *CDKAL1* was associated with marked decrease of the *CDKAL1-v1* expression level, possibly by influencing proper splicing. The reduction in the level of *CDKAL1-v1* causes a decrease in the CDKAL1 level by enhancing miRNA-mediated suppression of *CDKAL1*, which ultimately contributes to the development of T2D. Our results present a clear correlation between T2D-associated SNPs and *CDKAL1* gene expression for the first time, and reveal a novel molecular mechanism underlying the development of T2D. Moreover, our results suggest that restoration of the homeostasis between miRNA and its cognate transcripts may be a new approach for treatment of T2D.

MATERIALS AND METHODS

Participant recruitment

Blood samples were obtained from non-diabetic volunteers at Kumamoto University. Informed consent was obtained before collecting blood. The Ethics Committee of Kumamoto University approved the experiments regarding the analysis of human genetic materials and the collection of human blood (approval ID: Genome 159).

Oral glucose tolerance test

Individuals were fasted beginning at 21:00 the night before the test day. These individuals were then given a 75 g glucose load (Ajinomoto) anytime between 9:00 and 11:00. Peripheral blood samples were obtained when the individuals were fasting and 30 min after the glucose load. Plasma glucose and serum insulin concentrations were measured by enzyme-linked immunosorbent assay (SRL Inc.). Corrected insulin response (CIR30) was calculated as I30/[G30 × (G30−70)] as described previously (1).

Genotyping of SNPs

Genomic DNA was purified from 200 μl of peripheral human blood using the QIAamp DNA Blood Mini Kit (Qiagen) and adjusted to 10 ng/μl with distilled water. The SNPs (rs10946398, rs7756992 and rs7754840) in *CDKAL1* and the SNP in *KCNQ1* (rs2237892) were examined using the TaqMan SNP Genotyping Assay kit (Life Technologies).

Isolation of RNA and gene expression assay

Total RNA fractions of cultured cell lines were isolated using TRIzol reagent (Life Technologies) according to the manufacturer's protocol. For isolation of total RNA from peripheral blood, 1.5 ml fresh peripheral blood was immediately mixed with the erythrocyte lysis buffer provided in the QIAamp RNA Blood Mini Kit (Qiagen). Total RNA fractions of leukocytes were isolated according to the manufacturer's protocol. Total RNA of normal human tissues was derived from commercial available Human Total RNA Master Panel II (TAKARA). A PrimerScript RT reagent kit (TAKARA) was used to generate cDNA. Quantitative real-time PCRs were performed using either a Taqman Gene Expression Kit (Applied Biosystems) or SYBR Premix Ex Taq (TAKARA). The results were normalized to the level of GAPDH or 18S rRNA. Information regarding the probes and primers used in this study is described in Supplementary Material, Table S1.

Measurement of ms^2-modification

Measurement of ms^2-modification in total RNA isolated from peripheral blood specimens or cultured cell lines was performed using the qPCR-MtR method as described previously (8). Total RNA was adjusted to 50 ng/μl in RNase-free water. Two microliters (100 ng) of RNA were digested by five units of DNase I (Roche Applied Science) in a 20 μl reaction at 37°C for 20 min. DNase I was inactivated by heat-inactivation at 75°C for 10 min. After the DNase treatment, 2.5 μl of digested RNA was subjected to reverse transcription (Transcriptor, Roche Applied Science) using reverse primer r1 (CCTGGACCCTCAGATTAAAA) or reverse primer r2 (GAACAGGGACTTGAACCCTG). Two microliters of synthesized complementary DNA (cDNA), forward primer (GTCGGTAGAGCATCAGACTT) and reverse primer r1 were subjected to quantitative PCR using the SYBR Premix Ex Taq Kit (Takara) and the ABI PRISM 7300 Real-time PCR System (Life Technologies) according to the manufacturer's protocol.

Relative modification was calculated as described previously (8). Briefly, the threshold cycle number (referred to as CTr1 or

CTr2) was obtained from quantitative PCR using a cDNA template that was generated by primer r1 or primer r2. The difference between CTr1 and CTr2 (hereafter dCTr2r1) was calculated by subtracting CTr1 from CTr2. The dCTr2r1 number for individual samples was used to compare the relative modification rate between samples.

MicroRNA selection

To search the common miRNA that can target both 3′UTR of *CDKAL1* and 3′ UTR of *CDKAL1-v1*, an online search program (www.microrna.org) was utilized with the most stringent criteria (conserved and with good mirSVR scores). The candidate miRNAs were further validated using another online search program (www.targetscan.org). The common candidate miRNA (miR-494) was selected for experiments.

Cell culture

HeLa cells and HEK293 cells were cultured in Dulbecco's modified Eagle's medium (Life Technologies) containing 10% fetal bovine serum. QGP-1 cells were obtained from the Japanese Collection of Research Bioresources Cell Bank and cultured according to a previous report (46). Human embryonic primary fibroblast cell lines were provided by the RIKEN BRC through the National Bio-Resource Project of the MEXT, Japan. The fibroblast cells were cultured in minimal essential medium containing 10% fetal bovine serum, 1 mM sodium pyruvate, 0.2 mM serine and 0.2 mM aspartate. Plasmids and siRNAs were transfected using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol.

For western blotting, cells were lysed in lysis buffer containing 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, pH 7.2, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid, 1% NP-40, 1% sodium deoxycholate, 0.1% sodium dodecyl sulfate, protease inhibitor cocktail (Roche). The protein concentration was adjusted to 1 mg/ml by bicinchoninic acid reagent (Piece). CDKAL1 was detected by homemade antibody targeting the *N*-terminus of human CDKAL1.

For *in situ* hybridization, DNA regions of putative ORF (208–542 of Ak310219), regions flanking ORF and the 3′UTR (428–943) and the regions of 3′UTR of *CDKAL1-v1* (555–952) were cloned into pCRII-TOPO. A sense probe and an antisense probe were transcribed using either the T7 promoter or the Sp6 promoter using PCR DIG Labeling Mix (Roche). Hybridization was performed according to the manufacturer's protocol (Roche). Digoxigenin (DIG) was detected by anti-digoxigenin antibody (Roche) followed by Alexa488 anti-mouse secondary antibody. Images were acquired using Fluoview 300 (Olympus).

Plasmids and siRNA

All siRNAs were purchased from Life Technologies. For knockdown of *CDKAL1*, an siRNA targeting exon 16 in *CDKAL1* was selected. For knockdown of *CDKAL1-v1*, siRNAs targeting the 3′UTR of *CDKAL1-v1* were selected. For knockdown of both *CDKAL1 and CDKAL1-v1*, an siRNA targeting exon 4 was selected. A morpholino antisense oligo targeting U1 snRNA (U1 AMO) was designed as described previously (19) and purchased from Gene Tool, LLC. Human miR-494 (miRIDIAN Mimic) was purchased from Thermo Scientific. Inhibitor for miR-494 (mirVana miRNA inhibitor) was purchased from Life Technologies. The information regarding siRNAs or U1 AMO is provided in Supplementary Material, Table S1.

Luciferase assay

For the Luciferase-based assay, the 5′UTR or ORF regions of *CDKAL1-v1* and *CDKAL1* were cloned into multi-cloning sites in pGL4.14 (Promega). For the 3′UTR assay, 3′UTR of *CDKAL1-v1* or *CDKAL1* was cloned into the 3′ downstream of *Luciferase* in pGL4.14. Luciferase reporter vectors (100 ng) and the *Renilla* luciferase reporter vector (10 ng) were co-transfected into HEK293 cells for 16 h. Relative Luciferase activity to *Renilla* Luciferase activity was examined using the dual Luciferase reporter assay system (Promega). For examining the mRNA level of *Luciferase*-conjugated *CDKAL1* or *CDKAL1-v1* transcripts, report vectors were transfected into HEK cells, and total RNAs were purified by TRIzol (Invitrogen). To eliminate contamination with the plasmid vector, 2 μg total RNA was digested with DNase I (Roche) for 30 min and then subjected to quantitative PCR. The relative mRNA levels of reporters were examined by normalizing the level of *Luciferase* to that of *Renilla Luciferase*.

Genome editing by TALEN

Specific TALEN constructs for editing nucleotides in SNP sites (rs10946398) were designed and purchased from Cellectis Bioresearch. A DNA template containing 1.5 kb genomic regions from the 5′ direction and the 3′ direction of the SNP site (3 kb in total) was amplified from human DNA carrying risk alleles (C/C) or non-risk alleles (A/A), and cloned into pCRII-TOPO vector (Life Technologies). HeLa cells were seeded in 35 mm dishes at a density of 3×10^5 cells/ml. The next day, 2 μg of each TALEN construct and 4 μg of DNA template were co-transfected in HeLa cells using Lipofectamine 2000. Four days after transfection, HeLa cells were trypsinized and diluted to 10 cells/ml. One hundred microliters of cell suspension was transferred to each well of the 96-well plate. Colony grown from a single cell was isolated and expanded. To confirm genomic editing, the genomic region containing the SNP site was amplified by PCR followed by DNA sequencing. The expression levels of *CDKAL1* and *CDKAL1-v1* were examined as described above.

Statistics

Statistical analyses were performed using Prism 6 Software (GraphPad Software). Student's *t*-test was used to test the difference between two groups. Analysis of variance (one-way ANOVA) was used to test the difference among multiple groups followed by a *post hoc* examination of the *P*-value between two groups. A two-tailed *P*-value of 0.05 was considered statistically significant.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Y. Ueda and N. Maeda for technical assistance.

Conflict of Interest Statement. None declared.

FUNDING

This work was supported by a Grant-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by the Japan Society for the Promotion of Science (JSPS) through its "Funding Program for Next Generation World-Leading Researchers", and by the Takeda Science Foundation.

REFERENCES

1. Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S. *et al.* (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.*, **39**, 770–775.
2. Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. *et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, **316**, 1331–1336.
3. Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341–1345.
4. Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. *et al.* (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**, 1336–1341.
5. Dehwah, M.A., Wang, M. and Huang, Q.Y. (2010) CDKAL1 and type 2 diabetes: a global meta-analysis. *Genet. Mol. Res.*, **9**, 1109–1120.
6. Arragain, S., Handelman, S.K., Forouhar, F., Wei, F.Y., Tomizawa, K., Hunt, J.F., Douki, T., Fontecave, M., Mulliez, E. and Atta, M. (2010) Identification of eukaryotic and prokaryotic methylthiotransferase for biosynthesis of 2-methylthio-*N*6-threonylcarbamoyladenosine in tRNA. *J. Biol. Chem.*, **285**, 28425–28433.
7. Wei, F.Y., Suzuki, T., Watanabe, S., Kimura, S., Kaitsuka, T., Fujimura, A., Matsui, H., Atta, M., Michiue, H., Fontecave, M. *et al.* (2011) Deficit of $tRNA^{Lys}$(UUU) modification by Cdkal1 causes the development of type 2 diabetes in mice. *J. Clin. Invest.*, **121**, 3598–3608.
8. Xie, P., Wei, F.Y., Hirata, Y., Kaitsuka, T., Suzuki, T., Suzuki, T. and Tomizawa, K. (2013) Quantitative PCR measurement of tRNA 2-methylthio modification for assessing Type 2 diabetes risk. *Clin. Chem.*, **59**, 1604–1612.
9. McCarthy, M.I. (2010) Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.*, **363**, 2339–2350.
10. Chistiakov, D.A., Potapov, V.A., Smetanina, S.A., Bel'chikova, L.N., Suplotova, L.A. and Nosikov, V.V. (2011) The carriage of risk variants of CDKAL1 impairs beta-cell function in both diabetic and non-diabetic patients and reduces response to non-sulfonylurea and sulfonylurea agonists of the pancreatic KATP channel. *Acta Diabetol.*, **48**, 227–235.
11. Rong, R., Hanson, R.L., Ortiz, D., Wiedrich, C., Kobes, S., Knowler, W.C., Bogardus, C. and Baier, L.J. (2009) Association analysis of variation in/near FTO, CDKAL1, SLC30A8, HHEX, EXT2, IGF2BP2, LOC387761, and CDKN2B with type 2 diabetes and related quantitative traits in Pima Indians. *Diabetes*, **58**, 478–488.
12. Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., Hirota, Y., Mori, H., Jonsson, A., Sato, Y. *et al.* (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.*, **40**, 1092–1097.
13. Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jørgensen, T. *et al.* (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.*, **40**, 1098–1102.
14. Isken, O., Kim, Y.K., Hosada, N., Mayeur, G.L., Hershey, J.W. and Maquat, L.E. (2008) Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell*, **133**, 314–327.
15. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. and Pandolfi, P.P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.
16. Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F. *et al.* (2011) Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*, **147**, 344–357.
17. Salmena, L., Poliseno, L., Tay, Y., Kats, L. and Pandolfi, P.P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
18. Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A. and Yoshiki, A. (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature*, **423**, 91–96.
19. Klein, D., Misawa, R., Bravo-Egana, V., Vargas, N., Rosero, S., Piroso, J., Ichii, H., Umland, O., Zhijie, J., Tsinoremas, N. *et al.* (2013) MicroRNA expression in alpha and beta cells of human pancreatic islets. *PloS One*, **8**, e55064.
20. van de Bunt, M., Gaulton, K.J., Parts, L., Moran, I., Johnson, P.R., Lindgren, C.M., Ferrer, J., Gloyn, A.L. and McCarthy, M.I. (2013) The miRNA profile of human pancreatic islets and beta-cells and relationship to type 2 diabetes pathogenesis. *PLoS One*, **8**, e 55272.
21. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L. and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468**, 664–668.
22. Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell*, **150**, 53–64.
23. Rung, J., Cauchi, S., Albrechtsen, A., Shen, L., Rocheleau, G., Cavalcanti-Proença, C., Bacot, F., Balkau, B., Belisle, A., Borch-Johnsen, K. *et al.* (2009) Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.*, **41**, 1110–1115.
24. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
25. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G. *et al.* (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.*, **40**, 638–645.
26. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
27. Sim, X., Ong, R.T., Suo, C., Tay, W.T., Liu, J., Ng, D.P., Boehnke, M., Chia, K.S., Wong, T.Y., Seielstad, M. *et al.* (2011) Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet.*, **7**, e1001363.
28. Kuo, J.Z., Sheu, W.H., Assimes, T.L., Hung, Y.J., Absher, D., Chiu, Y.F., Mak, J., Wang, J.S., Kwon, S., Hsu, C.C. *et al.* (2013) Trans-ethnic fine mapping identifies a novel independent locus at the 3′ end of CDKAL1 and novel variants of several susceptibility loci for type 2 diabetes in a Han Chinese population. *Diabetologia*, **56**, 2619–2628.
29. Pagani, F. and Baralle, F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet.*, **5**, 389–396.
30. Visel, A., Rubin, E.M. and Pennacchio, L.A. (2009) Genomic views of distant-acting enhancers. *Nature*, **461**, 199–205.
31. Ragvin, A., Moro, E., Fredman, D., Navratilova, P., Drivenes, O., Engstrom, P.G., Alonso, M.E., de la Calle Mustienes, E., Gomez Skarmeta, J.L., Tavares, M.J. *et al.* (2010) Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc. Natl. Acad. Sci. U S A.*, **107**, 775–780.
32. Mondal, A.K., Das, S.K., Baldini, G., Chu, W.S., Sharma, N.K., Hackney, O.G., Zhao, J., Grant, S.F. and Elbein, S.C. (2010) Genotype and tissue-specific effects on alternative splicing of the transcription factor 7-like 2 gene in humans. *J. Clin. Endocrinol. Metab.*, **95**, 1450–1457.
33. Pang, D.X., Smith, A.J. and Humphries, S.E. (2013) Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes. *Nutr. Metab. Cardiovasc. Dis.*, **23**, 550–556.