Biochemical and Biophysical Research Communications 392 (2010) 89–94

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Screening of Kozak-motif-located SNPs and analysis of their association with human diseases

Heng Xu [a,1], Ping Wang [a,1], Jin You [a,1], Yufang Zheng [c], Yujie Fu [a], Quan Tang [a], Li Zhou [a], Zejun Wei [a], Bin Lin [a], Yang Shu [a], Yufei Zhu [a], Landian Hu [a,b,*], Xiangyin Kong [a,b,*]

[a] The Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS) and Shanghai Jiao Tong University School of Medicine (SJTUSM), 225 South Chong Qing Road, Shanghai 200025, PR China
[b] State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiaotong University, 197 Rui Jin Road II, Shanghai 200025, PR China
[c] Department of Physiology and Biophysics, School of Life Sciences, Fudan University, Shanghai 200433, PR China

ARTICLE INFO

Article history:
Received 30 December 2009
Available online 6 January 2010

Keywords:
Kozak motif
Start codon
Translation
Single nucleotide polymorphism
Disease

ABSTRACT

The Kozak motif, which is located near the translational start codon, often regulates the protein translation. Moreover, it is believed that the conserved positions −3 and +4 contribute the most. Since changes that occur in this motif have a great impact on protein yield and in some cases are associated with disease, we screened the human SNP database for all Kozak-motif-located SNPs (kSNPs) and focused on the strong-changed kSNPs (sckSNPs). Many intron-located and synonymous SNPs are reported to be associated with disease, though the mechanisms underlying these associations are poorly understood. Here, we performed haplotype analysis on sckSNP-containing genes and found that there are some sckSNPs that exist in the same haplotype blocks of reported intron-located and synonymous disease-associated SNPs, indicating that those kSNPs could be a true risk factor for disease-association by affecting the efficiency of protein expression. Our findings provide a candidate explanation for how diseases are associated with intron-located and synonymous SNPs.

Introduction

The Kozak motif, the context sequence around the mRNA start codon, plays an important role in protein synthesis by regulating translation efficiency. In eukaryotes, start codons are recognized most efficiently within the Kozak motif GCCRCCaugG. By defining the A of the AUG codon as +1, positions −3 and +4 are highly conserved. More than 90% of mRNAs have a purine (usually A) at −3 and around 50% have G at +4 [1]. The efficiency of protein expression is strongly impacted when −3R or +4G is substituted by other nucleotides. This motif configuration is known as a weak Kozak motif [2,3]. When the transcripts contain a weak Kozak motif, nucleotides at positions −6, −5, −4, −2 and −1 of the Kozak motif could also influence translation efficiency [4]. Therefore, diseases can be caused by mutations in the Kozak motif, especially at positions −3 and +4, such as partial androgen insensitivity syndrome and α-thalassemia [5]. SNPs located in a Kozak motif (kSNPs) could also be associated with disease, but so far, only four genes, *CHAT*, *SCARB1*, *CD40* and *GP1BA*, have been reported. A C/T polymorphism at position −1 of CD40 and position −5 of GP1BA (glycoprotein Ib, α polypeptide) was reported to be associated with Graves' disease [6,7] and ischemic stroke [8,9], respectively. Similarly, an A/G polymorphism at position +4 of *CHAT* (choline acetyltransferase) and *SCARB1* (scavenger receptor class B, member 1) were reported to be associated with Alzheimer's disease (AD) [10] and type 2 diabetes [11], respectively. However, only the association between *CD40* and Graves' disease has been experimentally confirmed for its translational pathology, in which the kSNP at position −1 of *CD40* is a risk factor, as it affects the efficiency of protein expression [7].

Besides kSNP, diseases could also be caused by start-codon-located SNPs (sSNPs), which results in the translation of a truncated protein that is initiated from a second in-frame AUG start site, which will be expressed and might induce dysfunction as dominant negative. Take *PSCA* (prostate stem cell antigen) as an example: sSNP at position +2 is associated with urinary bladder cancer [12] and gastric cancer [13].

However, only very few kSNPs and sSNPs have been reported to be directly associated with disease, compared with many other SNPs that are located in other positions, such as those located in introns. However, the mechanism by which these intron-located and synonymous SNPs impact the morbidity of diseases is still poorly understood. Since SNPs display a pattern of linkage disequilibrium (LD), nearby SNP alleles tend to be associated with each other to form a series of haplotype blocks [14]. We suspect that kSNPs or sSNPs could be in the same haplotype block of reported

* Corresponding authors. Fax: +86 21 64678976.
E-mail addresses: ldhu@sibs.ac.cn (L. Hu), xykong@sibs.ac.cn (X. Kong).
[1] These authors contribute equally to this work.

0006-291X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2010.01.002

disease-associated intron-located and synonymous SNPs, in turn becoming the cause of the disease.

In our study, we first screened the NCBI SNP database for all reported kSNPs and sSNPs. After haplotype block analysis with the extracted SNP haplotype data provided by the Hapmap Project, we found that some sckSNPs were in the same haplotype blocks of the reported disease-related SNPs, such as those in *DBI* and *PTGS2*, indicating that sckSNPs could be the true risk factors for various diseases. Our study provides a candidate explanation for how diseases are caused by the reported disease-associated SNPs, especially those located in introns and synonymous coding SNPs.

Materials and methods

Bioinformatics screening. All of our information on SNPs and the human translation start codon was downloaded from the NCBI dbSNP build 127 and Genbank build 36.3, respectively. Since the contexts of the recorded SNPs were provided with their genome sequences, SNPs were mapped to their corresponding DNA contigs. After comparing the positions of each SNP with their same-contig-located start codons, we recorded the detailed information of the SNPs that were located within the −9 to +10 region of the DNA, where the A of the AUG codon is designated as +1. The status of the genes that were screened out was also recorded, including "REVIEWED", "VALIDATED", "MODEL", "PROVISIONAL", and so on. Some adjacent sequences near the SNPs in the database were recorded on reversed complementary pattern compared with the according transcripts; we changed these SNPs into their complementary pattern in order to make them consistent with the sequence of the transcripts in which they were located. The number of SNPs was calculated based on their positions and SNP types.

Haplotype block analysis. According to the positions of the SNPs we analyzed, the adjacent SNP information from four populations was downloaded from the Hapmap Project website (www.hapmap.org). SNP haplotype blocks were analyzed with the software Haploview (downloaded from http://www.broadinstitute.org/mpg/haploview).

Plasmid construction. The pcDNA3.1(+)-GFP expression plasmid was constructed by cloning the GFP coding sequence (without the AUG initiation codon) into the EcoR I and Not I sites of the pcDNA3.1(+) mammalian expression vector. The first coding exons (i.e., the exon containing the translation initiating AUG) of human *CD40*, *APOA5*, *DBI* and *PTGS2* were then amplified by PCR and cloned into the Hind III and EcoR I sites of pcDNA3.1(+)-GFP. Different SNP alleles were then introduced into the corresponding vectors using the Quikchange Multi Site-Directed Mutagenesis Kit (Stratagene) and verified by sequencing.

Cell culture and transfection. 293T-cells were maintained in DMEM containing 10% FBS (Invitrogen) at 37 °C in a humidified 5% CO_2 atmosphere incubator. For transient transfections, 293T-cells were plated at 80% density and were transfected the following day with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Cells were harvested 30 h after transfection and divided into two even parts, for either RNA or protein expression analysis (Western or FACS).

RNA isolations and semi-quantitative RT-PCR analysis. Total RNA from 293T transiently transfected cells was prepared with Trizol reagent (Invitrogen) according to the manufacturer's protocol. The same amount of each RNA sample (500 ng) was used for single-stranded cDNA synthesis with the Reverse Transcription System kit (Invitrogen), according to the manufacturer's protocol. The GFP (forward: 5′ CAAGTTCAGCGTGTCCG 3′; reverse: 5′ GGTT CACCAGGGTGTCG 3′) and β-actin (forward: 5′ CTCCATCCTGGC CTCGCTGT 3′; reverse: 5′ GCTGTCACCTTCACCGTTCC 3′) primers were taken to proceed with the semi-quantitative RT-PCR at 30 and 35 cycles, respectively.

Western blot and FACS analysis. The other even part of transfected cells was either lysed and subjected to western blotting with an antibody against GFP (Clonetech) and tubulin (Sigma), or digested with 0.05% trypsin (BBI), washed and resuspended with PBS, and subjected to FACS to measure the mean fluorescent intensity of GFP.

Results

Searching for the kSNPs and sSNPs by screening the NCBI SNP database

SNPs located in the context of AUG (−9 to +10) were sorted by mapping all human SNPs published on the NCBI website to the reported mRNAs. The number of SNPs on these sites was listed based on the location and SNP type (Fig. 1A and Supplementary Table 1). We noticed that some genes were annotated with "MODEL", or "PROVISIONAL", and so on. To avoid false positives, total genes and the genes annotated with "REVIEWED" and "VALIDATED" were counted and listed, respectively (Fig. 1A). All the reported disease-associated kSNPs, such as those in *CD40*, *GP1BA*, *CHAT* and *SCARB1*, were contained in our list (Supplementary Table 1).

We focused on the kSNPs at positions −3 and +4 that have the strongest influence on protein translation efficiency. According to our results, there are significantly fewer SNPs located at position +4 compared with other coding positions. However, there is only a little difference between the number of SNPs at position −3 and other 5′-UTR positions (Fig. 1A). We noticed that nearly half of the SNPs at position −3 were A/G polymorphisms which barely affect protein translation. When the A/G SNPs were excluded in all positions, the number of SNPs at −3 was also significantly less than SNPs at other 5′-UTR positions (Fig. 1B). One explanation is that these SNPs are unlikely to be preserved through evolution, since mutations at position −3 and +4 are even more likely to be deleterious.

Since a purine-to-pyrimidine polymorphism at position −3 and a G-switch at position +4 have strong impact on protein translation efficiency, we named the A/C, A/T, C/G or G/T at position −3 and A/G, C/G or G/T at position +4 SNPs as strong-changed kSNPs (sckSNPs), while naming other SNPs as weak-changed kSNPs (wckSNPs). The genes with sckSNP were listed in Table 1. Within these sckSNPs, the A/G polymorphisms at position +4 of *CHAT* and *SCARB1* have been reported to be associated with AD risk [10] and type 2 diabetes [11], respectively. Carefully looking at Table 1, we noticed that some transcripts with sckSNP at −3 do not have a G at position +4 and some transcripts with sckSNP at +4 do not have R at position −3. These genes are marked in bold in Table 1. For example, there is a T but not a G at position +4 of *DBI*, which has an A/T SNP at its −3 position. According to the Kozak rules, these kinds of SNPs could impact the translation efficiency most strongly. However, no *DBI*-like sckSNP has, to date, been reported as disease-associated.

Meanwhile, we also screened the sSNPs (Fig. 1A and Supplementary Table 1). In most cases, a truncated protein product initiated from a second in-frame AUG is expressed. The length of original proteins and peptides lost in truncated proteins were listed in Supplementary Table 1.

Similarly, the reported disease-associated sSNPs were screened out and listed in our table, such as *PSCA*. SNP rs2294008 of *PSCA*, which was bladder and gastric cancer associated [12,13], alters the start codon and was predicted to cause a nine amino acid truncation at the N-terminal, disturbing the potential signal peptide (http://www.cbs.dtu.dk/services/SignalP/).

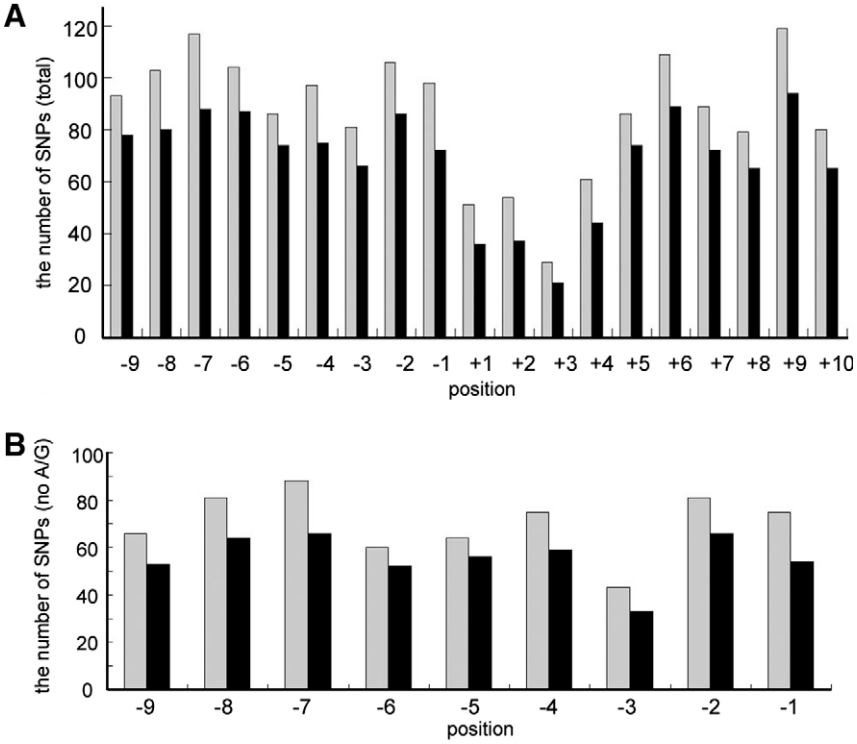

Fig. 1. The number of SNPs in different positions. (A) The number of all types of SNPs within the Kozak motif and adjacent positions; (B) the number of SNPs without an A/G type SNP in the position from −1 to −9. The gray bars and the black bars indicate the number of all genes and the genes annotated with "REVIEWED" or "VALIDATED", respectively.

In our list, we noticed that a relatively long peptide could be lost in many cases. Nearly half of the sSNP-containing transcripts may lose more than 10% of their amino acids. Meanwhile, more than 20% of such transcripts lose 30% of their amino acids (if the first downstream in-frame AUG is used as the start codon). Interestingly, many proteins could lose their conserved domains. A good example is a transcript of *POU5F1* (also known as *OCT4*), where there is a common G/T SNP that is located at position +2. As a consequence of this SNP, the truncated protein is 101 amino acids shorter than the original product, which has 265 amino acids. The lost peptide contains a highly conserved POU domain and a nuclear location signal (NLS) with the sequence of LKQKRITL (http://www.predictprotein.org/cgi/var/nair/resonline.pl). Therefore, we suspect that this sSNP impact on *POU5F1* function greatly, especially in cases with a GG genotype. Additionally, six sSNP-containing genes even have no downstream in-frame AUG.

Haplotype block analysis on sckSNPs and their adjacent disease-associated SNPs

We analyzed the sckSNP-containing genes screened out above and noticed that some of them contained synonymous or intron-located, disease-associated SNPs. Since only genetic associations were reported (i.e., no mechanistic explanation was provided), we suspected that the sckSNPs in those genes could also be disease-associated and may in fact be a true risk factor by impacting protein expression. After analyzing the sckSNPs and their adjacent disease-associated SNPs, we found that some of them were exactly in the same haplotype block. We listed the brief information of the sckSNP-containing genes, including their associated diseases and corresponding disease-associated SNPs, the genomic distance between the disease-associated SNPs and the sckSNPs and results of the haplotype block analysis. The sckSNP-containing genes with a weak context at their corresponding −3 or +4 (*DBI*-like genes) are listed in Supplementary Table 2, while those with a strong context are listed in Supplementary Table 3.

For example, *DBI* is a gene with two intron-located SNPs (rs2084202 and rs8192503) and a non-synonymous coding SNP A/G (rs8192506). These *DBI* SNPs were suggested to be associated with anxiety disorder and type II diabetes [15–17] without mechanistic explanation. Based on the theory of haplotypes and linkage disequilibrium (LD), if one SNP is associated with a phenotype, the other SNPs that are located in the same block might be as well [14]. Therefore, we suspected that those three SNPs of *DBI* could be in the same haplotype block of the sckSNP A/T (rs3209864) at position −3 (Supplementary Table 2). To test this hypothesis, we searched out SNPs that were located in the region of *DBI* from the Hapmap Project data (www.hapmap.org). There is no information of the sckSNP (rs3209864 in Chr2: 119841095) and the reported disease-associated SNP (rs2084202 in Chr2: 119841499) of *DBI* in the Hapmap data. Thus, we analyzed the neighboring SNPs of rs3209864, including rs6756192, rs3731607, rs8192503, rs2276596, rs11123519, rs956309, rs8192506, rs2289948, and rs2289946 in Chr2 from 119840174 to 119846807, which are extracted from the data of different populations within the Hapmap Project. Our results showed that all these SNPs were indeed in the same haplotype block (Fig. 2A and data not shown). Therefore, we suspected that the sckSNP of *DBI* (rs3209864) could also be disease-associated and may in fact be the true risk factor, inducing diseases by affecting protein yield.

Another example is *PTGS2*, which has been found to be associated with many diseases, such as asthma [18], inflammation [19], and gastric cancer [20]. We noticed that rs689466 (Chr1: 184917374) rather than rs20417 (Chr1: 184916944) of *PTGS2* was reported to be associated with asthma in an Australian population [18], while rs20417 was reported to be associated with inflammation in a Spanish population [19] and gastric cancer in a Dutch population [20]. With European population data, we found

Table 1
Detailed information of sckSNPs in the reviewed or validated genes list.

sckSNP position	SNP ID	Gene name	SNP type	Corresponding +4 or −3
−3	5848372	*DLEC1*	−/C*	G
	3209253	***COX7A1***	**A/C**	**C**
	41305775	*GABRG1*	A/C	G
	41535344	***SCGB1A1***	**A/C**	**A**
	7094259	*SLIT1*	A/C	G
	12891217	*AK7*	A/T	G
	11554349	*CMAS*	A/T	G
	3209864	***DBI***	**A/T**	**T**
	56203218	*GNGT2*	A/T	G
	1056544	*KLK7*	A/T	G
	2975116	*IGF2R*	C/G	G
	11573020	*IGFBP7*	C/G	G
	11804508	*OBSCN*	C/G	G
	5914273	***PAGE5***	**C/G**	**C**
	3134591	***PTGS2***	**C/G**	**C**
	2568372	*REL*	C/G	G
	59065783	***CDK5RAP2***	**G/T**	**A**
	4457918	***DHRS4***	**G/T**	**C**
	1051187	***ICAM2***	**G/T**	**T**
	7189194	***LOC283902***	**G/T**	**C**
	17523387	*MRPS35*	G/T	G
	12912647	*RYR3*	G/T	G
	8113914	*TOMM34*	G/T	G
+4	35254068	*FCER1A*	−/G*	A
	60325461	***ACCN1***	**A/G**	**C**
	58552831	*ANXA5*	A/G	G
	17855555	*BRUNOL6*	A/G	G
	17851210	*C19orf36*	A/G	G
	3810950	*CHAT*	A/G	A
	1049367	*EBP*	A/G	G
	6053	***FGB***	**A/G**	**T**
	11554308	*KIAA0101*	A/G	A
	41285023	*PHACTR2*	A/G	G
	11551969	***RCL1***	**A/G**	**C**
	4238001	*SCARB1*	A/G**	G
	11555894	*SLC25A4*	A/G	A
	4904448	*SPATA7*	A/G	A
	34587547	*CALCA*	C/G	G
	10401800	*DHDH*	C/G	A
	17581133	***GCM2***	**C/G**	**C**
	5931	*LDLR*	C/G	A
	13412879	*LOC100131211*	C/G	G
	11554447	*RPLP1*	C/G	G
	11558171	*GOT2*	G/T	A
	36006195	*HBG1*	G/T	G
	12711	*PRDX4*	G/T	G
	409782	*RBM11*	G/T	G
	55674308	*SGSM3*	G/T	A

Note: sckSNPs with a weak context in their corresponding +4 or −3 position are in bold.

* These two sckSNPs are both equal to C/G in −3 and +4, respectively.

** Status of this gene is annotated with "PROVISIONAL"; however, this gene is described in many papers.

that rs689466 and rs20417 are indeed not in the same haplotype block, and the sckSNP (rs3134591) of *PTGS2* is in the same block with rs20417 but not rs689466 (Fig. 2B). Therefore, we concluded that the sckSNP of *PTGS2* could be the true risk factor for inflammation and gastric cancer but not asthma in European population.

Importantly, some other sckSNP-containing genes are oncogenes or tumor suppressor genes (Supplementary Table 4). Since the concentration of their protein products is reported to be involved in tumorigenesis, we suspected that sckSNPs in these genes could impact one's susceptibility to cancers.

Experimental confirmation on sckSNPs' impact on protein expression efficiency

Our above analyses have shown that sckSNPs could be a true risk factor for diseases by affecting protein expression level. To confirm this, two sckSNP-containing genes (*DBI* and *PTGS2*) were investigated to determine the influence of sckSNPs on translation efficiency. According to the previous reports, kSNP of *CD40* [7] and *APOA5* [21] was used as positive and negative control, respectively. The sequences around the first coding exon of *CD40*, *APOA5*, *DBI* and *PTGS2* were inserted in the pcDNA3.1(+)-GFP expression plasmid. As a result, about 3–15 amino acids of each gene will be translated and then fused to GFP. Different kSNP alleles of each gene were generated by Site-Directed Mutagenesis. 293T-cells transfected with these plasmids were harvested and subjected to western blotting. Consistent with previous reports, the C allele resulted in increased protein expression of *CD40* compared to the T alleles and there was no significant difference between A-*APOA5* and G-*APOA5*. Importantly, our results also showed that the sckSNPs in *DBI* and *PTGS2* could indeed significantly affect protein yield (Fig. 3A). Semi-quantitative RT-PCR results showed that there was no significant difference at the RNA expression level (Fig. 3B), indicating that the difference in protein expression occurred at the translational stage. The protein expression level was also measured by mean fluorescent intensity as determined by FACS analysis; similar results were obtained as shown in Fig. 3C. The mean fluorescent intensity of C-*CD40* was higher than that of T-*CD40,* but the difference was not statistically significant, with a *p*-value of 0.055 (Fig. 3C). This result is likely due to the fact that the fluorescent method that was used to quantify differences is less accurate than western blot quantization. These results confirmed that sckSNPs have a strong impact on protein translation efficiency.

Discussion

In our study, we screened Kozak-motif-located SNPs (kSNPs) and focused on strong-changed kSNPs (sckSNPs), which can strongly impact protein translation efficiency.

Some sckSNPs we screened out were reported to be directly disease-associated in genetic level, such as *CHAT* and *SCARB1*, but the biological mechanism behind these associations has not been explained. We suspected that changes in protein expression induced by sckSNPs could be the cause of those diseases. In the case of *DBI* and *CDK5RAP2*, the protein concentrations of these two sckSNP-containing genes were reported to be important risk factors for disease. Although not strong enough to induce disease per se, the change in protein concentration of *DBI* in the brain, which can be impacted by its sckSNP, has been reported to be correlated with high and low anxiety levels [22]. While in the case of *CDK5RAP2, CDK5RAP2*-knockdown cells results in increased resistance to paclitaxel and doxorubicin, which are chemotherapy drugs for cancers [23]. We hypothesize that differences in *CDK5RAP2* expression induced by sckSNP could impact the effect of those drugs on cancer treatment.

In our study, we found that some sckSNPs were located in the exact same haplotype block of the reported disease-associated SNPs. Therefore, we suspected that in these cases, sckSNPs could be the true risk factor. This hypothesis brings about the intriguing question of why these other position-located SNPs, instead of sckSNPs, were reported to be disease-associated. There are two possibilities. First, in the previous studies, researchers biased their investigations by focusing on non-synonymous SNPs, since these have a higher chance of disrupting protein function, such as rs8192506 in *DBI*. Second, when a SNP was used as the screening disease-associated marker, this SNP sites should be widely reported (those reported by Hapmap Project were commonly used). The disease-related allele of sckSNPs could be rare in healthy people but rich in patients because of their strong effects. As a result, they could not be used as the common screening marker.

Some factors may disrupt the haplotype analysis, such as allele frequency and populations. In many reported cases, not all SNPs in

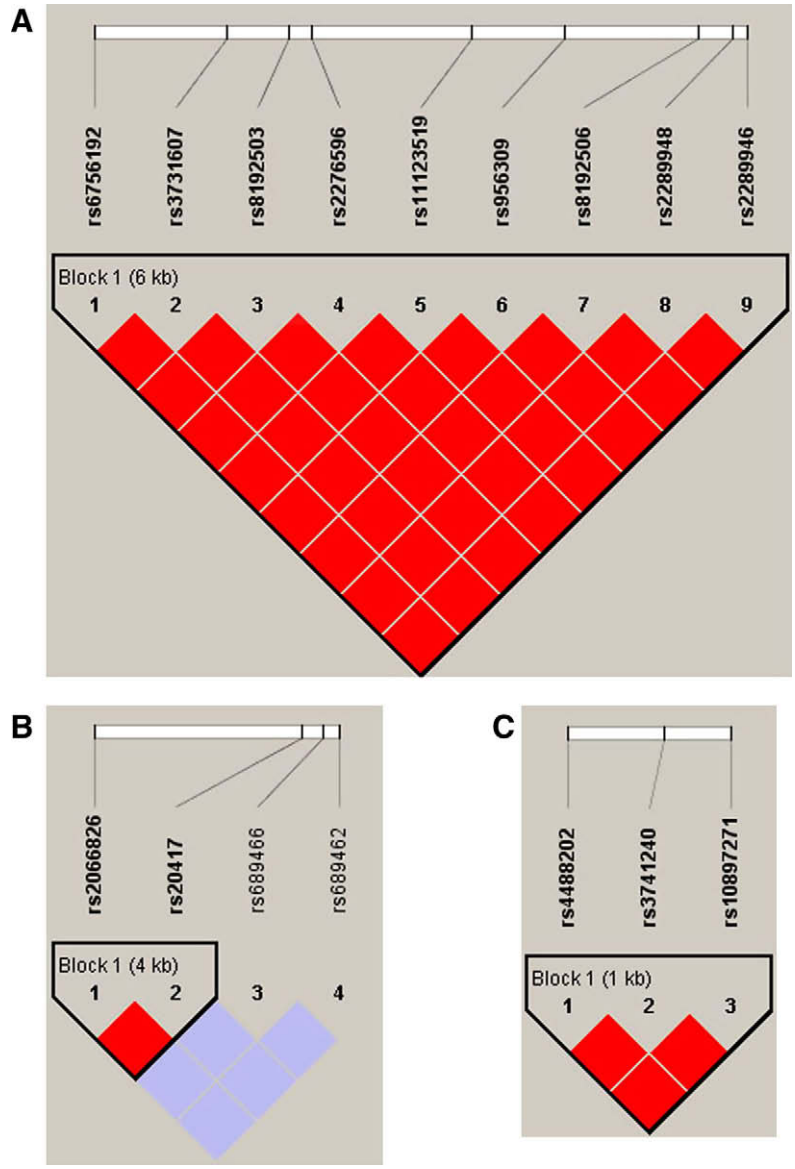

Fig. 2. Haplotype block analysis of sckSNP-containing regions in *DBI* (A), *PTGS2* (B) and *SCGB1A1* (C) with the European data from the Hapmap Project. Red and purple squares indicate two SNPs that are in tight and low LD, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

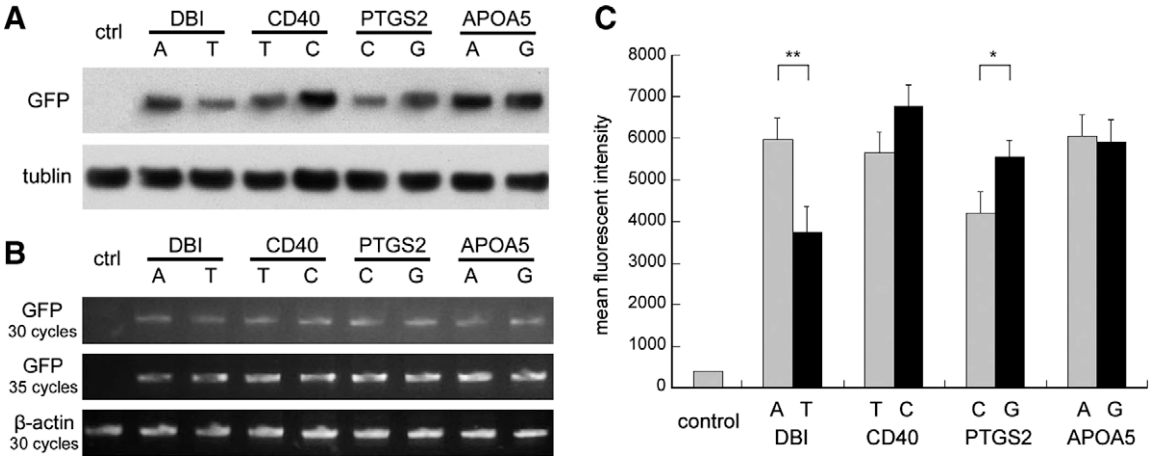

Fig. 3. Different kSNP-containing genes were transfected, and expression was detected at the protein (A) and RNA (B) level; mean fluorescent intensity was also detected by FACS analysis (C).

the same haplotype block were associated with disease at same time; one impact factor is the different frequencies of alleles at an SNP site. For example, in the study of age-related macular degeneration (AMD), all SNPs except rs7524776 in a block of *CFH* were reported to be strongly AMD-associated [24]. This was mainly because of the very low frequency of the C allele of rs7524776 (about 10% in a normal Caucasian population and 5% in a normal East Asian population according to the Hapmap Project database). Therefore, more evidence is needed to verity whether those sckSNPs in the same block of the reported disease-associated SNPs display a similar association. On the other hand, haplotype blocks are not always consistent in different populations, as observed for *PTGS2*. SNPs of *PTGS2* reported to be associated with asthma, inflammation, and gastric cancer were all analyzed in a white European population. However, after analyzing the information of African and East Asian populations in the Hapmap Project, differences in the haplotype blocks were revealed (Supplementary Fig. 1). Thus, the association of the SNPs of *PTGS2* with diseases may change in different populations.

Although we believed that the rest of the consensus sequence contributed only marginally if an AUG codon was flanked by R in position −3 and G in +4 [3], some exceptions were reported. *CD40* is an example that contains a G in both the −3 and +4 positions. The C/T kSNP at the −1 position of *CD40* was reported to be associated with Graves' disease, probably because its protein expression levels were affected by the kSNP [7]. In our study, we also found that the kSNP of *CD40* affected protein expression levels. However, this effect was to a much lesser extent, compared to the effect induced by the sckSNPs in *DBI* and *PTGS2*. Therefore, we considered that some *CD40*-like kSNPs could also be disease-associated.

To summarize, we listed all the reported human kSNPs and sSNPs, and provided a candidate explanation of how diseases are caused by kSNP-containing genes, via haplotype analysis. However, more evidence is needed to further confirm our speculations.

Acknowledgments

This work is supported by the National High Technology Research and Development Program of China (2006AA02Z330, 2006AA02A301, 2008ZX10001-001), the National Basic Research Program of China (No. 2007CB512202), the National Natural Science Foundation of China, Key Program (No. 30530450), and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX1-YW-R-74, KSCX1-YW-22).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2010.01.002.

References

[1] M. Kozak, An analysis of vertebrate mRNA sequences: intimations of translational control, J. Cell Biol. 115 (1991) 887–903.
[2] M. Kozak, Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems, Mol. Cell Biol. 9 (1989) 5073–5080.
[3] M. Kozak, Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6, Embo J. 16 (1997) 2482–2492.
[4] M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, Cell 44 (1986) 283–292.
[5] M. Kozak, Emerging links between initiation of translation and human diseases, Mamm. Genome 13 (2002) 401–410.
[6] Y. Tomer, E. Concepcion, D.A. Greenberg, A C/T single-nucleotide polymorphism in the region of the CD40 gene is associated with Graves' disease, Thyroid 12 (2002) 1129–1135.
[7] E.M. Jacobson, E. Concepcion, T. Oashi, Y. Tomer, A Graves' disease-associated Kozak sequence single-nucleotide polymorphism enhances the efficiency of CD40 gene translation: a case for translational pathophysiology, Endocrinology 146 (2005) 2684–2691.
[8] J.M. Maguire, A. Thakkinstian, J. Sturm, et al., Polymorphisms in platelet glycoprotein 1balpha and factor VII and risk of ischemic stroke: a meta-analysis, Stroke 39 (2008) 1710–1716.
[9] R.I. Baker, J. Eikelboom, E. Lofthouse, et al., Platelet glycoprotein 1balpha Kozak polymorphism is associated with an increased risk of ischemic stroke, Blood 98 (2001) 36–40.
[10] A. Ozturk, S.T. DeKosky, M.I. Kamboh, Genetic variation in the choline acetyltransferase (CHAT) gene may be associated with the risk of Alzheimer's disease, Neurobiol. Aging 27 (2006) 1440–1444.
[11] D. Osgood, D. Corella, S. Demissie, et al., Genetic variation at the scavenger receptor class B type I gene locus determines plasma lipoprotein concentrations and particle size and interacts with type 2 diabetes: the framingham study, J. Clin. Endocrinol. Metab. 88 (2003) 2869–2879.
[12] X. Wu, Y. Ye, L.A. Kiemeney, et al., Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer, Nat. Genet. (2009).
[13] H. Sakamoto, K. Yoshimura, N. Saeki, et al., Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer, Nat. Genet. 40 (2008) 730–740.
[14] T.A. Manolio, L.D. Brooks, F.S. Collins, A HapMap harvest of insights into the genetics of common disease, J. Clin. Invest. 118 (2008) 1590–1605.
[15] E. Fisher, I. Nitz, C. Gieger, et al., Association of acyl-CoA-binding protein (ACBP) single nucleotide polymorphisms and type 2 diabetes in two German study populations, Mol. Nutr. Food Res. 51 (2007) 178–184.
[16] E. Fisher, I. Nitz, I. Lindner, et al., Candidate gene association study of type 2 diabetes in a nested case-control study of the EPIC-Potsdam cohort–role of fat assimilation, Mol. Nutr. Food Res. 51 (2007) 185–191.
[17] C.K. Thoeringer, E.B. Binder, D. Salyakina, et al., Association of a Met88Val diazepam binding inhibitor (DBI) gene polymorphism and anxiety disorders with panic attacks, J. Psychiatr. Res. 41 (2007) 579–584.
[18] J. Shi, N.L. Misso, M.A. Kedda, et al., Cyclooxygenase-2 gene polymorphisms in an Australian population: association of the −1195G>A promoter polymorphism with mild asthma, Clin. Exp. Allergy 38 (2008) 913–920.
[19] D. Corella, J.I. Gonzalez, M. Bullo, et al., Polymorphisms cyclooxygenase-2–765G>C and interleukin-6–174G>C are associated with serum inflammation markers in a high cardiovascular risk population and do not modify the response to a Mediterranean diet supplemented with virgin olive oil or nuts, J. Nutr. 139 (2009) 128–134.
[20] R. Sitarz, R.J. Leguit, W.W. de Leng, et al., The COX-2 promoter polymorphism −765G>C is associated with early-onset conventional and stump gastric cancers, Mod. Pathol. 21 (2008) 685–690.
[21] P.J. Talmud, J. Palmen, W. Putt, et al., Determination of the functionality of common APOA5 polymorphisms, J. Biol. Chem. 280 (2005) 28215–28220.
[22] S.K. Sudakov, O.F. Medvedeva, I.V. Rusakova, et al., Differences in genetic predisposition to high anxiety in two inbred rat strains: role of substance P, diazepam binding inhibitor fragment and neuropeptide Y, Psychopharmacology (Berl) 154 (2001) 327–335.
[23] X. Zhang, D. Liu, S. Lv, H. Wang, et al., CDK5RAP2 is required for spindle checkpoint function, Cell Cycle 8 (2009) 1206–1216.
[24] A.E. Hughes, N. Orr, H. Esfandiary, M. Diaz-Torres, T. Goodship, U. Chakravarthy, A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration, Nat. Genet. 38 (2006) 1173–1177.