Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press



Alternative approach to a heavy weight problem

Amir Goren, Eddo Kim, Maayan Amit, et al.

Genome Res. published online December 20, 2007
Access the most recent version at doi:10.1101/gr.6661308

P<P Published online December 20, 2007 in advance of the print journal.

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © 2007, Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Letter

Alternative approach to a heavy weight problem

Amir Goren,[1,4] Eddo Kim,[1,4] Maayan Amit,[1,4] Ron Bochner,[1] Galit Lev-Maor,[1] Nadav Ahituv,[2,3] and Gil Ast[1,5]

[1]*Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv 69978, Israel;* [2]*Department of Biopharmaceutical Sciences, University of California, San Francisco, California 94143, USA;* [3]*Institute for Human Genetics, University of California, San Francisco, California 94143, USA*

Obesity is reaching epidemic proportions in developed countries and represents a significant risk factor for hypertension, heart disease, diabetes, and dyslipidemia. Splicing mutations constitute at least 14% of disease-causing mutations, thus implicating polymorphisms that affect splicing as likely candidates for disease susceptibility. A recent study suggested that genes associated with obesity were significantly enriched for rare nucleotide variants. Here, we examined these variants and revealed that they are located near splice junctions and tend to affect exonic splicing regulatory sequences. We also show that the majority of the exons that harbor these SNPs are constitutively spliced, yet they exhibit weak splice sites, typical to alternatively spliced exons, and are hence suboptimal for recognition by the splicing machinery and prone to become alternatively spliced. Using ex vivo assays, we tested a few representative variants and show that they indeed affect splicing by causing a shift from a constitutive to an alternative pattern, suggesting a possible link between extreme body mass index and abnormal splicing patterns.

[Supplemental material is available online at www.genome.org.]

The sequencing of the human genome (Lander et al. 2001) and the recent completion of the International HapMap Project (International HapMap Consortium 2003) significantly improved the ability to identify possible gene variants in different human individuals and to associate them with individual phenotypes and diseases (Crawford et al. 2005). The single nucleotide polymorphism (SNP) database is the most abundant source of genetic variation in the human genome (Sherry et al. 2001). Associating a SNP with a disease could improve the development of specific diagnostic tools and possibly provide the basis for direct clinical intervention.

Splicing is a process, by which introns are removed from precursor messenger RNA (pre-mRNA) transcripts and the protein coding elements assemble into mature mRNAs. Alternative splicing is a process by which more than one type of mRNA transcripts are produced from the same pre-mRNA (Black 2003). Alternative splicing increases transcriptomic diversity by adding a new regulatory dimension to the expression pattern of an organism (Graveley 2001; Ast 2004). Previous studies revealed the existence of exonic and intronic *cis*-acting regulatory sequences, which bind *trans*-acting factors and thus influence splice-site recognition and selection. These motifs are classified as exonic or intronic splicing regulatory sequences (ESR and ISR, respectively) and are required for the regulation of both constitutive and alternative splicing (Blencowe 2000; Graveley 2000). Binding of splicing regulatory proteins (such as SR proteins) to ESRs assists the placement of the spliceosome on the appropriate splice sites (Cartegni et al. 2002; Sanford et al. 2005).

Splicing mutations constitute at least 14% of disease-causing mutations (Krawczak et al. 1992; Cooper and Mattox 1997; Blencowe 2000; Fairbrother et al. 2004; Steiner et al. 2004). Hence, polymorphisms that affect splicing would be good candidates for association studies intended to identify genetic contributors to quantitative traits or diseases. Indeed, previous studies have identified a number of disease alleles, where a mutation in a conserved sequence, which performs as an exonic splicing enhancer, corresponds with splicing defects, such as in the *BRCA1* and *SMN* genes (Liu et al. 2001; Orban and Olah 2001; Cartegni and Krainer 2002; Kashima and Manley 2003; Chamary et al. 2006).

Obesity is becoming increasingly prevalent in developed countries and significantly elevates human disease and mortality risk factors (Bell et al. 2005). While environmental factors contribute to the increase in the percentage of obese individuals in the population, heredity is also a very strong component (Friedman 2003; Flier 2004). The most compelling evidence that obesity is genetically determined comes from twin, adoption, and family studies suggesting that hereditary factors could contribute to the development of obesity, with heritability estimates of 40%–70% (Comuzzie and Allison 1998; Friedman 2003). Hence, the majority of the variance in the incidence of obesity is thought to be attributed to genetic factors (Stunkard et al. 1986, 1990; Allison et al. 1996; Friedman 2004).

A recent medical sequencing study compiled a large data set of SNPs found in 757 individuals with extreme body mass index values (BMI). This data set contains 496 SNPs in the coding region (extreme BMI SNPs) of genes that are associated with obesity or that function in body-weight–related pathways (Ahituv et al. 2007). Interestingly, 38.1% of the SNPs are in positions that do not alter the protein coding sequence (189 synonymous SNPs), suggesting that these SNPs might affect other steps in gene expression apart from altering protein sequence. In this study, we undertook a large-scale analysis to examine the potential effect of both synonymous and nonsynonymous SNPs on mRNA splicing in multiple candidate genes associated with human extreme BMI susceptibility. Our results correlate a broad spectrum of SNPs with alterations in mRNA splicing.

[4]These authors contributed equally to this work.
[5]Corresponding author.
E-mail gilast@post.tau.ac.il; fax +972-3-640-9900.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.6661308.

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Results and Discussion

Genes that contain one or more SNPs can give rise to more than one allelic form of mRNA. It has been shown that, in addition to changes in the amino acid sequence, SNPs could also alter essential splicing elements, important for the processing or translation of human mRNA. Specifically, SNPs can be responsible for disruption of splice sites and splicing regulatory sequences, creation of different structural mRNA folds, and changes in the rate of protein synthesis due to skews in tRNA abundances (Shen et al. 1999; Chamary et al. 2006), which might, in turn, affect protein folding (Kimchi-Sarfaty et al. 2007). We were thus interested to study whether extreme BMI SNPs affect the mRNA splicing process of the corresponding exons. Therefore, we first set out to examine the spatial distribution of the extreme BMI SNPs along the exons and whether they are located within splicing regulatory motifs.

Synonymous extreme BMI SNPs are more prevalent near the splice sites

Ahituv et al. (2007) compiled a data set of SNPs that were found in extreme BMI individuals. From this data set we extracted 258 SNPs that appear within internal coding exons (see Methods), of which 98 are synonymous SNPs and 160 are nonsynonymous SNPs. For these 258 extreme BMI SNPs, we extracted the sequence of the corresponding exons, in which they resided. As a control data set, we extracted 64,509 SNPs (appearing in internal coding exons) from the UCSC genome browser (see Methods), of which 30,903 are synonymous and 33,606 are nonsynonymous SNPs. Next, we set out to examine the SNPs' spatial distribution along the exons. Previous studies indicated that ESR density is highest and SNP density is lowest near splice sites (Majewski and Ott 2002; Fairbrother et al. 2004). In fact, the increased density of regulatory motifs near splice sites forces SNPs that occur in exons near splice sites to be selected against because of their higher likelihood to disrupt the functionally important regulatory elements. To this end, we subdivided each of the UCSC SNP containing exons into quartiles and summed the number of synonymous and nonsynonymous SNPs in each quartile. As expected, the results revealed an unequal distribution of SNPs along the exons for both synonymous and nonsynonymous SNPs ($P = 2.49 \times 10^{-29}$ and $P = 1.6 \times 10^{-61}$, respectively), with lower density of SNPs near splice sites (Fig. 1A). Next, we set out to examine the positions of the 258 extreme BMI SNPs (98 synonymous and 160 nonsynonymous) along the exons. The results reveal that the density of synonymous, but not nonsynonymous SNPs, is significantly different from the one observed using the control SNPs group ($P = 1.03 \times 10^{-3}$; by χ^2 test, df = 3; Fig. 1B). Specifically, synonymous SNPs were found to be located near splice sites ($P = 2.44 \times 10^{-3}$, by χ^2 test, df = 1; Fig. 1B, left panel). Hence, surprisingly and opposed to previous results, the extreme BMI synonymous SNPs do not tend to be located in the middle of exons, as could be expected to prevent disruption of splicing regulatory elements. We were concerned that the difference we observed could be attributed to the fact that the extreme BMI SNPs are mostly rare SNPs (<1% minor allele frequency), whereas the UCSC SNPs are mostly common ones (>1% minor allele frequency). Therefore, we extracted another data set of 484 rare SNPs (<1% minor allele frequency) from the UCSC database (see Methods), of which 182 are synonymous SNPs and 302 are nonsynonymous ones, and repeated the above analysis. The results revealed that both the UCSC rare synonymous and nonsynonymous SNPs are less prevalent near the splice junctions ($P = 0.026$ and $P = 0.016$, respectively), which is similar to the observation for all UCSC SNPs (Fig. 1, cf. A and C), and is opposed to the observation for the extreme BMI SNPs (Fig. 1, cf. B and C). This suggests that these extreme BMI SNPs could possibly lead to the disruption of the splicing regulatory sequences.

Synonymous extreme BMI SNPs are located within splicing regulatory motifs

We next set out to analyze the prevalence of these extreme BMI SNPs within splicing regulatory motifs. To this end we searched

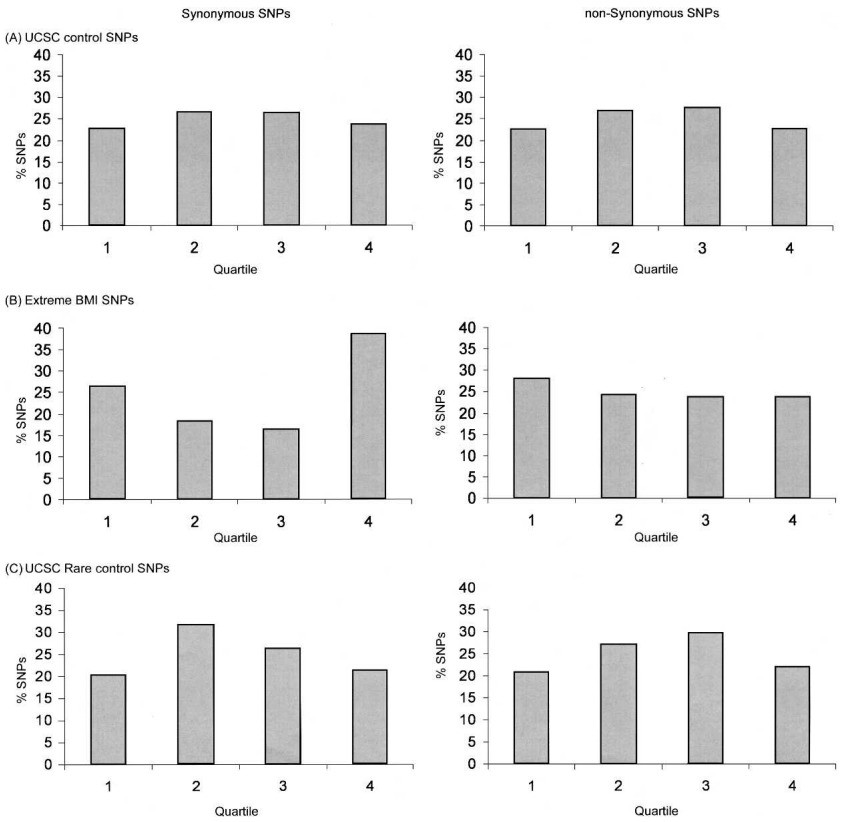


Figure 1. Distribution of extreme BMI SNPs and UCSC SNPs along exons. The distribution of SNPs was calculated for (*A*) UCSC SNP containing exons; for (*B*) extreme BMI SNP containing exons; and for (*C*) UCSC rare SNP containing exons. These calculations were performed for both synonymous (*left* panels) and nonsynonymous (*right* panels) SNPs. Each exon was divided into four quartiles (*X*-axis). The percentage of SNPs in each of the four quartiles is indicated (*Y*-axis).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

both groups of SNP-containing exons (extreme BMI and UCSC) for putative regulatory binding motifs based on the known matrices for the four SR proteins: SFRS1, SFRS5, SFRS6, and SFRS2, which are available on the ESEfinder server (Cartegni et al. 2003). We first calculated the number of ESEfinder putative motifs per number of coding nucleotides, and revealed that extreme BMI SNP containing exons do not differ significantly from UCSC SNP containing exons in this aspect. We next compared the extreme BMI SNPs with the UCSC SNPs to examine whether they are located within ESEfinder putative binding sites more frequently than could be expected. The results reveal that synonymous extreme BMI SNPs have a significant tendency to be located within the potential binding sites of these four SR proteins, compared with the UCSC SNPs ($P = 0.015$; by χ^2 test, df = 1), while nonsynonymous SNPs do not. This is in contrast to previous results indicating that synonymous sites in putative regulatory sequences evolve more slowly than the remaining exonic sequence (Parmley et al. 2006). We next repeated the above analysis, this time for each of the four splicing regulatory sequences, separately. The results reveal that the prevalence of synonymous SNPs within the putative binding sites of SFRS1 is significantly higher, compared with UCSC SNPs ($P = 6.23 \times 10^{-3}$; by χ^2 test, df = 1), but not within the putative binding sites of SFRS5, SFRS6, or SFRS2. For comparison, we examined the rare UCSC SNPs. The results revealed that these rare UCSC SNPs do not have a significant tendency to be located within the potential binding sites of SR proteins ($P = 0.49$). Testing each of the four potential binding sites separately yielded insignificant results as well. These results suggest that the synonymous extreme BMI SNPs may affect splicing regulatory elements and could potentially disrupt the splicing pattern.

Exons containing extreme BMI SNPs exhibit weak splice sites

Previous studies have shown that alternatively spliced exons have weaker splice signals compared to constitutively spliced ones (Itoh et al. 2004; Zheng et al. 2005). These differences presumably indicate that alternatively spliced exons are suboptimal for detection by the splicing machinery (Ast 2004). In addition, orthologous human-mouse alternatively spliced exons were found to be more conserved than constitutively spliced ones (Sorek et al. 2004), particularly in the wobble position (Goren et al. 2006), indicating a higher selective pressure to conserve regulatory sequences. Hence, the role of splicing regulatory sequences in assisting the splicing machinery to recognize the appropriate exon–intron junctions is more essential in alternatively spliced exons. Indeed, recent results found that a constitutive exon with strong splice sites was not dependent on the presence of splicing regulatory elements for its constitutive recognition, while the weakening of the splice sites, which shifted the splicing pattern to alternative, resulted in its reliance on these elements for proper splicing (Lev-Maor et al. 2007). These findings led us to investigate the correlation between the extreme BMI SNPs and the splice sites' strength. Specifically, we were interested to find whether the exons that harbor these extreme BMI SNPs have weak splice sites, suggesting that SNPs that alter splicing regulatory sequences might directly affect their splicing pattern. We examined the splicing pattern of the 130 SNP-containing exons (all in the coding sequence), and based on EST data from UCSC we identified 113 that are constitutively spliced. We then compared the splice donor and acceptor sites of the constitutively spliced SNP-containing exons with splice sites flanking constitutively and alternatively spliced exons (both data sets were extracted from Goren et al. [2006]). The results revealed that the splice sites of exons that harbor synonymous extreme BMI SNPs exhibit significantly weaker splice sites than constitutive exons ($P = 0.033$, and $P = 0.038$, Mann–Whitney test, for the splice donor and splice acceptor sites, respectively). These results also show that exons that harbor nonsynonymous extreme BMI SNPs exhibit a significantly weaker splice acceptor site, compared with constitutively spliced exons ($P = 2.04 \times 10^{-4}$, Mann–Whitney test), but not the splice donor site. The splice sites' strengths of exons that harbor extreme BMI SNPs and those of alternatively spliced exons were not found to be significantly different, for both synonymous and nonsynonymous SNPs containing exons (Fig. 2).

Most of the exons that harbor the extreme BMI SNPs are constitutively spliced. This finding, together with the finding that exons that harbor the extreme BMI SNPs have weaker splice sites compared to constitutively spliced exons, implies that these exons are suboptimal for recognition by the splicing machinery and hence are predisposed to become alternatively spliced due to alterations of splicing regulatory elements. Thus, the extreme BMI SNPs have a higher probability to shift these exons' splicing pattern from constitutive to alternative splicing. It is noteworthy that the definition of the extreme BMI exons as being constitutively spliced relies on the EST sequences that support them. Hence, there is a possibility that some of these exons are actually alternatively spliced (in certain tissues or conditions) but lack EST sequences to support it. However, even if this is the case, this does not change the fact that mutations in exonic splicing regulatory sequences could change the splicing pattern, since alternative exons were shown to be even more dependent on splicing regulatory sequences, which also determine the inclusion level of the exon (Goren et al. 2006).

Different constraints are acting on different positions of ESEfinder motifs

ESEfinder splicing regulatory motifs are short sequences, 6–8 bp long, that serve as binding sites for four SR proteins. However, the relative effect of each of the nucleotide positions on the binding affinity of the cognate SR protein, and hence on the splicing pattern, is not well understood. The above results suggest that the synonymous BMI SNPs, in particular, are involved in the modification of exonic splicing regulatory motifs. Thus, we set

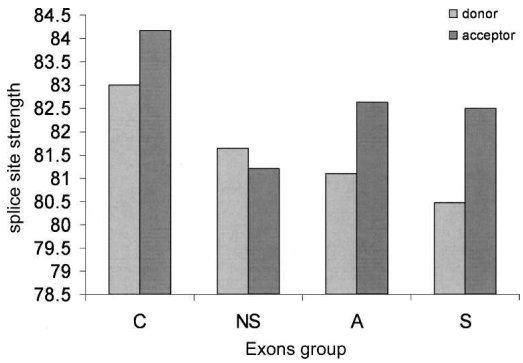


Figure 2. Exons containing extreme BMI SNPs exhibit weak splice sites. Splice acceptor (dark gray) and donor (light gray) site strengths were calculated based on the algorithm by Shapiro and Senapathy (1987) (*Y*-axis) for constitutive exons (C), alternative exons (A), and exons that harbor synonymous (S) and nonsynonymous (NS) extreme BMI SNPs (*X*-axis).

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

out to examine whether specific positions within these motifs are "preferred" by the SNPs more than others, suggesting they are key positions. To this end, we examined the positions of the synonymous SNPs within the ESEfinder splicing regulatory sequences, and compared them to a random background model (Fig. 3, light gray line). These results revealed that in the examined SFRS1, SFRS5, and SFRS6 potential binding sites there is an unequal distribution of SNPs among the positions ($P = 7.05 \times 10^{-3}$, $P = 2.36 \times 10^{-3}$, $P = 4.31 \times 10^{-4}$, respectively, by χ^2 test, df = motif length − 1). Further examination revealed that in each of the above three motifs there is one position that exhibits overabundance of SNPs, and one position that is underrepresented and thus probably selected against the presence of SNPs (Fig. 3). Specifically, in the SFRS1 sequence motif, there are more SNPs than expected in position 5 and less in position 7 ($P = 7.02 \times 10^{-3}$ and $P = 2.74 \times 10^{-3}$, respectively, by χ^2 test, df = 1; Fig. 3A, marked by circles). In the SFRS5 sequence motif, there are more SNPs than expected in position 7 and less in position 2 ($P = 8.29 \times 10^{-4}$ and $P = 5.75 \times 10^{-3}$, respectively, by χ^2 test, df = 1; Fig. 3B, circles). In the SFRS6 sequence motif, there are more SNPs than expected in position 3 and less in position 5 ($P = 5.73 \times 10^{-5}$ and $P = 0.012$, respectively, by χ^2 test, df = 1; Fig. 3C, circles; see Supplemental material for more information on the SNPs within the ESEfinder potential binding sites). To further examine constraints that act on different positions of these putative binding sites we set to compare the extreme BMI SNPs with another background model. We therefore used the data set of rare SNPs that were extracted from UCSC. This background model might represent the more "natural" occurrence of SNPs along the regulatory sequences (Fig. 3, black line). To this end, we compared the distribution of the extreme BMI synonymous SNPs to the synonymous UCSC rare SNPs and revealed somewhat different results. We found that in the examined SFRS1, SFRS5, and SFRS2 potential binding sites there is an unequal distribution of SNPs among the positions ($P = 1.20 \times 10^{-5}$, $P = 1.21 \times 10^{-4}$, $P = 4.00 \times 10^{-11}$, respectively, by χ^2 test, df = motif length − 1). We further revealed that in each of the above three motifs there is one position that exhibits overabundance of SNPs and one position that is underrepresented (Fig. 3). Specifically, in the SFRS1 sequence motif, there are more SNPs than expected in position 4, and less in position 2 ($P = 0.015$ and $P = 0.029$, respectively, by χ^2 test, df = 1; Fig. 3A, squares). Interestingly, in the SFRS5 sequence motif, there are more SNPs than expected in position 7, and less in position 2, similar to what we found using the random background model ($P = 0.022$ and $P = 1.40 \times 10^{-3}$, respectively, by χ^2 test, df = 1; Fig. 3B, squares). In the SFRS2 sequence motif, there are more SNPs than expected in position 2, and less in position 4 ($P = 1.11 \times 10^{-3}$ and $P = 0.011$, respectively, by χ^2 test, df = 1; Fig. 3C, squares). These findings suggest the existence of different regulatory constraints acting on different positions along the SR binding sites. It is worth noting that the unequal distribution of SNPs was also found in potential binding sites that did not exhibit a statistically significant higher prevalence of SNPs (potential binding sites of SFRS5 and SFRS6), implying that the lack of significance might be a result of insufficient data.

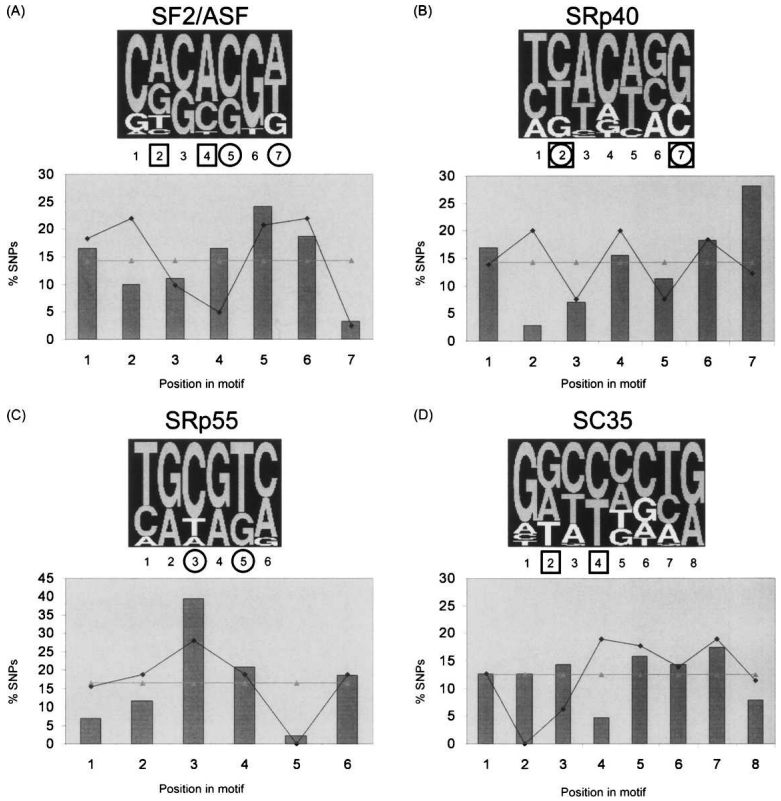


Figure 3. Extreme BMI synonymous SNPs are located in specific positions within ESEfinder motifs. The distribution of the synonymous extreme BMI SNPs within motifs of (*A*) SFRS1, (*B*) SFRS5, (*C*) SFRS6, and (*D*) SFRS2 is plotted. For each panel, the percentage of SNPs (*Y*-axis) in each position (*X*-axis) is indicated. For comparison, the light gray and black lines indicate the expected percentages for random data and the percentages of the rare UCSC SNPs, respectively. The pictogram was adopted from Cartegni et al. (2003) and reflects the consensus motif sequence. Positions with a significant percentage of SNPs, compared with the random data or with the rare UCSC SNPs, are marked in circles or squares, respectively.

Experimental validation

The results of the bioinformatic analyses suggested a link between extreme BMI SNPs and alterations of splicing patterns. We thus set out to examine whether the extreme BMI SNPs actually affect the splicing pattern ex vivo. Exon 4 of *PRKAG3* gene (Refseq NM_017431) was selected for biological examination because it contains a high number of SNPs: nine SNPs, four of which are synonymous SNPs, and the other five are nonsynonymous SNPs (see Methods and Supplemental material). Moreover, its length (404 bp) and its weak splice acceptor site score (77.76; Shapiro and Senapathy 1987) are suboptimal for recognition by the splicing machinery, and hence this exon may be prone to splicing pattern alterations as a result of a SNP. This exon, which is constitutively spliced according to current EST data, was cloned from the human genome along with the flanking introns and exons (a minigene encompassing exons 3 to 5). The minigene was transfected into

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

293T cells, total cytoplasmic RNA was extracted, and splicing products were separated on a 1.5% agarose gel after reverse transcription–polymerase chain reaction (RT-PCR; see Methods). The results revealed that in three out of the four synonymous SNPs, the splicing pattern shifted from constitutive to alternative splicing as a result of introducing the SNP (Fig. 4A, cf. lanes 2, 4, and 5 with lane 3). Moreover, although effects of nonsynonymous SNPs are usually attributed to a change in protein sequence, the results show that one (out of five) of the nonsynonymous SNPs also causes a shift from constitutive to alternative splicing (Fig. 4A, cf. lane 6 with lanes 7–10). The fact that nonsynonymous SNPs affect the protein sequence, while synonymous SNPs do not, correlates with the result that three out of four synonymous SNPs affect the splicing pattern, while only one out of the five nonsynonymous SNPs do so. Since this exon is not symmetrical, the exon-skipping isoform will contain a premature termination codon and potentially reduce the level of the normal protein within the cells. These results suggest that extreme BMI SNPs in exon 4 of the *PRKAG3* gene shift the splicing pattern from constitutive to alternative.

We next examined exon 3 of the *LIPC* gene, which is constitutively spliced according to current EST data but exhibits a very weak splice donor site score (68.57; Shapiro and Senapathy 1987). This exon, which contains a synonymous SNP and a nonsynonymous SNP (see Supplemental material), was cloned from the human genome along with the flanking introns and exons (a minigene encompassing exons 2 to 4; see Methods). The results revealed that both the synonymous SNP and the nonsynonymous SNP shifted the splicing pattern of the exon from constitutive to alternative splicing (Fig. 4B, cf. lanes 2 and 3 with lane 1).

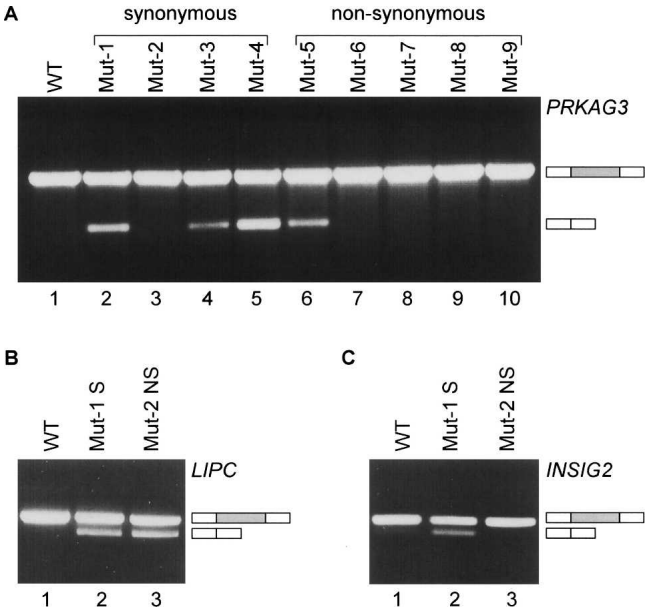


Figure 4. Extreme BMI SNPs affect the splicing pattern. Plasmid mutants were introduced into cells by transfection. Total cytoplasmic RNA was extracted, and splicing products were separated on a 1.5% agarose gel after reverse transcription–polymerase chain reaction (RT-PCR). (*A*) Lane *1*, wild-type (WT) *PRKAG3*; lanes *2–5*, synonymous mutations; lanes *6–10*, nonsynonymous mutations. (*B*) Lane *1*, WT *LIPC*; lane *2*, synonymous mutation; lane *3*, nonsynonymous mutation. (*C*) Lane *1*, WT *INSIG2*; lane *2*, synonymous mutation; lane *3*, nonsynonymous mutation. The two minigene mRNA products are shown on the *right* of each panel; empty and filled boxes define constitutive and alternative exons, respectively.

Next, we examined exon 5 of the *INSIG2* gene. Current available EST data suggest that this exon is alternatively spliced, with a very low skipping level. This exon, which contains a synonymous SNP and a nonsynonymous SNP (see Supplemental material), was cloned from the human genome along with the flanking introns and exons (a minigene encompassing exons 4 to 6; see Methods). The results revealed that the wild type exon is indeed alternatively spliced with low skipping levels, as was observed from EST data. Introducing the specific mutations revealed that both SNPs change the splicing pattern of the exon. The nonsynonymous SNP decreased the skipping level of the exon, while the synonymous SNP significantly increased its skipping level (Fig. 4C, cf. lanes 2 and 3 with lane 1). Overall, these results suggest that several of these extreme BMI SNPs could be functionally related to splicing abnormalities

In conclusion, in this study we investigated SNPs found in extreme BMI individuals. We revealed that most of the exons that harbor these SNPs are constitutively spliced, yet exhibit weak splice sites, typical to alternatively spliced exons, and are hence suboptimal for recognition by the splicing machinery and prone to become alternatively spliced. Moreover, we showed that these extreme BMI SNPs are located near the splice junctions, and specifically within putative splicing regulatory sequences, and hence could potentially affect the splicing activity. We then tested a few of these SNPs experimentally and demonstrated ex vivo that they affected splicing by shifting the splicing pattern, mainly from constitutive to alternative. This shift might interrupt the normal gene function by reducing the level of normal mRNA in the cells beneath the minimal required amount or by introducing a new undesirable transcript variant, which might act in a dominant negative manner. Finally, we would like to note, that the results presented in this paper, which link SNPs found in extreme BMI individuals and aberrant splicing, may reflect a broader scenario. It may be that deep resequencing of a random set of individuals would result in a similar finding, due to the ability of deep resequencing to detect more deleterious mutations as compared with resequencing of few chromosomes, as was the case in the majority of the UCSC SNPs that we compared. This possibility could be tested by evaluating more deep resequencing efforts or by examination of a similar population with normal BMI for the same exact genes, both of which are not currently available. If this is indeed the case, it might indicate that the role of splicing in shaping the phenotypic differences between individuals could be quite substantial.

Methods

Extracting data sets of extreme BMI and UCSC SNPs

The 496 extreme BMI SNPs (189 synonymous and 307 nonsynonymous) were extracted from Ahituv et al. (2007). In Figure 1, since we were interested in studying the potential effect of these SNPs on splicing, and to prevent from introducing a bias into the data set, we removed SNPs located in first/last exons, as well as SNPs located in exons containing UTRs (resulted in 98 synonymous and 160 nonsynonymous SNPs). UCSC SNPs were downloaded using the SNPs track, version hg17, table snp125 (Karolchik et al. 2004). Similarly, we discarded UCSC SNPs located in first/last exons, as well as SNPs located in exons containing UTRs (resulted in 30,903 synonymous and 33,606 nonsynonymous SNPs). We then also created a data set of rare UCSC SNPs (version

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

hg17, table snp125), exhibiting allele frequency of less than 1% and that do not reside in first/last exons or in exons that contain UTRs (in total 182 synonymous SNPs and 302 nonsynonymous SNPs).

Extreme BMI SNPs are located within splicing regulatory motifs

Extreme BMI SNP-containing exons and UCSC SNP-containing exons were analyzed and splicing regulatory sequences were marked for SFRS6, SFRS5, SFRS2, and SF2, using ESEfinder with default parameters. The number of nucleotides covered by ESEfinder motifs versus the number of nucleotides not covered was calculated to identify whether there is a significant difference in the number of ESEfinder putative motifs per number of coding nucleotides between extreme BMI and UCSC SNP-containing exons. Next, the number of SNPs that are located in ESEfinder motifs was calculated accordingly.

Construction of the minigenes

A human cytomegalovirus (CMV) immediate early promoter in location 1–589 of pEGFP-N3 vector (BD Biosciences Clontech) was replaced with a 250-nt promoter of human *SLU7* (Alberstein et al. 2007), using the following primers: 5′-CTTGAAACTC CCCACAAGCC-3′ and 5′-CCAGCCCCGCCGCAGCTAGCC-3′. Next, a genomic DNA of 2.8 kb comprising exons 3 through 5 of the PRKAG3 gene (AMP-activated protein kinase, noncatalytic; NM_017431) was cloned into EcoRI and BamHI sites with the following primers: 5′-AAAGAATTCCTAACCTCTTCCCC CACCTC-3′ and 5′-AAAGGATCCCAAAGCTCTGCTTCTTGCTG-3′, respectively. Similarly, a genomic DNA of 4.2 kb comprising exon 2 through 4 of the *LIPC* gene (lipase C precursor; NM_000236) was cloned into XhoI and BamHI sites of a similar vector with the following primers: 5′-AAAAAACTCGAGA CAAAACGCTGCATGAGATG -3′ and 5′-AAAGGATCCACCCAG GCTGTACCCAATTA-3′, respectively. In the same way, a genomic DNA of 2.8 kb comprising exons 4 through 6 of the *INSIG2* gene (insulin-induced protein 2; NM_016133) was cloned into EcoRI and BamHI sites of the modified pEGFP-N3 vector using the following primers: 5′-AAAGAATTCTCTCTCACACTGGCTG CACT-3′ and 5′-AAAGGATCCATCTTGCCCTTGTGGATCAG-3′, respectively.

Mutating the reporting exon

Nine positions in exon 4 of *PRKAG3* minigene were mutated, either by DpnI-mediated site-directed mutagenesis (Stratagene) or by substitution with a primer tail followed by ligation. The four synonymous mutations (mutations 1 through 4) are as follows: Mut-1 (32C→T), 5′-TGGGCTGGAGGCCACATTCC-3′ and 5′-GTGGACTCAGCAGCTGGCCT-3′; Mut-2 (308C→T), 5′-GAA CTGCGGAAACCTGGCGCCCAGATCTAC-3′ and 5′-GTAGA TCTGGGCGCCAGGTTTCCGCAGTTC-3′; Mut-3 (311C→T), 5′-CTGCGGAAACCCGGTGCCCAGATCTACATG-3′ and 5′-CATGTAGATCTGGGCACCGGGTTTCCGCAG-3′; Mut-4 (353C→T), 5′-GAGCACACCTGCTATGATGCCATGGCAAC-3′ and 5′-GTTGCCATGGCATCATAGCAGGTGTGCTC-3′. The five nonsynonymous mutations (mutations 5 through 9) are as follows: Mut-5 (79A→G), 5′-GTCCTGCCGGGGTGGGCACT-3′ and 5′-CAGCTTGAGCCAAGGGTGTG-3′; Mut-6 (228C→G), 5′-CT AGAAGGCCTGGTGGAAGAGAGGC-3′ and 5′-GCCTCTCTTCC ACCAGGCCTTCTAG-3′; Mut-7 (253T→C), 5′-CGTCCCCGCAG GCCCCATTT-3′ and 5′-GGCACAGGGCAGGCCTCTCT-3′; Mut-8 (361T→C), 5′-GCTACGATGCCACGGCAACTAGCTC-3′ and 5′-GAGCTAGTTGCCGTGGCATCGTAGC-3′; Mut-9 (402G→C), 5′-CAGGTGAGGCCACGGCTCTG-3′ and 5′-CAGCATGGTGTCGAAGATGAC-3′; Similarly, two positions in exon 3 of LIPC minigene were mutated. The synonymous mutation is as follows: Mut-1 (159C→T), 5′-TGCGGCTCTTCTCCG GTG-3′ and 5′-ACCTCCTTGCCCACAAGGCG-3′; and the nonsynonymous mutation is as follows: Mut-2 (10G→A), 5′-CTCAGGTGGACGGCATGCTAGAAAACTG-3′ and 5′-CAGT TTTCTAGCATGCCGTCCACCTGAG-3′. Two positions in exon 5 of the *INSIG2* minigene were mutated. The synonymous mutation is as follows: Mut-1 (91A→G), 5′-GAAACATTGGTCGA CAGCTGGCAATGGTAAGC-3′ and 5′-GCTTACCATTGCCAGCT GTCGACCAATGTTTC-3′; and the nonsynonymous mutation is as follows: Mut-2 (66T→C), 5′-CAACAATGGGAAACATTGGTCG-3′ and 5′-TGCCTCCAGCAAAAAATATACATGG-3′. All plasmids were confirmed by sequencing.

Transfection, RNA isolation, and RT-PCR amplification

The 293T and U2OS cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 4.5 g/mL glucose (Renium) and 10% fetal calf serum (Biological Industries). Cells were grown to 50% confluence in a 10-cm culture dish, under standard conditions, at 37°C with 5% CO_2. Cells were split at a 1:8 ratio to 6-well plates 24 h prior to transfection. Transfection was performed using 3 µL of FuGENE6 (Roche) with 1 µg of plasmid DNA. Cells were harvested 48 h after transfection. *PRKAG3* and *LIPC* experiments were performed on 293T cells, and the *INSIG2* experiment was performed on U2OS cells. Total cytoplasmic RNA was extracted using TriReagent (Sigma), followed by treatment with 2 U of RNase-free DNase (Ambion). Reverse transcription (RT) was performed on 2 µg of total cytoplasmic RNA for 1 h at 42°C, using an oligo(dT) and 2 U of reverse transcriptase of avian myeloblastosis virus (RT–AMV, Roche). The spliced cDNA products derived from the expressed minigenes were detected by PCR, using a forward primer which is specific to exon 3 of *PRKAG3* gene or to exon 2 of *LIPC* gene and a reverse primer specific to the pEGFP-N3 expression vector; *PRKAG3* forward primer, 5′-TGTGACCAGCAGCTCAGAAA-3′; *LIPC* forward primer, 5′-ACCCAGGCTGTACCCAATTA-3′, and pEGFP-N3 reverse primer, 5′-GAACTTCAGGGTCAGCTTGC-3′. For exon 5 of the *INSIG2* minigene, pEGFP-N3 forward primer, 5′-TAGCG CTACCGGACTCAGAT-3′, and *INSIG2* reverse primer, 5′-TG CATCCTGGCAATTTTACA-3′, were used. Amplification was performed for 30 cycles, consisting of 94°C for 30 sec, 60°C for 45 sec, and 72°C for 1 min, using ReadyMix (Bio-Lab). The products were resolved on a 1.5% agarose gel. PCR products were eluted from gel and confirmed by DNA sequencing after purification (Wizard, Promega).

Acknowledgments

This work was supported by a grant from the Israel Science Foundation (1449/04 and 40/05), MOP Germany–Israel, the German-Israeli Foundation, the Deutsch-Israel Project, the Israel Cancer Association through the Ber–Lehmsdorf Memorial Fund, and EURASNET. A.G. is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, and E.K. is a fellow of the Clore Scholars Programme.

References

Ahituv, N., Kavaslar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hebert, S., Doelle, H., Ersoy, B., Kryukov, G., Schmidt, S., et al. 2007. Medical sequencing at the extremes of human body mass. *Am. J. Hum. Genet.* **80:** 779–791.

Alberstein, M., Amit, M., Vaknin, K., O'Donnell, A., Farhy, C., Lerenthal, Y., Shomron, N., Shaham, O., Sharrocks, A.D.,

Downloaded from genome.cshlp.org on November 18, 2026 . Published by Cold Spring Harbor Laboratory Press

Ashery-Padan, R., et al. 2007. Regulation of transcription of the RNA splicing factor hSlu7 by Elk-1 and Sp1 affects alternative splicing. *RNA* **13:** 1988–1999.
Allison, D.B., Kaprio, J., Korkeila, M., Koskenvuo, M., Neale, M.C., and Hayakawa, K. 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int. J. Obes. Relat. Metab. Disord.* **20:** 501–506.
Ast, G. 2004. How did alternative splicing evolve? *Nat. Rev. Genet.* **5:** 773–782.
Bell, C.G., Walley, A.J., and Froguel, P. 2005. The genetics of human obesity. *Nat. Rev. Genet.* **6:** 221–234.
Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72:** 291–336.
Blencowe, B.J. 2000. Exonic splicing enhancers: Mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25:** 106–110.
Cartegni, L. and Krainer, A.R. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* **30:** 377–384.
Cartegni, L., Chew, S.L., and Krainer, A.R. 2002. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3:** 285–298.
Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31:** 3568–3571.
Chamary, J.V., Parmley, J.L., and Hurst, L.D. 2006. Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7:** 98–108.
Comuzzie, A.G. and Allison, D.B. 1998. The search for human obesity genes. *Science* **280:** 1374–1377.
Cooper, T.A. and Mattox, W. 1997. The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.* **61:** 259–266.
Crawford, D.C., Akey, D.T., and Nickerson, D.A. 2005. The patterns of natural variation in human genes. *Annu. Rev. Genomics Hum. Genet.* **6:** 287–312.
Fairbrother, W.G., Holste, D., Burge, C.B., and Sharp, P.A. 2004. Single nucleotide polymorphism–based validation of exonic splicing enhancers. *PLoS Biol.* **2:** E268. doi: 10.1371/journal.pbio.0020268.
Flier, J.S. 2004. Obesity wars: Molecular progress confronts an expanding epidemic. *Cell* **116:** 337–350.
Friedman, J.M. 2003. A war on obesity, not the obese. *Science* **299:** 856–858.
Friedman, J.M. 2004. Modern science versus the stigma of obesity. *Nat. Med.* **10:** 563–569.
Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T., and Ast, G. 2006. Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Mol. Cell* **22:** 769–781.
Graveley, B.R. 2000. Sorting out the complexity of SR protein functions. *RNA* **6:** 1197–1211.
Graveley, B.R. 2001. Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* **17:** 100–107.
International HapMap Consortium. 2003. The International HapMap Project. *Nature* **426:** 789–796.
Itoh, H., Washio, T., and Tomita, M. 2004. Computational comparative analyses of alternative splicing regulation using full-length cDNA of various eukaryotes. *RNA* **10:** 1005–1018.
Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32:** D493–D496.
Kashima, T. and Manley, J.L. 2003. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* **34:** 460–463.
Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315:** 525–528.
Krawczak, M., Reiss, J., and Cooper, D.N. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum. Genet.* **90:** 41–54.
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
Lev-Maor, G., Goren, A., Sela, N., Kim, E., Kerem, H., Doron-Faigenboim, A., Leibman-Barak, S., Pupko, T., and Ast, G. 2007. The "alternative" choice of constitutive exons throughout evolution. *PLoS Genet.* **3:** e203. doi: 10.1371/journal.pgen.0030203.
Liu, H.X., Cartegni, L., Zhang, M.Q., and Krainer, A.R. 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat. Genet.* **27:** 55–58.
Majewski, J. and Ott, J. 2002. Distribution and characterization of regulatory elements in the human genome. *Genome Res.* **12:** 1827–1836.
Orban, T.I. and Olah, E. 2001. Purifying selection on silent sites—A constraint from splicing regulation? *Trends Genet.* **17:** 252–253.
Parmley, J.L., Chamary, J.V., and Hurst, L.D. 2006. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* **23:** 301–309.
Sanford, J.R., Ellis, J., and Caceres, J.F. 2005. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem. Soc. Trans.* **33:** 443–446.
Shapiro, M.B. and Senapathy, P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15:** 7155–7174.
Shen, L.X., Basilion, J.P., and Stanton Jr., V.P. 1999. Single-nucleotide polymorphisms can cause different structural folds of mRNA. *Proc. Natl. Acad. Sci.* **96:** 7871–7876.
Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. 2001. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29:** 308–311.
Sorek, R., Shemesh, R., Cohen, Y., Basechess, O., Ast, G., and Shamir, R. 2004. A non-EST–based method for exon-skipping prediction. *Genome Res.* **14:** 1617–1623.
Steiner, B., Truninger, K., Sanz, J., Schaller, A., and Gallati, S. 2004. The role of common single-nucleotide polymorphisms on exon 9 and exon 12 skipping in nonmutated CFTR alleles. *Hum. Mutat.* **24:** 120–129.
Stunkard, A.J., Foch, T.T., and Hrubec, Z. 1986. A twin study of human obesity. *JAMA* **256:** 51–54.
Stunkard, A.J., Harris, J.R., Pedersen, N.L., and McClearn, G.E. 1990. The body-mass index of twins who have been reared apart. *N. Engl. J. Med.* **322:** 1483–1487.
Zheng, C.L., Fu, X.D., and Gribskov, M. 2005. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA* **11:** 1777–1787.

Received May 1, 2007; accepted in revised form November 14, 2007.