Premature termination codon rs67047829 in ERV3-1 may be protective against obesity

Jeremy S.C. Clark (Jeremy.Clark@pum.edu.pl)
Pomeranian Medical University in Szczecin
Konrad Podsiadło
Pomeranian Medical University in Szczecin
Marta Sobalska-Kwapis
University of Lodz
Błażej Marciniak
University of Lodz
Kamila Rydzewska
Pomeranian Medical University in Szczecin
Andrzej Ciechanowicz
Pomeranian Medical University in Szczecin
Thierry Wetering
Pomeranian Medical University in Szczecin
Dominik Strapagiel
University of Lodz

Article

Keywords: body mass index, HERV, pretermination codon, RNA expression, stop codon

Posted Date: April 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1509014/v1

Abstract

Premature termination codons are usually single nucleotide polymorphisms which can have high, but ethnically-dependent, prevalence in a population. The aim of this fishing study was to assess all such premature termination codons (n=140) from a database (with 551915 exonic variants; HumanCoreExome-24 v1.0 and v1.1 BeadChip) with association with obesity in a large sample (n=5757) of the Polish population. Statistical analyses were conducted using regression models with body mass index (BMI) as a continuous variable. Interactions with year of birth and sex; and recessive, dominant and heterozygote models for males and females together and then separately were assessed. A significant association, which was interaction independent, was found between the *ERV3-1*: rs67047829 heterozygote model and BMI with both sexes together (p-value adjusted for year of birth and sex: 0.001; BMI for AA, GA, GG: means (standard deviations): 22.9 (3.68), 24.7 (4.44), 24.9 (4.59)). Corresponding values for males, but not for females, were also significant: (p=0.003; AA 22.2 (3.96), GA 23.8 (4.44), GG 24.3 (4.96)). As rs67047829 produces a stop codon in the Endogenous retrovirus group 3 member 1, envelope protein, the mRNA of which is highly expressed in adipose tissue, it is possible this premature termination codon provides protective effect against obesity.

Introduction

A premature termination codon (PTC) can result from a change which gives either of the mRNA translation stop codons UGA, UAG or UAA found in the standard genetic code ("premature termination" here being synonymous with "stop" or "pretermination"). Most are formed from single nucleotide polymorphisms (SNPs). Premature termination of translation often leads to a truncated protein unable to fulfill its functions, might result in mRNA retention in the nucleus,[1] and promotes mRNA instability via the nonsense-mediated mRNA decay (NMD) pathway. [2]

(Alternatively, however, it is possible that protein function may be only modified rather than eradicated despite premature termination, because functional domain(s) might remain intact and/or if the variant only affects particular splice forms. Theoretically a membrane domain might be eliminated from a protein resulting in enhanced, and altered, function of the now water-soluble truncated peptide. It is also possible that occasional read-through of an altered stop codon might occur.[3,4])

In a genome-wide study of pretermination codons, treated as a class, they were (1) found to be common and (2) found overall to have been disadvantageous over evolutionary timescales, with only some indicated to have possible beneficial effects.[5] Pretermination codons are so prevalent in the human population that MacArthur et al. [6] have suggested unexpected widespread redundancy in the human genome in order to cope with loss of function of particular proteins. They estimated that a human genome typically contains around 20 genes which have been completely inactivated, presumably compensated by the presence of other proteins with compensatory function (it should be noted, however, that they also suggest that further validation is needed for these loss-of-function mutations).

In most cases it is expected that pretermination mutations result in loss of function which could then be associated with disease. Macarthur et al. [6] identified 26 recessive disease-causing mutations associated with severe early-onset conditions such as Leber congenital amaurosis, harlequin ichthyosis, osteogenesis imperfecta and Tay-Sachs disease, and a further 20 strong candidates for dominant Mendelian disease, including adult-onset muscular dystrophy, Charcot-Marie-Tooth disease and mucolipidosis. They also predicted that loss-of-function variants might be associated with the risk of common, complex diseases such as Crohn's disease and rheumatoid arthritis.

It is therefore possible, given the known complexity of the genetics of obesity, that association might be found between this phenotype and presence of one or more pretermination codons. It is thought that genetics could contribute up to 70% of the risk for obesity [7] and over 100 genes or genetic variants have been found to contribute to the risk for obesity, with most forms being strongly influenced by an obesogenic environment i.e. that the presence of obesity-related genetic combinations does not necessarily result in obesity without such an environment. (There are also some serious forms of childhood monogenic obesity.)

A genome-wide association (GWAS) study by Albuquerque et al. [8] found that the *FTO* gene, which encodes for the protein alpha-ketoglutarate-dependent dioxygenase FTO, had a strong influence on polygenic susceptibility to obesity, and is believed to affect the ratio between food intake and energy expenditure.[9] Some *FTO* intronic variants (block 8) were found to be strongly associated with being overweight in males but not females.[10] Other genes associated with obesity include the *MC4R* gene which is synergistic with *FTO;* the leptin gene and its receptor, and genes which encode for ectoenzyme nucleotide pyrophosphate phosphodiesterase 1, tumor necrosis factor alpha, interleukin-6, peroxisome proliferator–activated receptor gamma, angiotensin-converting enzyme, and glutathione S-transferase genes.[7] These are thought to influence one or more of the following "food intake control, appetite behaviour, energy balance, insulin signalling, glucose and lipid metabolism, adipocyte (...) differentiation, and metabolic disorders".[7] Muller et al. [11] have argued that obesity is such a complex phenotype that studies should concentrate on more specific phenotypes such as one or more in this list, but Speakman et al. [9] have counter-argued that an increase in sample numbers might well still provide insights into obesity as a whole. In 2017, Albuquerque et al. [8] suggested that the genetic variants found so far contribute only a small percentage of the total estimated body mass index (BMI)-heritability (which can be assumed to be closely related to the risk of obesity and have a large genetic component), and in any case probably cannot account for the rapid spread of obesity. Therefore, despite the fact that new SNPs associated with obesity have been found since then (e.g. see Sobalska-Kwapis et al. [10]), a search for further genetic variants is still warranted.

The aim of the present study was to assess possible association between BMI and the pretermination codons found in a large exomic SNP database ("POPULOUS") collected from the entire geographical region of the Polish population. The POPULOUS database contained around 550 000 SNPs each from 5757 subjects who declared themselves healthy. Based on data from the Polish Central Statistical Office

in the year 2012, this group represented ~ 0.015% of the Polish population. The hypothesis was that a pretermination codon might be found to be associated with obesity.

Material And Methods

Access to the POPULOUS database was granted for this study after a licence agreement was signed (PUM_UL_001). The POPULOUS database was the outcome of the project TESTOPLEK (funded by the Innovative Economy Operational Programme provided by the European Regional Development Fund 2007–2013), which was approved by the regional ethical committee (Institutional Review Board of the University of Łódź) and all procedures were in accordance with the current Declaration of Helsinki (2013). Genetic data was made available from anonymous Polish unrelated volunteers, who had declared themselves as healthy and signed written informed consent. Procedures for collecting samples, DNA isolation and genetic analysis can be found in Sobalska-Kwapis et al..[10]Exome SNP beadchips (HumanCoreExome-24 v1.0 and v1.1; Illumina, San Diego, CA, USA) were used giving SNP allele values for 551915 SNPs.

The SNP rs numbers (defined at www.ncbi.nlm.nih.gov/snp) of all 246 SNPs found in Table S1 of Fujikura [12] were searched for and a total of 141 premature termination codons resulting from single nucleotide polymorphisms were identified as having rs numbers in the Illumina beadchip lists. A full list of all SNP values (including duplicates) for all 5757 subjects is given in Supplemental_Table_S2 (this table is also found at https://github.com/ Abiologist/PTCobesity.git). Two SNPs were removed from analyses: rs7120775: allele G gives a pretermination codon, but the beadchips analyse alleles C and T; and rs545652: the gene *C17orf77* or *CD300LD* now recognised as giving anti-sense RNA only and no protein (see www.ncbi.nlm.nih.gov/snp/rs545652 and www.ncbi.nlm.nih.gov/gene/146723). The remaining 139 SNPs are also listed in Supplemental_Table_S1 of the present article. Two SNPs, rs497116 and rs35032582, were found to have only one allele (A and C, respectively) for those subjects with BMI data, leaving 137 SNPs for analysis.

All statistical analyses were performed using the R statistical platform (version 4.1.0, 2021; https://cran.r-project.org [13]). Preliminary regressions used function R [SNPassoc] *WGassociation* [14] for five models (codominant, dominant, recessive, overdominant (= heterozygote model), log-additive) using body mass index (BMI, mass/height^2) as a continuous variable. Additional preliminary tests were performed (only comparing groups 2 and 4) using BMI as a categorical variable with four groups as defined by the World Health Organisation: BMI group 1: "underweight", < 18.5 kg/m^2; group 2: "normal weight", 18.5 to 24.99 kg/m^2; group 3: "overweight", 25 to 29.99 kg/m^2; group 4: "obese", > 30 kg/m^2 [15]; with Fisher's tests: R [stats] *fisher.test (simulate.p.value = TRUE*); and Cochran-Armitage trend tests: R [DescTools] *CochranArmitageTest*.[16] All preliminary tests could not include adjustments for interactions.

Further regressions were therefore performed using R [stats] *lm* using *anova* to compare e.g. with or without adjustment for interaction with sex and/or year of birth. This is important as it is well known, and has been shown for the database used, that year of birth is associated with obesity with increase over

time. [10] All statistical tests were two-tailed with cut-off defined as p = 0.05, with or without Bonferroni or false discovery rate correction.

Results

For each of the 139 premature termination codons resulting from single nucleotide polymorphisms identified in the Illumina exome beadchip lists, body mass index information for three possible genotypes is given in Supplemental_Table_S3. and for rs67047829 is summarized, including according to sex and with age data, in Table 1.

Table 1
Characteristics of subjects with body mass index from the POPULOUS database, according to sex and rs67047829 genotype. m.a.d. = median absolute deviation; s.d. = standard deviation; n = number of subjects. [a,b]Significant difference among genotypes: heterozygote model.

Group	Genotype	n	Body mass index (kg/m^2)		Age (years)	
			median (m.a.d.)	mean (s.d.)	median (m.a.d.)	mean (s.d.)
All subjects		5095	24 (3)	24.8 (4.56)	41 (12)	42.0 (15.5)
Both sexes together	**AA**	56	23 (3)	22.9[a] (3.68)	39.5 (13.5)	42 (15.3)
	GA	842	24 (3)	24.7[a] (4.44)	41 (12)	41.8 (15.4)
	GG	4197	24 (3)	24.9[a] (4.59)	41 (12)	42 (15.6)
Males	**AA**	32	20.5 (2.5)	22.2[b] (3.96)	40.5 (12)	43.3 (14.9)
	GA	433	23 (3)	23.8[b] (4.44)	40 (12)	41.9 (15.2)
	GG	2140	23 (3)	24.3[b] (4.96)	41 (12)	42.3 (15.9)
Females	**AA**	24	23.5 (2.5)	23.8 (3.11)	36.5 (12)	40.2 (15.9)
	GA	409	25 (3)	25.6 (4.24)	41 (13)	41.7 (15.5)
	GG	2057	25 (3)	25.5 (4.08)	40 (12)	41.7 (15.2)

Initial regression results (without interactions) are given in Supplemental_Table_S3 for all SNPs, and a Manhattan plot is shown in Fig. 1. Several SNPs gave low p values with regressions, but rs67047829 (chromosome 7:64992360, GRCh38.p13) located in *ERV3-1* (NC_000007.14) gave the lowest values overall and $p < 0.01$ with four models (Supplemental_Table_S3) and therefore was chosen for further study.

A full regression study was then conducted with SNP rs67047829 in order to evaluate the effects of year of birth (which is known to be associated with obesity) on the association with BMI: results are shown in Table 1 and Supplemental_file_OR1. Results showed that a significant association, which was interaction independent, was found between the rs67047829 heterozygote model and body mass index of both sexes together: p value adjusted for year of birth and sex, but interactions removed: $p = 0.00145$; BMI for AA, GA, GG: means (standard deviations, n): 22.9 (3.68, n = 56), 24.7 (4.44, n = 842), 24.9 (4.59, n = 4197)).

Corresponding values for males were also significant: ($p = 0.00349$; AA 22.2 (3.96, n = 32), GA 23.8 (4.44, n = 433), GG 24.3 (4.96, n = 2140)) but were not significant for females: $p = 0.165$, AA 23.8 (3.11, n = 24), GA 25.6 (4.24, n = 409), GG 25.5 (4.08, n = 2057) (Fig. 2); note this lack of significance with females might simply be because of the smaller number of female AA homozygotes. For Bonferroni correction see Discussion.

Discussion

The SNP rs67047829 was found in the present study to be associated with obesity in males and females together or in males alone, using a heterozygote model adjusted for year of birth (and interaction with year of birth not significant). Note that positive results from other models disappeared when adjustment for interaction with year of birth was included (Supplemental_file_OR1) and therefore no relevant conclusions could be drawn from other models.

For the heterozygote model the fact that year of birth did not disturb these results is indicated by the lack of interaction in the regressions and also from Table 1: median age from the AA genotype was slightly lower than for the other genotypes for both sexes together, but mean age was slightly higher for males alone.

The individuals' years of birth distributed between the genotypes in the heterozygote model in a way which gave a non-significant interaction adjustment i.e. that year of birth could be discounted as giving the effect observed. It should not be concluded from this that the heterozygote model is the best model to describe the effects of genotype on obesity (which is obvious from Fig. 2), only that for this model the distribution of years of birth was conducive to maintain the result in the data analysed. It is well known, and has be shown from this data,[10] that obesity in the population has increased with time, and more data from a large cohort at the same age might be needed to extend this result, in order to avoid further interaction with year of birth, to other models (but see discussion of ethnic background below).

Note also it appears from the BMI distributions (Fig. 2) that results for males and females were rather similar, with BMI decreasing from GG to GA to AA, but significance disappeared for females: possibly indicating insufficient data for this sub-group.

The final p value for genotype:BMI (adjusted for year of birth but not interaction) for males and females together was 0.00145 and for males alone was 0.00349. Although these are low note they do not survive Bonferroni (0.05/137 = 0.000365) or false discovery rate correction (see Supplemental_Table_S3) and the results remain as "fishing trip" results which await confirmation from a larger study or from a different genetic background (see below).

The SNP rs67047829 (*ERV3-1*:NC_000007.14:g.64992360[*G > A*]; NM_001007253.4:c.667[C > T]; NP_001007254.2:p.Arg223Ter) A allele provides a premature termination codon in the SU domain of the "Endogenous retrovirus group 3 member 1, envelope" (UNIPROT short name: ERV3-1 envelope protein, ERV3-1env), probably, but not necessarily, resulting in degradation of mRNA before a protein can be produced. If ERV3-1env is normally responsible for stimulating adipose cell proliferation, then removal via this SNP might confer a protective effect.

ERV3-1env is composed of two major domains, SU and TM (which split from each other during processing and are held together by non-covalent bonds giving a heterodimer). In a virus, SU would mediate receptor recognition whereas TM would be a transmembrane domain. If *ERV3-1* with the rs67047829 A allele did produce a protein then this would be without the TM domain and would have 222 rather than 604 amino acids.

Numerous retroviral introgressions into primate or human nuclear DNA have occurred throughout evolution, to create the class of human endogenous retroviruses (HERV). HERVs, along with long terminal repeat (LTR) elements, constitute nearly 8% of the human genome.[17] To produce an intact virus, a HERV genome would need active *gag*, *pol* and *env* genes flanked by LTRs.[18] However, although most HERV remnants (which are still referred to as human endogenous retroviruses) are now virally inactive, protein expression often still occurs and HERVs have been associated with several autoimmune diseases.[19]

The ERV3-1 envelope protein (ERV3-1env) has not been found to compose an element of a virus and it is well-established that it has lost its fusogenic properties.[20] However it is recognised that, as its open reading frame has been conserved through 30 million years of primate evolution, and as full-length proteins (from four exons) are expressed in many tissues,[21] it likely has a biological function beneficial to the host.[22] In a placental trophoblast model with BeWo cells stably transfected with ERV3-1env, beta-human chorionic gonadotrophin (hCG) expression, which positively regulates the cell cycle,[23] was increased; cyclin B expression, which promotes cell cycling, was reduced; while p21 expression, which negatively regulates the cell cycle, was up-regulated.[24]

Of probable direct relevance to the result found here is the fact that, although ERV3-1env mRNA is found in all tissues, it is highly expressed in adipose tissue. According to the databases found in the Human

Protein Atlas (https://www.proteinatlas.org [21]), adipose tissue had the highest ERV3-1env mRNA expression in one database (number of tissues: 45) and was found in the top four mRNA expression levels in the other databases (note that, conversely and rather strangely, protein expression was found to be low in adipose tissue). (As an aside it is interesting to note that Luteinising hormone plays a cardinal role in androgen production; in three databases the highest expression of ERV3-1env was found in adrenal glands, and a combination of LH/human chorionic gonadotropin (hCG) was found to directly stimulate T-I cell proliferation.)

Lastly, the SNP rs67047829 A allele shows high ethnic dependence with a prevalence of 4% in people of relatively-recent African descent, 9% with European descent and 19% with Asian descent,[25] and analysis of people with Asian descent might possibly provide an easier route to the elucidation of relationships between this SNP and obesity phenotypes.

In conclusion, this fishing study of premature termination codons using a large sample of the Polish population (n = 5757) has allowed discovery of a possible association with a single nucleotide polymorphism, rs67047829 in *ERV3-1*, a gene known to be highly expressed as mRNA in adipose tissue, with obesity. This pretermination SNP is highly likely to change ERV3-1 envelope protein expression and activity, although it is not known whether remnant proteins would be active or not. The association found with body mass index, independent from date of birth, indicates a possible protective effect against obesity. Further study involving an even larger cohort or from another (e.g. Asian) ethnic group is needed to confirm and extend this result and in vitro studies could determine whether remnant protein survives in adipose cell cytoplasm or is destroyed by nonsense-mediated mRNA decay. Functional studies are also needed.

Declarations

Acknowledgements: We would like to thank Marcin Słomka for technical assistance.

Authors' contributions: JSCC and KP contributed to concept and design, literature search, data analysis, statistical analysis, manuscript preparation; MS and BM to data acquisition and data analysis; KR and TW to data analysis and statistical analysis; AC and DS to interpretation, manuscript editing and manuscript review.

Availability of data and materials: All data is found in Supplemental materials or https://github.com/Abiologist/PTCobesity.git.

Competing interests: The authors have declared that no competing interests exist.

Ethics approval and consent to participate: Database production was approved by the regional ethical committee (Institutional Review Board of the University of Łódź). Genetic data was from anonymous healthy Polish unrelated volunteers who signed written informed consent (see Sobalska-Kwapis et al. [10]).

Consent for publication: consent from relevant parties has been obtained according to licence agreement PUM_UL_001.

Funding: The POPULOUS database was the outcome of the project TESTOPLEK which was funded by the Innovative Economy Operational Programme provided by the European Regional Development Fund 2007-2013. This source had no involvement in the study other than funding for the database.

References

1. Shi, M. *et al.* Premature Termination Codons Are Recognized in the Nucleus in A Reading-Frame Dependent Manner. Cell Discovery **1**, 15001 (2015).
2. Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. Nature reviews Molecular cell biology **13**, 700–712 (2012).
3. Jungreis, I. *et al.* Evidence of abundant stop codon readthrough in Drosophila and other metazoa. Genome Research **21**, 2096–2113 (2011).
4. Wills, N. M. (eds Atkins, J. F. & Gesteland, R. F.) 365–381 (Springer, New York, NY, 2010).
5. Yngvadottir, B. *et al.* A Genome-wide Survey of the Prevalence and Evolutionary Forces Acting on Human Nonsense SNPs. American Journal of Human Genetics **84**, 224–234 (2009).
6. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. Science (New York, N.Y.) **335**, 823–828 (2012).
7. Golden, A. & Kessler, C. Obesity and genetics. Journal of the American Association of Nurse Practitioners **32**, 493–496 (2020).
8. Albuquerque, D., Nóbrega, C., Manco, L. & Padez, C. The contribution of genetics and environment to obesity. British Medical Bulletin **123**, 159–173 (2017).
9. Speakman, J. R., Loos, R. J. F., O'Rahilly, S., Hirschhorn, J. N. & Allison, D. B. GWAS for BMI: a treasure trove of fundamental insights into the genetic basis of obesity. International Journal of Obesity **42**, 1524–1531 (2018).
10. Sobalska-Kwapis, M. *et al.* Genetic association of FTO/IRX region with obesity and overweight in the Polish population. PLoS One **12**, e0180295 (2017).
11. Müller, M. J. *et al.* The case of GWAS of obesity: does body weight control play by the rules? International Journal of Obesity (2005) **42**, 1395–1405 (2018).
12. Fujikura, K. Premature termination codons in modern human genomes. Scientific reports **6**, 1–9 (2016).
13. R CORE TEAM, A. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. *URL* http://www. R-*project. org* (2020).
14. Victor, M. & Juan, R. G. SNPassoc: SNPs-Based Whole Genome Association Studies 2021).
15. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic* (World Health Organization, 2000).

16. Signorell, A. E. M. A. *DescTools: Tools for Descriptive Statistics* 2021).
17. Kim, T.-H. *et al.* The distribution and expression of HERV families in the human genome. Molecules and Cells **18**, 87–93 (2004).
18. Nelson, P. N. *et al.* Human endogenous retroviruses: transposable elements with potential? Clinical and Experimental Immunology **138**, 1–9 (2004).
19. Krzysztalowska-Wawrzyniak, M. *et al.* The distribution of human endogenous retrovirus K-113 in health and autoimmune diseases in Poland. Rheumatology (Oxford, England) **50**, 1310–1314 (2011).
20. Blaise, S., de Parseval, N., Bénit, L. & Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proceedings of the National Academy of Sciences of the United States of America **100**, 13013–13018 (2003).
21. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. Science (New York, N.Y.) **347**, 1260419 (2015).
22. Venables, P. J., Brookes, S. M., Griffiths, D., Weiss, R. A. & Boyd, M. T. Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. Virology **211**, 589–592 (1995).
23. Palaniappan, M. & Menon, K. M. J. Human Chorionic Gonadotropin Stimulates Theca-Interstitial Cell Proliferation and Cell Cycle Regulatory Proteins by a cAMP-Dependent Activation of AKT/mTORC1 Signaling Pathway. Molecular Endocrinology **24**, 1782–1793 (2010).
24. Lin, L., Xu, B. & Rote, N. S. The cellular mechanism by which the human endogenous retrovirus ERV-3 env gene affects proliferation and differentiation in a human placental trophoblast model, BeWo. Placenta **21**, 73–78 (2000).
25. Phan, L. *et al. ALFA: Allele Frequency Aggregator.* (National Center for Biotechnology Information, U.S. National Library of Medicine, 2020).

Figures

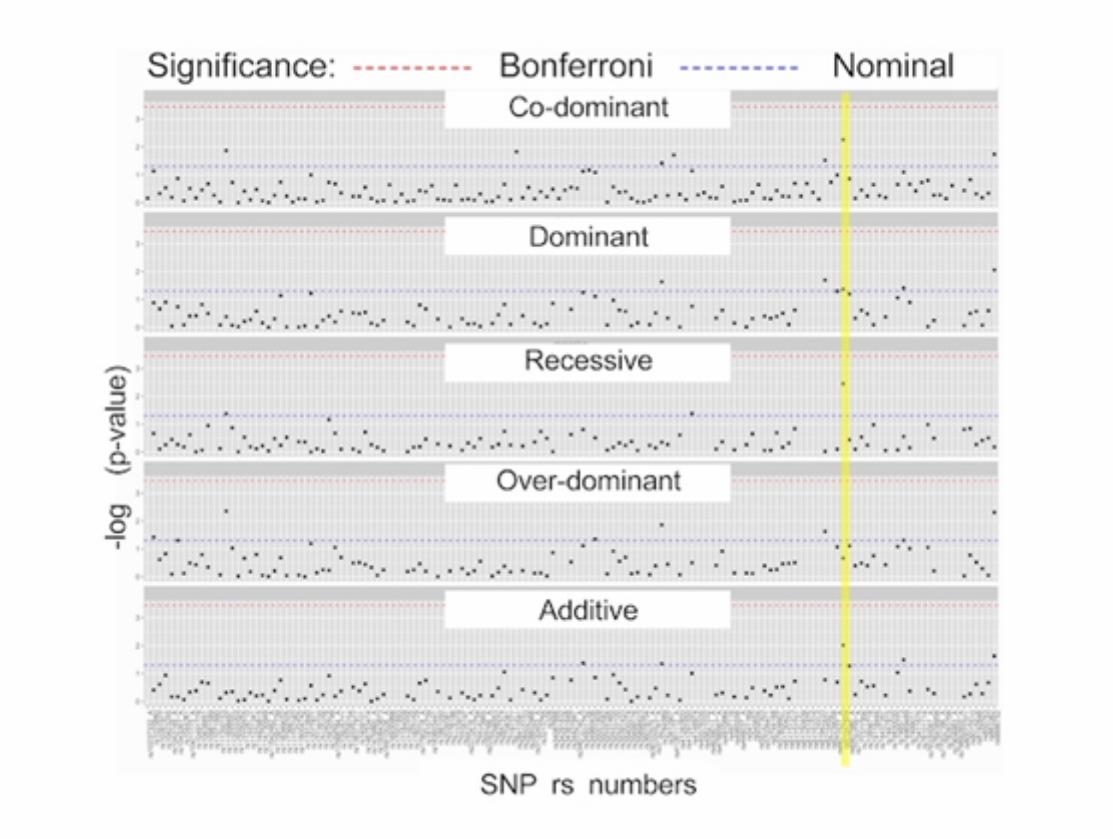

Figure 1

Manhattan plot for several regressions relating pretermination single-nucleotide polymorphism (SNP) genotypes with obesity. Points for SNP rs67047829 are highlighted in yellow.

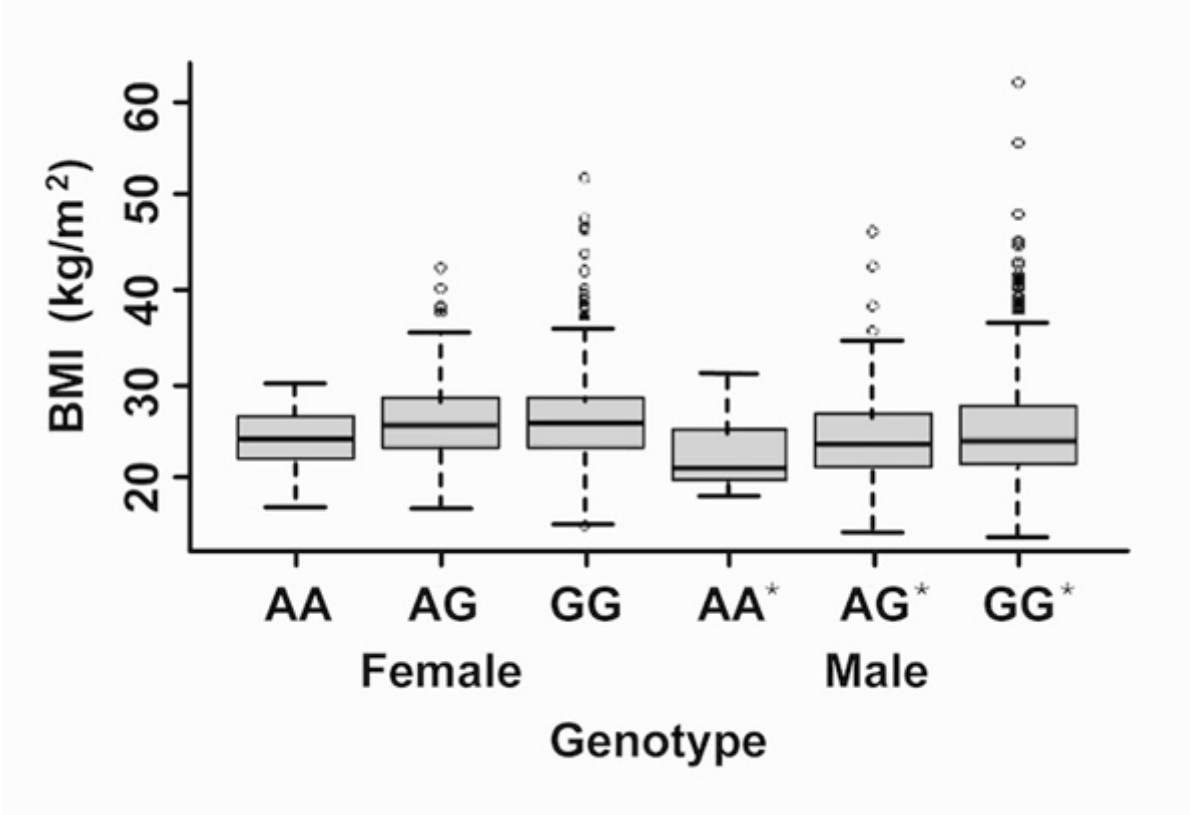

Figure 2

Body mass index (BMI; kg/m^2) versus genotype of pretermination single nucleotide polymorphism rs67047829. Box plots show medians (m), interquartile range (IQR: $q_{0.25}$ to $q_{0.75}$), range between $q_{0.25}$-1.5(m-$q_{0.25}$) to $q_{0.75}$+1.5($q_{0.75}$-m), and outliers. *Significant different among genotypes with heterozygote model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalTableS1E.xlsx
- SupplementalTableS2B.xlsx
- SupplementalTableS3E.xlsx
- SupplementalfileOR1.docx