



OPEN

Concurrent outcomes from multiple approaches of epistasis analysis for human body mass index associated loci provide insights into obesity biology

Sheldon D'Silva¹, Shreya Chakraborty^{1,2} & Bratati Kahali¹✉

Genome wide association studies (GWAS) have focused on elucidating the genetic architecture of complex traits by assessing single variant effects in additive genetic models, albeit explaining a fraction of the trait heritability. Epistasis has recently emerged as one of the intrinsic mechanisms that could explain part of this missing heritability. We conducted epistasis analysis for genome-wide body mass index (BMI) associated SNPs in Alzheimer's Disease Neuroimaging Initiative (ADNI) and followed up top significant interacting SNPs for replication in the UK Biobank imputed genotype dataset. We report two pairwise epistatic interactions, between rs21177596 (*RHBDD1*) and rs17759796 (*MAPK1*), rs1121980 (*FTO*) and rs6567160 (*MC4R*), obtained from a consensus of nine different epistatic approaches. Gene interaction maps and tissue expression profiles constructed for these interacting loci highlights co-expression, co-localisation, physical interaction, genetic interaction, and shared pathways emphasising the neuronal influence in obesity and implicating concerted expression of associated genes in liver, pancreas, and adipose tissues insinuating to metabolic abnormalities characterized by obesity. Detecting epistasis could thus be a promising approach to understand the effect of simultaneously interacting multiple genetic loci in disease aetiology, beyond single locus effects.

Susceptibility to complex human diseases, example, type 2 diabetes and Alzheimer's disease, among others are conferred by more than one gene as well as environmental influences. Technological advances in genomics have equipped researchers to make substantial progress in the field of genome wide association studies (GWAS) which is a robust hypothesis-free approach to scan entire genomes of individuals for identifying genetic loci that define susceptibility to complex diseases. GWAS have identified thousands of loci implicated in complex diseases unequivocally^{1,2}. Nevertheless, complete understanding of the pathophysiology of complex diseases remains elusive through GWAS, and the genetic variations identified through GWAS explains only a fraction of the heritability of the disease³. GWAS carried out for adult human height, detected associations with common variants that could explain about 60% of population variation^{3,4}. Near-independent genome-wide significant SNPs explain about ~6.0% of the variance of BMI (by 785 SNPs) in ~700,000 European ancestry individuals⁵. However, by carrying out GWAS on a relatively small cohort size of 96 cases and 50 controls, a variant causing an amino-acid replacement in the complement factor H gene (*CFH*) was found to increase the risk of age-related macular degeneration (AMD) by sevenfold when occurring in homozygous state⁶. Also, just 52 genome wide associated variants explain more than half of the genomic heritability for AMD⁷. Thus, GWAS has been a huge or moderate success depending on the disease trait or phenotype of interest. It has also been proposed that undetermined heritability is due to incomplete linkage disequilibrium (LD) between low frequency or rare causal variants and genotyped SNPs, and thus not detected by stringent significance tests⁸. Another proposition is that rare variants contribute to the missing heritability and optimally designed rare variant association studies would highlight important biological pathways for understanding disease aetiology⁹. Furthermore, evaluating each SNP individually and enforcing stringent correction for multiple testing can result in crucial variants to

¹Centre for Brain Research, Indian Institute of Science, Bangalore 560012, India. ²Interdisciplinary Mathematical Sciences, Indian Institute of Science, Bangalore 560012, India. ✉email: bratati@iisc.ac.in

Interaction RSID		Coordinates				Interaction results from SNPRuler, AntEpiSeeker, MDR, GMDR	
		SNP1		SNP2		ADNI	UKB
SNP1	SNP2	Chromosome	Position	Chromosome	Position	P-adj**	P-adj
rs2177596 (T: 46%: 42%)	rs17759796 (A:14%: 15%)	chr2	227,890,283	chr22	22,190,163	0.0324, 0.02005 (SNPRuler, AntEpiSeeker)	0.0 (SNPRuler)*
rs1121980 (A: 44%:42%)	rs6567160 (C: 23%: 23%)	chr16	53,809,247	chr18	57,829,135	0.0392, 0.0229 (SNPRuler, AntEpiSeeker)	(GMDR) CVC : 8 Testing Balanced Accuracy (TBA): 0.5346 Cutoff TBA ($p=0.05$): 0.5034 Cutoff TBA ($p=0.01$): 0.5044 p value : <0.001 (MDR) CVC: 9/10

Table 1. Significant epistatic pairwise SNP interactions for BMI associated loci. Chromosome and position with respect to GRCh37. * $P < 10^{-16}$. (Effect allele: Frequency in ADNI: Frequency in UKB). ** P -adj is based on 1000 permutation testing.

be overlooked. This is a severe problem given that these variants arise from components of pathways and genes that work in tandem to impact the concerned phenotype¹⁰.

Additionally, Bateson proposed that a variant or allele at one locus could prevent a variant at another locus from manifesting its effect¹¹. Bateson's theory was corroborated by Sewall Wright¹², where he acknowledged the relevance of gene–gene interaction by stating that selection pressure acts on the effects of the genetic background instead of individual genes. Shortly after, Sir Ronald Fisher described epistasis as the divergence from added effects on quantitative traits due to alleles at different loci¹³. Epistasis denotes the event in which the status of a trait arising from a genetic background deviates from the sum of its single locus effects, and can be interpreted as the effect of one locus being dependent on the allelic state of another or several other loci¹⁴. This idea has further been expounded when epistatic or gene–gene interactions among multiple loci in the genome have been observed for model organisms¹⁵. Even for human complex traits, it has been proposed that genetic interactions could explain some of the “missing heritability” undetected by GWAS^{16,17}, and can thus act as a governing factor underlying the biology of non-Mendelian phenotypes¹⁸. Epistasis from a biological perspective represents the synergy of gene products that are part of interconnected pathways associated with the concerned phenotype¹⁹. Significant epistasis analysis in recent years have furthered our understanding of complex traits, such as, atrial fibrillation²⁰, lung cancer²¹ and multiple sclerosis²², among others. In the pre-GWAS era, increased evidence for linkage at one locus had been observed when the interaction with another locus was considered for type 1²³ and type 2 diabetes²⁴ and inflammatory bowel disease^{25,26} showed that genetic variance for complex traits is predominantly additive and a large sample size of millions of unrelated individuals are needed for precise estimation of epistatic variance. Recently, successful epistasis in humans have been detected through genome screening projects^{27,28}, and by statistical methods of epistasis analysis^{29–31}. Campbell et al. has argued that studies about genotypic combinations in model organisms could facilitate our understanding for detection of epistasis in natural populations³². In this communication, we use the terms epistasis and genetic interactions interchangeably.

In this study, we have used population data from 785 individuals belonging to the Alzheimer's Disease Neuroimaging (ADNI) cohort in order to detect second order genetic interactions for SNPs known to be associated with body mass index (BMI). A recent exhaustive review meticulously details the journey of identifying single locus effects associated with human BMI, and how multidisciplinary analytical approaches could enhance our understanding of the biology of obesity and identify newer treatments³³. We employed nine different epistasis detection tools representing exhaustive, heuristics, and stochastic algorithms and identified 20 significant epistatic interactions for human BMI. We then take these top SNPs comprising the interactions for follow-up in the UKB imputed genotype dataset, constituting more than 188,000 individuals to carry out replication analysis. Our analysis shows that variants at genomic loci in or near *FTO* and *MC4R*; *RHBDD1* and *MAPK1*, exhibit reproducible pairwise interactions associated with human BMI.

Results

Statistical epistasis in human BMI-associated loci. We identified twenty significant pairwise interactions among BMI-associated loci (Supplementary Table S1), from the consensus results of nine tools in the ADNI dataset. We took forward the comprising SNPs in these interactions for replication in the independent imputed genotype dataset of UK Biobank for uncovering pairwise epistatic interactions associated with BMI.

We finally obtain two pairs of significant interactions in more than 188,000 individuals between rs2177596 (*RHBDD1*) and rs17759796 (*MAPK1*), rs1121980 (*FTO*) and rs6567160 (*MC4R*), from a consensus of nine different epistatic approaches. The detailed characterization of these interactions is given in the following sections.

Pairwise interactions detected. One of the significant interacting SNP pairs in ADNI replicated in the UKB dataset, as a consensus obtained from SNPRuler and AntEpiSeeker- rs2177596 (A/T) and rs17759796 (C/A), ($p_{adj} = 3.24E-02, 2.0E-02$) (Table 1), with corresponding BMI increasing alleles T ($\beta = 0.017, p = 5.61E-06$), freq = 46% (ADNI), 42% (UKB) and A ($\beta = 0.018, p = 4.2E-05$), freq = 14% (ADNI), 15% (UKB) respectively.

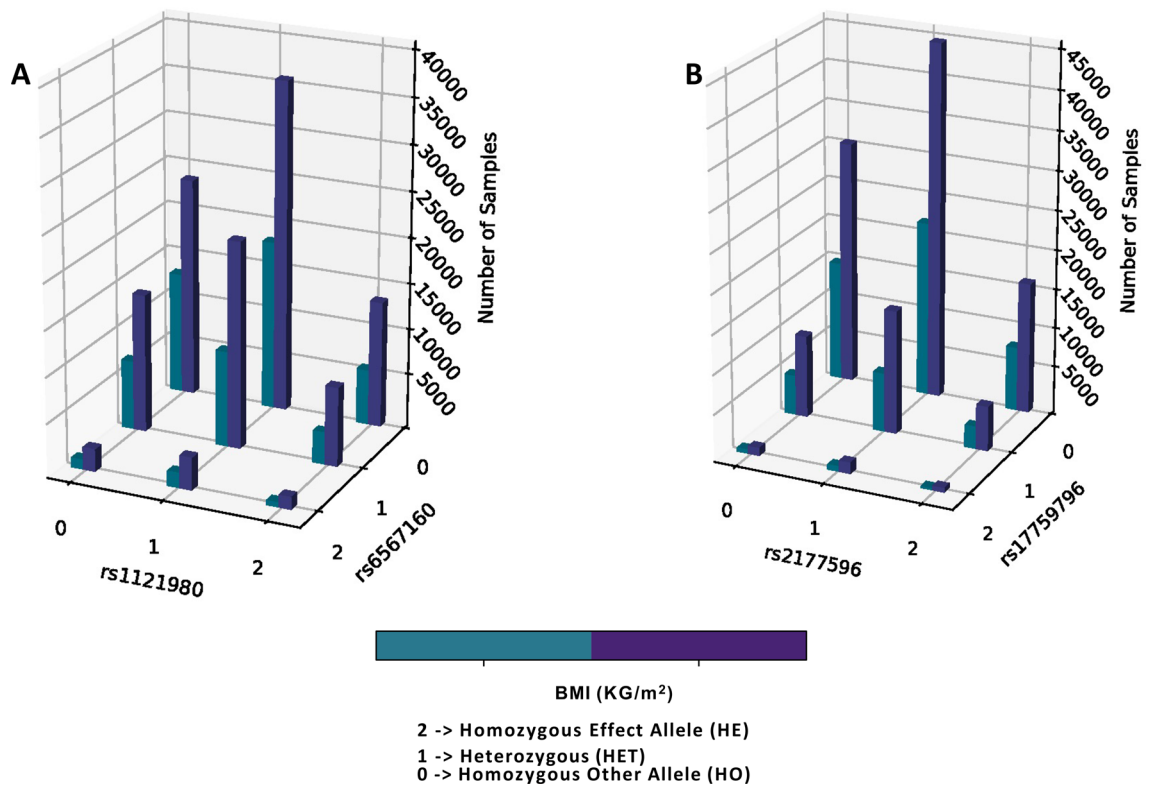


Figure 1. 3-D plot showing three genotypes at each locus for the interacting pairs (**A** rs1121980 and rs6567160, **B** rs2177596 and rs17759796) with respect to BMI prevalence in UKB population. 3D barplot showing prevalence of obesity against genotypes at each interacting locus for two SNP pair interactions: (**A**) rs1121980 and rs6567160, and (**B**) rs2177596 and rs17759796.

Additionally, one interaction, as a consensus of GMDR and MDR, between rs1121980(G/A) and rs6567160 (C/T) ($p_{\text{adj}}=0.001$) was replicated independently in the UKB genotype matrix, with corresponding BMI increasing alleles: A ($\beta=0.079$, $p=1.8\text{E}-142$), freq=44% (ADNI), 42% (UKB) and C ($\beta=0.056$, $p=3.93\text{E}-53$), freq=23% (ADNI), 23% (UKB) (Table 1, Table S1). Moreover, rs7613875 (C/A) (*MONIA/RBM5*) and rs8050136 (C/A) (*FTO/RPGRIP11*), identified as interacting in ADNI, failed to meet the criteria for permutation testing in MDR (cross validation consistency value of at least 7/10) in the UKB.

Allelic combinations observed in individuals for significantly interacting SNPs. We have investigated the genotype combinations of the two interacting SNPs against their prevalence in the UKB (Fig. 1). We observe that for every genotype combination in Fig. 1, the number of individuals is invariably more in the higher BMI group. This representation is based on absolute number of individuals carrying the said genotype combinations, and does not consider the total number of individuals in the high BMI and low BMI groups (126,651 and 61,969 respectively). There is marked difference in the total number of samples falling in the above-mentioned groups. When we closely observe the genotype combinations of effect and other alleles for the two interacting SNPs, individuals homozygous for BMI decreasing allele of both SNPs for the pair rs1121980 and rs6567160 were more likely to be predisposed to lower BMI as reflected in the UKB. This is because the proportion of individuals in the lower BMI group is significantly greater than the proportion in the higher BMI group; where proportions were assessed with respect to the total number of individuals in lower and higher BMI groups respectively (two sample z-test for proportions, $P<0.0001$) (Table 2). Similarly, proportion of individuals heterozygous for rs1121980 and homozygous for the BMI-decreasing allele of rs6567160 are significantly more in the lower BMI group ($P=0.0025$, Table 2). Proportion of carriers with heterozygous combination of rs1121980 and rs6567160 are observed to be predisposed to having increased BMI in UKB ($P=0.029$, Table 2). In UKB, individuals who are homozygous for the BMI increasing allele for rs1121980, however, either heterozygous ($P=0.014$) or homozygous carriers ($P=0.0185$) of BMI decreasing allele for rs6567160, are more likely to exhibit higher BMI for this SNP pair (Table 2). None of the genotypic combinations for interacting SNP pair rs2177596 and rs17759796 show any significant over or under-representation of proportional carriers towards high or low BMI.

This analysis shows the importance of deciphering the simultaneous effects from multiple SNPs acting in conjunction to influence the phenotypes of our interest in large-scale population datasets.

Further, we examine the interaction effects versus main effects for rs1121980 and rs6567160 where BMI is regressed for the SNPs with age and gender as covariates (Fig. S1A), and we see that the proportion of variance explained by main effects is 0.0122 and that by including interaction term the proportion of variance explained

rs1121980- rs6567160 (UKB)	ho/ho	ho/he t	ho/he	het/ho	het/he t	het/he	he/ho	he/het	he/he
BMI ≥ 25 kg/sq-m	0.187 98904 1	0.118 0093 3	0.019 07604 4	0.284 42728 4	0.176 99820 8	0.027 54814 4	0.108 07652 5	0.067 56362	0.010 31180 2
BMI < 25 kg/sq-m	0.212 75153 7	0.120 7216 5	0.016 57280 3	0.296 77742 1	0.167 16422 7	0.025 80322 4	0.096 91942 7	0.055 51162 7	0.007 77808 3
P-value	<0.00 01	0.554 3	0.545 8	0.002 5	0.029	0.72	0.018 5	0.014	0.63
rs2177596- rs17759796 (UKB)	ho/ho	ho/het	ho/he	het/ho	het/he t	het/he	he/ho	he/het	he/he
BMI ≥ 25 kg/sq-m	0.2429 11623	0.081 87065 2	0.0074 29866	0.3549 2811	0.1230 15215	0.0107 14483	0.1305 951	0.0445 16032	0.0040 18918
BMI < 25 kg/sq-m	0.2442 18884	0.081 71827 8	0.0074 23066	0.3576 94977	0.1203 98909	0.0100 05003	0.1307 42791	0.0446 67495	0.0031 30598
P-value	0.759	0.974 1	0.9989	0.4813	0.5704	0.8858	0.9742	0.9748	0.8637

Table 2. Combinations of allelic states for effect and other allele in individuals of high and low BMI. Column headers he, ho and het represent homozygous for the BMI increasing (effect) allele, homozygous for the BMI decreasing (other) allele, and heterozygous states respectively. Green cells represent the genotype combinations more likely to occur in low BMI rather than high BMI as per prevalence in UKB. Orange cells represent the genotype combinations more likely to occur in high BMI rather than low BMI as per prevalence in UKB. Level of significance: 0.05

becomes 0.0128. The interaction thus, explains 0.06% of the total phenotypic variance, which is not surprising given common variants like these can explain main-effects variance in the order of 0.008–0.3%³⁴. Similarly, we examine the interaction effects versus main effects for rs2177596 and rs17759796 (Fig. S1B), where the proportion of variance explained by main effects is 0.00897 and that by including interaction term the proportion of variance explained becomes 0.00896. The interaction thus, explains 0.005% of the total phenotypic variance.

Biological relevance of epistatic interactions. To systematically identify biological connections among the genetic loci uncovered for statistical epistatic interactions, we constructed interaction maps based on co-expression, co-localization, physical interaction, genetic interaction among the genes annotated to the interacting pair of SNPs.

The gene network map of interacting pair rs1121980 (chr16:53809247) and rs6567160 (chr18:57829135) is annotated to *FTO* and *MC4R* respectively, as shown in Fig. 2A. Interaction between *FTO* and *MC4R* has been documented in only a couple of studies for obesity in children and adolescents³⁵, and large artery atherosclerotic risk³⁶, both studies done in less than 500 subjects, and for gene-diet interactions in 7052 individuals with high cardiovascular risk³⁷.

Our results show for the first time that a pairwise interaction among SNPs mapping to *FTO* and *MC4R* is unequivocally associated with BMI in the UKB cohort of 188,620 individuals (after quality check). Our in-depth analysis reveals that *FTO* co-expresses with *MGRN1*³⁸ which in turn shares a physical interaction with *MC4R*, and it has been shown that *MGRN1* inhibits *MC4R* signalling by displacement of Galpha(s), accounting for coat colour and obesity, features of the mahoganoid phenotype in mouse, and plays a key role in insulin sensitivity^{39–42}. These results could further our understanding of the mechanism of energy intake and expenditure balance, with respect to satiety and weight loss or gain.

The gene network map for interacting pair rs2177596 (chr2:227890283) and rs17759796 (chr22:22190163) is shown in Fig. 2B. Genes annotated to rs2177596 are *RHBDD1* and *COL4A4*, the former is a functional component of endoplasmic reticulum-associated degradation for misfolded membrane proteins, and required for the degradation process of some specific misfolded endoplasmic reticulum (ER) luminal proteins, and apoptosis⁴³, and is found to genetically interact with *MAPK1*⁴⁴, a gene annotated to rs17759796. It is also known that many pathways in adipose tissue, liver, and pancreas can be disrupted during ER stress, and ER stress is one of the

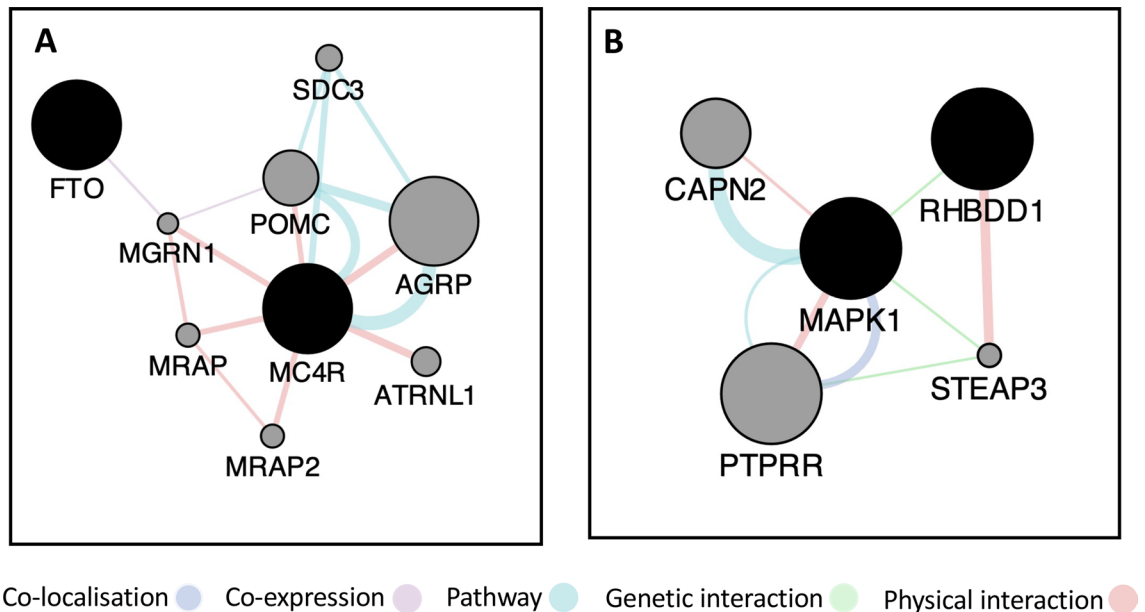
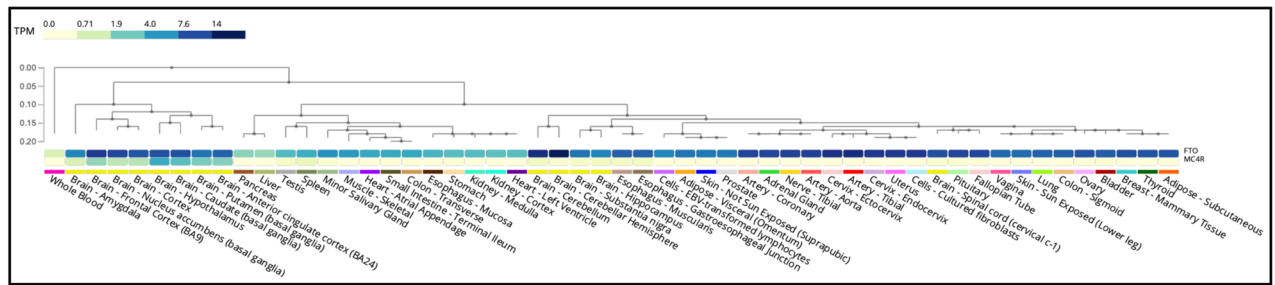


Figure 2. Gene networks. Nodes depicting genes are connected by edges formed on the basis of evidence for physical interaction (red), pathway (blue), co-expression (purple), co-localisation (violet) and genetic interaction (green). The maps for the detected interaction pairs are as follows: (A) rs1121980-rs6567160, (B) rs2177596-rs17759796. The solid black nodes are the query genes and the grey nodes are the resultant genes.

primary characteristics of obesity⁴⁵. We thus report here the tissue specific expression results in liver and pancreas from GTEx for these two interacting loci.

Expression profiles. Our results show a significant pairwise interaction between *FTO* (rs1121980) and *MC4R* (rs6567160) associated with human BMI in more than 188,000 individuals. Individually, both *FTO* and *MC4R* have emerged as strong signals for human obesity^{34,46}. In mice, studies have shown that *FTO* can regulate fasting and feeding, and therefore variation in *FTO* can result in changes of feeding behaviour and obesity⁴⁷. Recently, it has been shown that engineered deletion in the rs1421085 ($r^2 > 0.9$ with rs1121980 in Europeans) conserved cis-regulatory module can affect organismal phenotypes relevant to obesity in mouse⁴⁸. *MC4R* deficiency is known to be the most common cause of monogenic obesity⁴⁹, and gain-of-function variants in *MC4R* is associated with lower risk of obesity, type 2 diabetes, and coronary artery disease⁵⁰. We delve deeper into the individual-wise distribution of the *FTO* and *MC4R* alleles as explained in the previous section, and observe that carriers with rs6567160-TT (T is BMI decreasing allele) are predisposed to lower BMI in conjunction with rs1121980-AG, GG ($N = 91,407$). This is even more intriguing when we see that the entire UKB dataset has 61,969 individuals in the lower BMI category ($P < 2.2 \times 10^{-16}$ in a Binomial test). Thus, we posit that it might be possible that the BMI decreasing allele of rs6567160 near *MC4R*, contributing to its expression levels, has a significant mitigating effect on obesity in conjunction with the homozygous decreasing allele or heterozygous stature of the *FTO* SNP. *FTO* and *MC4R* are expressed in high levels in the hypothalamus and cortex in the brain as seen in GTEx tissue expression profiles (Fig. 3A). The eQTL analysis for BMI associated SNPs rs6567160 (BMI increasing allele = minor allele = C) and rs1121980 (BMI increasing allele = minor allele = A) in the hypothalamus and cortex is shown in Fig. 4A. They do not significantly regulate the expression of *MC4R* or *FTO*. Nevertheless, when we analyse the eQTL dataset for expressions of *FTO* and *MC4R* in cortex and hypothalamus, we notice that the lead SNP rs11873305 (Alt allele and minor allele = C) for *MC4R* significantly increases the expression ($\beta = 0.55$, $P = 3.5 \times 10^{-4}$) of *MC4R* in cortex, while the lead SNP rs10521305 (Alt allele and minor allele = C) for *FTO* significantly decreases the expression ($\beta = -0.55$, $P = 3.3 \times 10^{-13}$) of *FTO* in cortex. Similarly, significant increase and decrease of *MC4R* and *FTO* expression levels are observed in the hypothalamus for the corresponding lead SNPs rs8083758 ($\beta = 2.5$, $P = 1 \times 10^{-5}$) and rs10521305 ($\beta = -0.55$, $P = 5.1 \times 10^{-10}$) respectively (Fig. 4B). Thus, an increase in expression of *MC4R* in cortex and hypothalamus is accompanied by a decrease in expression to *FTO* in cortex and hypothalamus, with respect to the minor allele of these lead SNPs. However, these *FTO* and *MC4R* SNPs are low frequency (MAF = 1–6% in the European population) and not yet known to be associated with BMI, as well as not in LD (1000G European dataset) with the BMI-associated SNPs (rs6567160, rs1121980) for which significant epistasis has been identified in this study. rs6567160 and rs1121980 had been found to be associated with BMI from array-based genotypes and HapMap imputed studies. This makes us postulate that there could be another SNP in or near *MC4R* (say, m1) that is most likely to be causally associated with BMI, and in moderate LD with rs6567160 (BMI-associated), rs11873305 and rs8083758. Similarly, another SNP in or near *FTO* (say, f1) could possibly be the causal variant for BMI, and in moderate LD with eQTL lead SNPs rs10521305 and rs1121980 (BMI-associated). It would be motivating to uncover such SNPs f1 and m1 and investigate epistatic interaction between them, thus elucidating the network biology of obe-

A: Tissue specific expression profile of *FTO* (rs1121980) & *MC4R* (rs6567160).



B: Tissue specific expression profile of *RHBDD1* (rs2177596) & *MAPK1* (rs17759796).

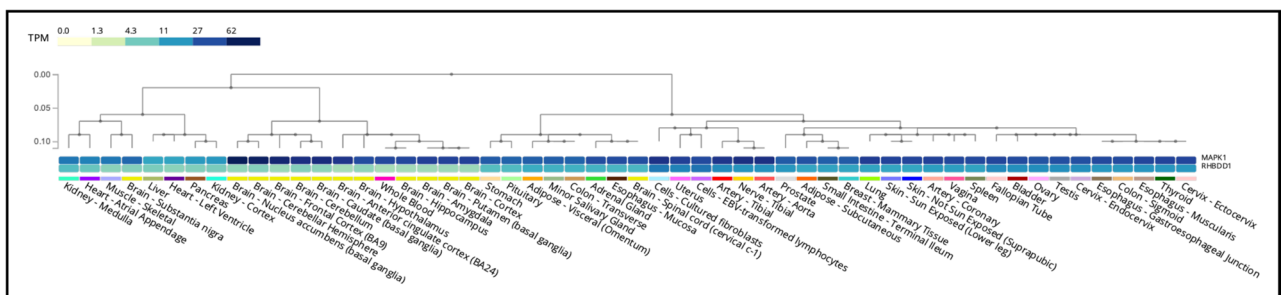


Figure 3. (A) Tissue specific expression profile of *FTO* (rs1121980) and *MC4R* (rs6567160). Comparison of expression (TPM) across various tissues, generated from GTEx for the genes *FTO* and *MC4R*. (B) Tissue specific expression profile of *RHBDD1* (rs2177596) and *MAPK1* (rs17759796). Comparison of expression (TPM) across various tissues, generated from GTEx for the genes *RHBDD1* and *MAPK1*.

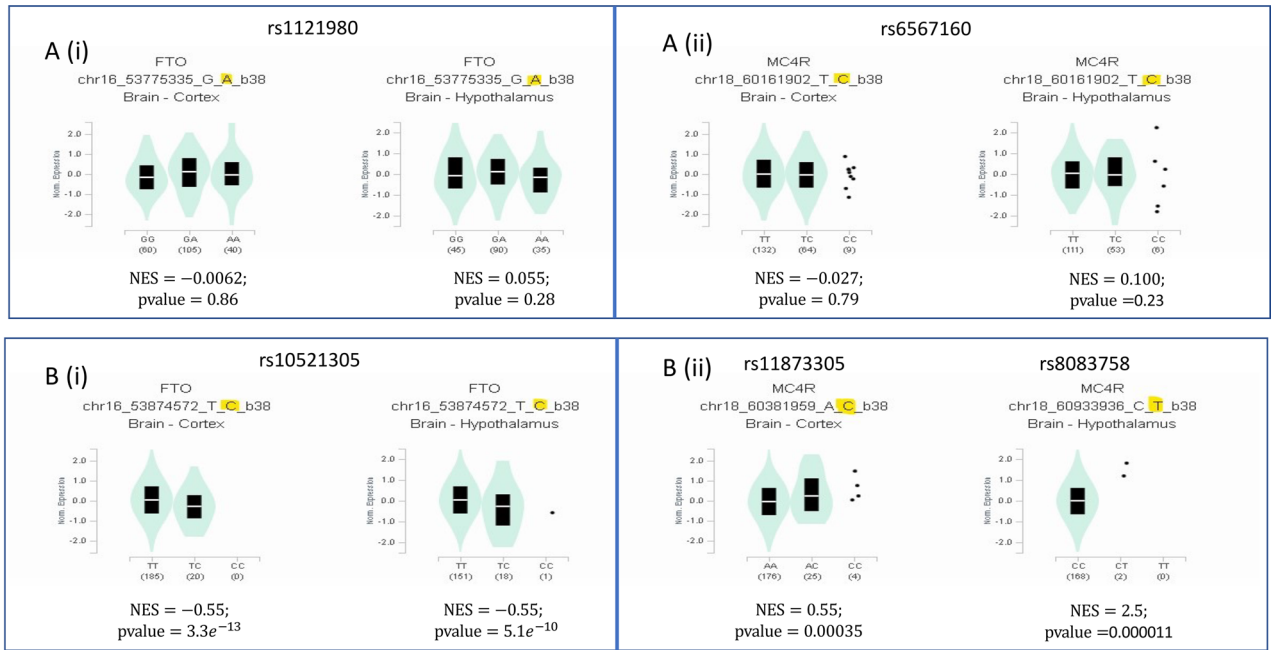
sity further (schema given in Fig. 4C). This observation from our results is especially interesting in the light of recent findings in an UKB-based study that gain-of-function variant [rs2229616 (coding change V103I), present in only 2% population] in *MC4R* can predispose individuals to lower obesity⁵⁰, and we observe that rs2229616 is in LD with the cortex eQTL lead SNP rs11873305 ($r^2=0.22$), which further provides evidence for the possible mechanisms we propose in Fig. 4C. With the technological advancements in genomics, when we uncover perfectly tagged variants by sequencing-based studies, discovery of epistatic effects would also be facilitated. Therefore, studying contribution from interaction effects of simultaneous genetic loci while considering the respective allelic combinations and uncovering eQTL SNPs in the same set of study individuals could facilitate our understanding of complex phenotypes.

rs17759796-A and rs2177596-T are respectively the BMI increasing alleles for *MAPK1* and *RHBDD1*. Our allelic combination analysis above do not reveal any statistically significant trend for belonging to higher or lower BMI for any of the nine possible genotype combinations for these two SNPs. We investigate the tissue-specific expression data for these two genes in GTEx. The two genes are expressed highly together in liver, pancreas, adipose-subcutaneous, and adipose-visceral tissues. In addition, *MAPK1* is widely expressed in several other tissues (Fig. 3B). Although the eQTL expression analysis of these genes in liver, pancreas, and adipose tissues (Fig. S2) could be indicative as to how they are modulated as a response to endoplasmic reticulum stress in obesity, the absence of a clear picture of any preferential allelic combination limits such in-depth investigation in this paper. There is also scarce genotype data for the lead SNPs for these two genes.

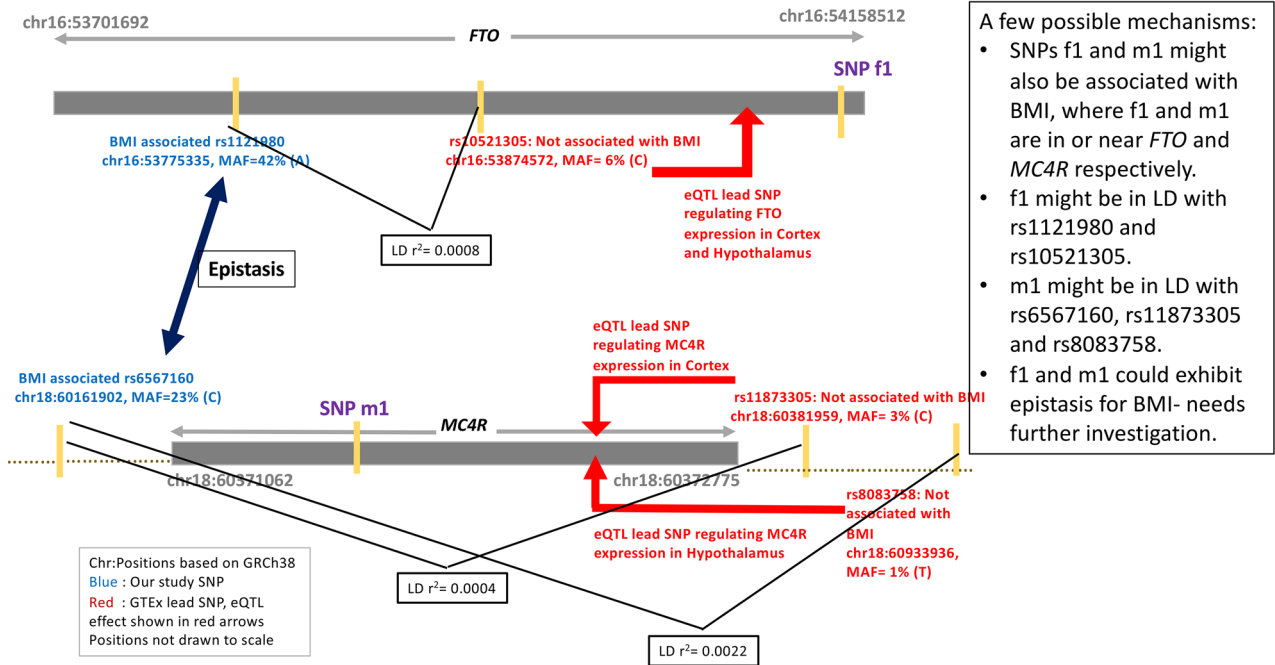
Discussion

We present here a comprehensive perspective of how allele combinations from interacting SNPs can influence complex traits like human BMI. Our epistasis analysis shows that a diverse combination of trait increasing and trait decreasing allele with respect to their homozygous and/or heterozygous status for both SNPs in the interacting pair lead to differential susceptibility to obesity in the population scale, as the frequency of occurrence and the underlying biological function of the interacting SNPs play crucial roles beyond the additive mechanism of individuals SNPs in the manifestation of the trait in population scale.

We have detected and replicated two significant pairwise interactions for human BMI between genome-wide associated loci in more than 188,000 individuals. Interestingly, we uncover for the first-time interactions among SNPs near *FTO* and *MC4R* associated to BMI. Although, these genes have been implicated in BMI, and *MC4R* is known to play significant role in energy homeostasis, the role of *FTO* in regulating energy intake and



A: Interacting SNPs. B: Lead SNP (rs10521305 for *FTO*, rs11873305 and rs8083758 for *MC4R*) for eQTL. NES: beta -> regression slope for highlighted allele. Positions with respect to GRCh38. C: eQTL profile of interacting SNPs and lead SNPs for *FTO* & *MC4R* in the genomic context.



A few possible mechanisms:

- SNPs f1 and m1 might also be associated with BMI, where f1 and m1 are in or near *FTO* and *MC4R* respectively.
- f1 might be in LD with rs1121980 and rs10521305.
- m1 might be in LD with rs6567160, rs11873305 and rs8083758.
- f1 and m1 could exhibit epistasis for BMI- needs further investigation.

Figure 4. eQTL profile of *FTO* (i) and *MC4R* (ii) SNPs in relevant tissues. (A) eQTL profile of *FTO* (rs1121980) and *MC4R* (rs6567160) in brain cortex and hypothalamus. (B) eQTL profile of lead SNP of *FTO* (rs10521305) and lead SNPs of *MC4R* (rs11873305 and rs8083758 expressed in brain cortex and hypothalamus respectively). (C) eQTL profile of interacting SNPs and lead SNPs for *FTO* and *MC4R* in the genomic context. Possible mechanisms of action of epistatic SNPs in regulating expression of relevant genes.

expenditure is an active area of research. We show in a first, that they can be part of biological interactions via *MGRN1*, depending on the genotype combinations and allelic status of the interacting SNPs. This interaction result detected by GMDR in elucidating epistasis mechanism leads to new insights in governing human BMI in populations, that could be tested in functional studies. Additionally, independent SNPs rs7613875 (*MON1A/RBM5*) and rs8050136 (*FTO/RPGRIP1*), identified as interacting in the discovery dataset, failed to meet the criteria for permutation testing in MDR (CVC value at least 7/10). Thus, future studies in *FTO* might shed more

light for understanding the biology of obesity, and lead to our precise understanding of how *FTO* plays a role in energy intake and expenditure.

Our input list consisted of SNPs in established genome wide significant thresholds as well as sub genome wide significance levels, yet significant epistasis interactions have been uncovered for SNPs that were associated with BMI for single locus effects within $P < 1E-04$. This suggests that there are loci waiting to be discovered for governing the biology of obesity, and probing upto a threshold of comparable significance for SNPs already achieved in GWAS to uncover such latent interactions is useful.

This epistasis study on BMI implicated loci helped us identify mechanisms potentially implicated in the biology of BMI that cannot be captured by single locus effects analysis. Some of the networks we uncovered, especially the connections to energy homeostasis, endoplasmic reticulum stress, concerted expression profiles of interacting genes in liver, pancreas, adipose tissue, and hypothalamus have direct implications in obesity biology. In statistical epistasis, interactions are defined by genetic variations and not by physical interaction of biomolecules in real systems, the latter is also hard to establish experimentally. Even though statistical models can implicate genetic variants that do not occur in transcribed regions, for which the functionality in biological mechanisms becomes hard to experimentally prove, they are still our most desired approach, especially in humans, to provide insights into unknown biological mechanisms occurring through interactions among loci across the genome. Additionally, statistical models can define instances where multiple genetic factors have a non-additive effect on a phenotype. It has already been acknowledged⁵¹ that biological interpretation of epistasis is usually easiest when the penetrance values all equal either 0 or 1, leading to a clear relationship between genotype and phenotype. However, this explanation is not straightforward for human complex genetic diseases due to their polygenic and multifactorial nature. In humans, due to the complexity of genomic architecture and multitudes of factors influencing complex traits like BMI, our results are probably the tip of the iceberg and many more indiscernible genetic interactions among human BMI associated loci are waiting to be uncovered.

Despite their widely accepted potential role in capturing the phenotypic variance in traits owing to genetic effects, epistatic interaction in human genetic analysis has been debated in recent times. This is largely because of imperfect linkage disequilibrium between causal variants with large additive effect size and nearby tagging loci resulting in inflated test statistics for interaction terms. Wood et al.⁵² reported that one more strongly associated variant uncovered from whole genome sequencing was moderately correlated with both of the interacting SNPs (showing very low levels of LD between them) identified by Hemani et al.⁵⁸, and the inclusion of this third variant as a covariate removed any evidence for interaction. This means that the apparent epistasis reported by Hemani et al. for SNPs on the same chromosome can actually be attributed to a single causal allele having moderate levels of LD with each of the two SNPs, and has recently been dubbed as phantom epistasis⁵³ between unlinked markers.

This is not the scenario in our results since we have detected inter-chromosomal interactions. On the contrary, our expression profile analysis reveals plausible mechanisms (Fig. 4C) involving causal variants at the epistatic loci that could be obtained by fine mapping.

Genetic variations making us susceptible to disease traits or determining the phenotypic variation in quantitative traits is a result of intricate selective forces at play striving to achieve the optimal balance in human systems. Epistasis is one of the intrinsic mechanisms on which such stabilising control ought to exist. Therefore, detecting statistical epistasis and deciphering biological implications for the same is of utmost importance in understanding disease mechanisms. Yet, we are cognizant that epistasis detection is limited by a number of factors. Statistical tests of interaction are limited to testing specific hypothesis in relation to genotypic measures in study subjects, and we will have insufficient statistical power to investigate all possible allelic combinations when dealing with rare variants. Additionally, when the causal variant or putative functional variant is in LD with another common variant, it is highly likely that the latter will show evidence for interaction in statistical tests due to linkage disequilibrium. In addition to these, inadequate controlling for overall genetic background, incidence of false positives, uncorrected environmental effects can constrain the identification of true epistatic effects present in human complex traits from statistical analysis.

Our results in discovery and replication dataset of more than 188,000 individuals highlight the importance of studying the interaction of SNPs for understanding the genetic architecture of complex traits. This is again more prominent in our in-depth analysis of the allelic combinations observed in individuals with high versus low BMI. We reiterate that our results are suggestively informing about plausible mechanisms to uncover putative interacting variants associated with obesity. We hope that our results will encourage large scale future studies to uncover epistasis in humans that will expound further the biology and genetic underpinnings of complex traits.

Methods

We considered BMI as our phenotype of interest in this work and set out to identify second order epistatic interactions among genetic loci associated with BMI at a genome wide level. The entire workflow is given in Fig. 5.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

Creating a phenotype and covariate file. The Alzheimer's Disease Neuroimaging Initiative (ADNI) along with UK BioBank (UKB) data repositories were used in this analysis, and we retrieved age, gender, height (in cm), and weight (in kg) for each participant. Among the 808 samples genotyped in the ADNI study, 785 had all the aforementioned fields. While 485,281 out of 487,405 samples with imputed genotypes in the UKB study

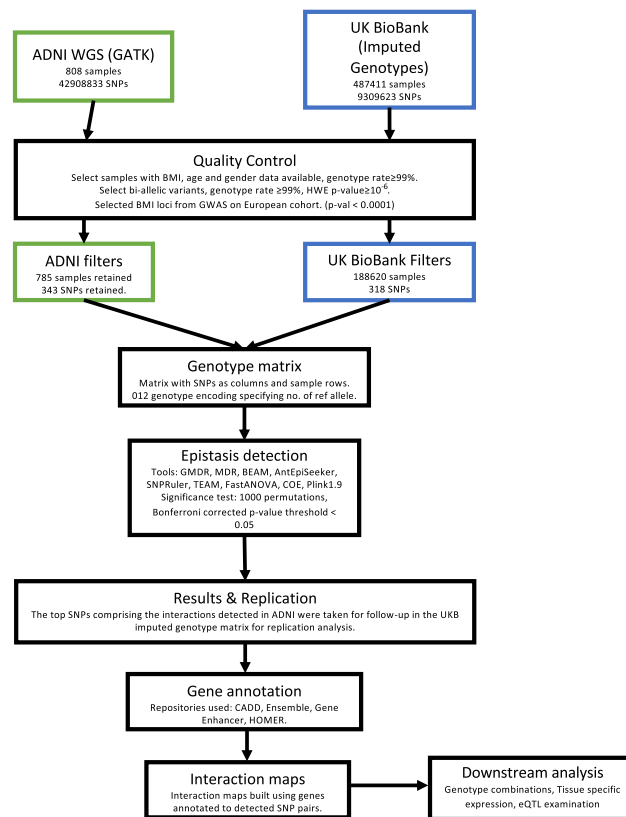


Figure 5. Study workflow. Schematic of procedures followed to detect and validate statistical epistatic interactions.

had the required phenotype fields. UKB samples that self-identified as 'White British' and belonged to a similar genetic ancestry based on a principal components analysis of the genotypes (407,560 samples) were retained. From this set, 188,620 samples did not possess any missing genotypes, and were used for further analysis.

BMI for these individuals was calculated by formula (weight in kilograms)/(height in metres)². Sample characteristics i.e., age and gender were parsed as covariates and BMI measure was selected as the phenotype. We have provided a threshold of BMI 25 kg/m² in order to suffice the condition of certain programs used in our study that requires the user to specify a class variable for the phenotype, although the approach is linear regression to identify interaction scores, based on the guidelines laid down by the World Health Organisation with respect to the European population (<https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>).

In order to ensure that BMI is not confounded by cognitive status in the ADNI sample set, we compared the BMI distributions in the ADNI study cohort among samples in these three categories of cognition i.e., Dementia, Minimal Cognitive Impairment (MCI) and Control Normal (CN). We did not find any significant difference (assessed by $p_{\text{adj}} < 0.01$) in BMI between CN and MCI, CN and Dementia, MCI and Dementia or MCI + CN versus Dementia in Mann–Whitney–Wilcoxon test (Fig. S3).

Selecting BMI Loci. We selected 466 genetic loci (Table S2) to be analysed for second order interactions for BMI by extracting SNPs associated with BMI from large scale meta-analysis studies^{34,46,54,55}. We retrieved 359 loci, after extracting the genome wide significant and sub genome wide significant loci (p -value < 0.0001) reported in these publications and retaining the most significant SNPs among those present within 500 kb of each other.

Generating genotype matrix. In order to rule out allelic bias due to any probable differential cognitive status among high and low BMI individuals, we checked for significant differences in allele frequencies of the 359 BMI-associated loci in a pairwise comparison of the CN, MCI, Dementia groups (Fisher's exact test, and adjusted by Benjamini–Hochberg Procedure). One SNP in the *TOMM40* region (rs2075650) with significantly different frequencies ($p_{\text{adj}} < 0.01$) was removed.

For the ADNI genotype matrix, only bi-allelic variants without missing genotype values and Hardy–Weinberg equilibrium exact test p -value > 10^{−6} were retained. Out of the 466 loci (Table S2), 343 variants (Table S3) were selected for further analysis. And for the UKB genotype matrix, the variants with imputation $r^2 \geq 0.7$ were extracted for all chromosomes (39091537 variants). Additionally, biallelic markers with genotyping rate > 99% and Hardy–Weinberg equilibrium exact test p -value > 10^{−6} (35937211 variants) were further selected.

We generated the ADNI genotype matrix with 343 variants, where individuals were represented as rows and the variants as columns. Of the possible 398 loci obtained from the four meta-analysis studies on BMI (p -value $\leq 10^{-4}$), 318 were present in the UKB imputed genotype set (Table S3). The allelic states within each cell are represented as 0: homozygous reference, 1: heterozygous and 2: homozygous alternate allele.

UKB replication set. Of the forty loci comprising the 20 interaction pairs obtained from the ADNI cohort, 8 loci (chr14:33302882, chr12:41948196, chr2:35404011, chr14:25928179, chr19:18454825, chr5:124330522, chr12:41921665, chr12:24075508) were not present in the UKB imputed genotype set, and the rest were investigated for interaction.

Epistasis detection tools. Multiple tools based on different classes of algorithms, for example, Exhaustive: FastANOVA⁵⁶, COE⁵⁷, TEAM⁵⁸, MDR⁵⁹, GMDR⁶⁰, PLINK⁶¹; MCMC: BEAM⁶² and Heuristic: AntEpiSeeker⁶³, SNPRuler⁶⁴; have been developed and are generally used to statistically define gene–gene interactions or multi-order epistasis from large scale population data.

FastANOVA. FastANOVA⁵⁶ relies on the application of ANOVA test for the detection of 2 locus interactions in a semi-exhaustive manner. It employs an upper bound which is the sum of one term derived from single-SNP ANOVA test, and the other derived from SNPs and independent of any phenotype permutation, to ensure effective SNP pair pruning.

COE. COE⁵⁷ employs a similar approach like FastANOVA, where the single locus test and the genotype of the loci pairs are required to compute the upper bound and reduces the number of computations pertaining to permutation testing by grouping the SNP pairs according to their genotypes. The approach carried out by COE also makes for a tighter upper bound which results in an efficient SNP pruning mechanism and can be applied on all statistical tests that are convex by nature.

TEAM: tree-based epistasis association mapping. TEAM⁵⁸ uses the construction of minimum spanning trees to define a contingency table that features the observed frequency of genotype permutations of each SNP pair related to its phenotype state. The table is then updated for every permutation of phenotype state across the population. Given that this method only considers those individuals with differing genotypes between SNP pairs, the computational burden to associate a SNP pair to the state of the phenotype is greatly reduced.

MDR: multifactor dimensionality reduction. MDR⁵⁹ is a nonparametric method that uses machine learning to reduce a 3×3 dimensional genotype matrix into a binary classification of case and control. MDR trains its models on 9/10th of the input genotype matrix and carries out the testing phase on the final (10th) partition. Finally, the MDR model with the least prediction error and the most cross validation consistency is selected.

GMDR: generalized multifactor dimensionality reduction. GMDR⁶⁰, based on the same principles as the MDR approach, can additionally handle quantitative phenotypes by taking into account a score statistic, generated by carrying out a generalized linear regression (GLM) with appropriate link function and suitable covariates, which is then used to map n -dimensional genotype combination into a 1-dimensional space. Final results are obtained after permutation testing.

*Plink 1.9*⁶¹.

Fast-epistasis. Calculates the odds ratio between loci A and B along with standard error. This procedure is performed for both cases and controls, and the epistasis test is defined as:

$$Z = (\log(R) - \log(S)) / \sqrt{SE(R) + SE(S)}$$

where R and S are the odds ratios in cases and controls respectively.

Epistasis. Can input quantitative phenotype values and uses linear regression to fit the model

$$Y = \beta_0 + \beta_1 g_A + \beta_2 g_B + \beta_3 g_A g_B$$

for each inspected variant pair (A, B), where g_A and g_B are allele counts; then the β_3 coefficients are tested for significance.

PLINK does not support including covariates or carrying out permutation testing for epistasis tests.

BEAM: Bayesian epistasis association mapping. BEAM⁶² is a Bayesian implementation that defines both single locus and multi-locus disease association from case–control datasets. It classifies a pair of loci as interacting if their combined distribution better generalises the phenotype data compared to considering the loci independently. It uses Markov Chain Monte Carlo on a case–control SNP matrix to iteratively build the posterior probabilities of phenotype association for each marker.

AntEpiSeeker. AntEpiSeeker⁶³ utilises Ant Colony Optimization (ACO) algorithm to identify statistical epistatic SNP pairs. Here, the SNPs being assessed form the probability distribution function that is updated through weights (pheromones). As the probability density function is fitted for the input genotype matrix, the epistatic interactions (paths) with higher corresponding scores get sampled by probes with increased frequency, resulting in these interactions being tagged with higher weight scores.

SNPRuler. SNPRuler⁶⁴ employs predictive rule learning to identify and assess SNP interactions. With the premise that n loci interactions and phenotype class are governed by rules which can be found and evaluated qualitatively more efficiently than the interactions themselves. SNPRuler builds an enumeration tree with SNPs as nodes and phenotypes representing the leaves, it then prunes the search space using an upper bound on χ^2 statistic.

All the tools described above and employed in this communication use permutation testing or multiple correction or in-built stringent P value threshold ($P < 1E-4$) to obtain robust epistasis results.

Hardware. We executed all the nine programs on a x86_64 architecture-based machine with an Intel(R) Xeon(R) Silver 4214 with 32 GB of memory (RAM) and CPU processing at a maximum clock speed of 3.2 GHz. The OS on the system was RHEL distribution CentOS 7.

Execution time. We directed COE, TEAM, FastANOVA, MDR and GMDR to carry out 1000 permutations. They clocked run times of 1 min 54 s, 16 s, 11 min 37 s, 16 min 8 s and 9 min 19 s respectively. BEAM, SNPRuler and AntEpiSeeker took 1 min 15 s, 1.14 s and 10 s respectively to output interactions at Bonferroni corrected $P < 0.05$. PLINK—fast-epistasis and—epistasis completed runs in 0.151 and 0.04 s respectively.

Memory utilisation. For executing the pairwise interaction analysis between 343 markers in 785 individuals, MDR had the highest memory consumption at 3200 MB. GMDR follows next with a consumption of 573.552 MB, followed by TEAM at 180 MB and SNPRuler at 180.06 MB. FastANOVA consumed 9.53 MB. BEAM, AntEpiSeeker had similar memory signatures with values of 2.84 MB, 2.18 MB respectively. COE consumed the least memory of 1.61 MB.

Biological relevance of interacting SNP pairs. *Generation of SNPs in linkage disequilibrium.* We obtained the variants in high LD with the SNPs that emerged significant in our epistatic models, using the LDproxy tool⁶⁵, from European sub-populations of 1000 Genomes Project (phase3). Only variants with $R^2 \geq 0.9$ were selected for further analysis.

Functional annotation of variants. We used each interacting SNP along with its corresponding LD variants as input for gene annotation in Combined Annotation Dependent Depletion (CADD)⁶⁶, SNP Nexus⁶⁷, Hypergeometric Optimization of Motif EnRichment (HOMER)⁶⁸ and gene enhancer regions⁶⁹.

Mapping the genes onto networks. The gene annotations were then parsed to generate a gene interaction network for the detected epistatic models. The gene interaction networks were generated using GeneMANIA⁷⁰, for the following subcategories- co-expression, co-localization, physical interaction, genetic interaction and shared pathways (Fig. 2).

Sample distribution by genotype states. For each of the detected SNP interactions, we extracted the number of samples for each genotype combination, thus resulting in nine combinations depending on the homozygous state for effect or other allele and heterozygous state of the two alleles for each pair of SNPs. We then compared the sample proportions in the high and low BMI groups for significant differences of these combinations using a two-sample z test for proportions (Table 2).

Data availability

All phenotype and genotype data used in this study for analysis are available at ADNI (<http://adni.loni.usc.edu/>) and UK BIOBANK (<https://www.ukbiobank.ac.uk>). We shall share the in-house scripts as required by other researchers.

Received: 13 August 2021; Accepted: 18 April 2022

Published online: 04 May 2022

References

1. Visscher, P. M. *et al.* 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
2. MacArthur, J. *et al.* The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucl. Acids Res.* **45**, D896 (2017).
3. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).
4. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
5. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
6. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science (80-)* **308**, 385–389 (2005).

7. Fritsche, L. G. *et al.* A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **48**, 134–143 (2015).
8. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
9. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl. Acad. Sci. U. S. A.* **111** (2014).
10. Shi, M. & Weinberg, C. R. How much are we missing in SNP-by-SNP analyses of GWAS?. *Epidemiology* **22**, 845 (2011).
11. Bateson, W., Bateson, W. & Mendel, G. *Mendel's Principles of Heredity* (University Press, 1913). <https://doi.org/10.5962/bhl.title.21089>.
12. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97 (1931).
13. Fisher, R. A. XV: The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **52**, 399–433 (1919).
14. Carlborg, Ö. & Haley, C. S. Epistasis: too often neglected in complex trait studies?. *Nat. Rev. Genet.* **5**, 618–625 (2004).
15. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353** (2016).
16. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
17. Koch, L. Insights into missing heritability. *Nat. Rev. Genet.* **15**, 218–218 (2014).
18. Moore, J. H. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum. Hered.* **56**, 73–82 (2003).
19. Moore, J. H. & Williams, S. M. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *BioEssays* **27**, 637–646 (2005).
20. Tsai, C. T. *et al.* Renin-angiotensin system gene polymorphisms and atrial fibrillation. *Circulation* **109**, 1640–1646 (2004).
21. Li, Y. *et al.* Genetic interaction analysis among oncogenesis-related genes revealed novel genes and networks in lung cancer development. *Oncotarget* **10**, 1760–1774 (2019).
22. Gregersen, J. W. *et al.* Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443**, 574–577 (2006).
23. Cordell, H. J., Todd, J. A., Bennett, S. T., Kawaguchi, Y. & Farrall, M. Two-locus maximum lod score analysis of a multifactorial trait: joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. *Am. J. Hum. Genet.* **57**, 920 (1995).
24. Cox, N. J. *et al.* Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat. Genet.* **21**, 213–215 (1999).
25. Cho, J. H. *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: Evidence for epistasis between 1p and IBD1. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7502–7507 (1998).
26. Hivert, V. *et al.* Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *Am. J. Hum. Genet.* **108**, 786–798 (2021).
27. Horlbeck, M. A. *et al.* Mapping the genetic landscape of human cells. *Cell* **174**, 953–967.e22 (2018).
28. Rauscher, B. *et al.* Toward an integrated map of genetic interactions in cancer cells. *Mol. Syst. Biol.* **14**, e7656 (2018).
29. Wei, W. H. *et al.* Genome-wide analysis of epistasis in body mass index using multiple human populations. *Eur. J. Hum. Genet.* **20**, 857–862 (2012).
30. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinform.* **12**, 1–16 (2011).
31. Wang, H., Bennett, D. A., De Jager, P. L., Zhang, Q. Y. & Zhang, H. Y. Genome-wide epistasis analysis for Alzheimer's disease and implications for genetic risk prediction. *Alzheimers Res. Ther.* **13**, 1–13 (2021).
32. Campbell, R. F., McGrath, P. T. & Paaby, A. B. Analysis of epistasis in natural traits using model organisms. *Trends Genet.* **34**, 883 (2018).
33. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet.* **23**, 120–133 (2021).
34. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
35. Yang, Y. *et al.* Combined effect of FTO and MC4R genepolymorphisms on obesity in children and adolescents in Northwest China: a case-control study. *Asia Pac. J. Clin. Nutr.* **28**, 177–182 (2019).
36. Song, Z. *et al.* Evaluation of the obesity genes FTO and MC4R for contribution to the risk of large artery atherosclerotic stroke in a Chinese population. *Obes. Facts* **9**, 353–362 (2016).
37. Ortega-Azorin, C. *et al.* Associations of the FTO rs9939609 and the MC4R rs17782313 polymorphisms with type 2 diabetes are modulated by diet, being higher when adherence to the Mediterranean diet pattern is low. *Cardiovasc. Diabetol.* **11**, 1–12 (2012).
38. Wang, Q. *et al.* Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res.* **66**, 6050–6062 (2006).
39. Pérez-Oliva, A. B., Olivares, C., Jiménez-Cervantes, C. & García-Borrón, J. Mahogunin Ring Finger-1 (MGRN1) E3 ubiquitin ligase inhibits signaling from melanocortin receptor by competition with gas. *J. Biol. Chem.* **284**, 31714 (2009).
40. Overton, J. D. & Leibel, R. L. Mahoguninoid and Mahogany Mutations rectify the obesity of the yellow mouse by effects on endosomal traffic of MC4R protein. *J. Biol. Chem.* **286**, 18914–18929 (2011).
41. Phan, L. K., Lin, F., LeDuc, C. A., Chung, W. K. & Leibel, R. L. The mouse mahoguninoid coat color mutation disrupts a novel C3HC4 RING domain protein. *J. Clin. Invest.* **110**, 1449–1459 (2002).
42. Phan, L. K., Chung, W. K. & Leibel, R. L. The mahoguninoid mutation (Mgnr1md) improves insulin sensitivity in mice with mutations in the melanocortin signaling pathway independently of effects on adiposity. *Am. J. Physiol. Endocrinol. Metab.* **291**, E611–E620 (2006).
43. Wang, Y. *et al.* A novel member of the Rhomboid family, RHBDD1, regulates BIK-mediated apoptosis. *Cell. Mol. Life Sci.* **65**, 3822–3829 (2008).
44. Lin, A., Wang, R. T., Ahn, S., Park, C. C. & Smith, D. J. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res.* **20**, 1122–1132 (2010).
45. Fernandes-da-Silva, A. *et al.* Endoplasmic reticulum stress as the basis of obesity and metabolic diseases: focus on adipose tissue, liver, and pancreas. *Eur. J. Nutr.* **60**, 2949–2960 (2021).
46. Hoffmann, T. J. *et al.* A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics* **210**, 499–515 (2018).
47. Fawcett, K. A. & Barroso, I. The genetics of obesity: FTO leads the way. *Trends Genet.* **26**, 266 (2010).
48. Laber, S. *et al.* Linking the FTO obesity rs1421085 variant circuitry to cellular, metabolic, and organismal phenotypes in vivo. *Sci. Adv.* **7**, eabg0108 (2021).
49. Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.* **348**, 1085–1095 (2003).
50. Lotta, L. A. *et al.* Human gain-of-function MC4R variants show signaling bias and protect against obesity. *Cell* **177**, 597–607.e9 (2019).
51. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
52. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
53. Hemani, G. *et al.* Retracted article: Detection and replication of epistasis influencing transcription in humans. *Nature* **508**, 249–253 (2014).

54. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948. <https://doi.org/10.1038/ng.686> (2010).
55. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2008).
56. Zhang, X., Zou, F. & Wang, W. Efficient algorithms for genome-wide association study. *ACM Trans. Knowl. Discov. Data (TKDD)* **3**, 1–28. <https://doi.org/10.1145/1631162.1631167> (2009).
57. Zhang, X., Pan, F., Xie, Y., Zou, F. & Wang, W. COE: a general approach for efficient genome-wide two-locus epistasis test in disease association study. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **5541 LNBI**, 253–269 (2009).
58. Zhang, X., Huang, S., Zou, F. & Wang, W. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics* **26**, i217 (2010).
59. Ritchie, M. D. *et al.* Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* **69**, 138–147 (2001).
60. Lou, X. Y. *et al.* A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am. J. Hum. Genet.* **80**, 1125 (2007).
61. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
62. Zhang, Y. & Liu, J. S. Bayesian inference of epistatic interactions in case-control studies. *Nat. Genet.* **39**, 1167–1173 (2007).
63. Wang, Y., Liu, X., Robbins, K. & Rekaya, R. AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm. *BMC Res. Notes* **3**, 1–8 (2010).
64. Wan, X. *et al.* Predictive rule inference for epistatic interaction detection in genome-wide association studies. *Bioinformatics* **26**, 30–37 (2009).
65. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
66. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **13**, 1–12 (2021).
67. Oscanoa, J. *et al.* SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucl. Acids Res.* **48**, W185–W192 (2020).
68. Heinz, S. *et al.* Article simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2010.05.004> (2010).
69. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017** (2017).
70. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucl. Acids Res.* **38**, W214 (2010).

Acknowledgements

We thank members of the BK lab: Mr. Vishak Madhwaraj Kadambalithaya for help in preparing Figure 1) and Ms. Sree Vishmaya V for formatting of references for the revised version of the manuscript. Figures 3 and 4 were generated with tissue expression data from GTEx. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on February 11 2021, and March 17 2022. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California. We are grateful to UK Biobank participants. Some data used in this research this research has been conducted using data from UK Biobank, a major biomedical database (www.ukbiobank.ac.uk). UK Biobank is a large-scale biomedical database and research resource containing genetic, lifestyle and health information from half a million UK participants. UK Biobank’s database, which includes blood samples, heart and brain scans and genetic data of the 500,000 volunteer participants, is globally accessible to approved researchers who are undertaking health-related research that’s in the public interest. UK Biobank recruited 500,000 people aged between 40 and 69 years in 2006–2010 from across the UK. With their consent, they provided detailed information about their lifestyle, physical measures and had blood, urine and saliva samples collected and stored for future analysis. UK Biobank’s research resource is a major contributor in the advancement of modern medicine and treatment, enabling better understanding of the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses—including cancer, heart diseases and stroke. Since the UK Biobank resource was opened for research use in April 2012, over 20,000 researchers from 90+ countries have been approved to use it and more than 2000 peer-reviewed papers that used the resource have now been published. UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the British Heart Foundation, Cancer Research UK, Department of Health, North-west Regional Development Agency and Scottish Government. The organisation has over 150 dedicated members of staff, based in multiple locations across the UK. Data used in preparation of this article were obtained from

the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Author contributions

B.K. conceived and designed the study. S.D.S., S.C., and B.K. performed the data analysis. S.D.S., S.C., and B.K. wrote the manuscript. S.D.S., S.C., and B.K. prepared the figures. All the authors read and approved the final manuscript.

Funding

We thank the funding agencies Science & Engineering Research Board (SERB), Government of India (ECR/2018/001429) and Department of Biotechnology, Government of India (BT/RLF/29/2016) for funding equipment and manpower for this study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11270-0>.

Correspondence and requests for materials should be addressed to B.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022