Polymorphic inversions underlie the shared genetic susceptibility of obesity-related diseases

Juan R González[1-4,*], Carlos Ruiz-Arenas[1,2], Alejandro Cáceres[1-3], Ignasi Morán[5], Marcos López-Sánchez[6,7], Lorena Alonso[5], Ignacio Tolosana[1], Marta Guindo-Martínez[5], Josep M Mercader[8-10,5], Tonu Esko[11,12], David Torrents[5,13], Josefa González[14], Luis A Pérez-Jurado[2,6,7,15]

1. Barcelona Institute for Global Health (ISGlobal), Barcelona 08003, Spain
2. IMIM (Hospital del Mar Research Institute), Barcelona 08003, Spain.
3. Centro de Investigación Biomédica en Red en Epidemiologia y Salud Pública (CIBERESP), Barcelona 08003, Spain.
4. Department of Mathematics, Universitat Autònoma de Barcelona (UAB), Barcelona 08193, Spain
5. Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center (BSC-CNS), Barcelona 08034, Spain.
6. Department of Experimental and Health Sciences (CEXS), Universitat Pompeu Fabra, Barcelona 08003, Spain.
7. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona 08003, Spain.
8. Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
9. Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
10. Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
11. Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.
12. Institute of Molecular and Cell Biology, University of Tartu 51010, Tartu. Estonia.
13. Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona 08003, Spain
14. Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona 08003, Spain
15. Women's and Children's Hospital, South Australian Health and Medical Research Institute & University of Adelaide, Adelaide 5005, Australia

***Corresponding author:**
Juan R González
Bioinformatics Research Group in Epidemiology
Barcelona Institute for Global Health (ISGlobal)
Avd. Dr Aiguader, 88
08003 Barcelona, Spain
Phone: +34 932147327
e-mail: juanr.gonzalez@isglobal.org

Abstract

The burden of several common diseases including obesity, diabetes, hypertension, asthma, and depression is increasing in most world populations. However, the mechanisms underlying the numerous epidemiological and genetic correlations among these disorders remain largely unknown. We investigated whether common polymorphic inversions underlie the shared genetic influence of these disorders. We performed an inversion association analysis including 21 inversions and 25 obesity-related traits, on a total of 408,898 Europeans, and validated the results in 67,299 independent individuals. Seven inversions were associated with multiple diseases while inversions at 8p23.1*,* 16p11.2 and 11q13.2 were strongly associated with the co-occurrence of obesity with other common diseases. Transcriptome analysis across numerous tissues revealed strong candidate genes of obesity-related traits. Analyses in human pancreatic islets indicated the potential mechanism of inversions in the susceptibility of diabetes by disrupting the cis-regulatory effect of SNPs from their target genes. Our data underscore the role of inversions as major genetic contributors to the joint susceptibility to common complex diseases.

KEYWORDS: genetic inversions, obesity-related diseases, common diseases, obesity, diabetes, hypertension, asthma, human traits, disease co-occurrence.

INTRODUCTION

Obesity is a disorder with increasing but non-uniform prevalence in the world population and one of the major public health burdens[1]. Obesity (MIM: 615812) derived morbidity and years of life lost strongly associate to a broad range of highly prevalent diseases, including type 2 diabetes (MIM:125853), cardiovascular disease (MIM: 608901), asthma (MIM: 600807) and (neuro)psychological disturbance such as depression (MIM:608516) or intellectual disability, among others[2]. While the causes underlying the multiple co-occurrences of obesity are likely complex and diverse, common mechanisms underlying these comorbidities, which are potential targets for preventive or therapeutic intervention, are largely unknown.

One of the possible genetic mechanisms of comorbidity can be through rare copy number variants (CNVs), which are more prevalent in people with some severe forms of obesity[3,4] and might confer at least part of the increased risk for obesity via developmental delay[5]. Most of these findings have been described in pediatric obesity.[6,7]

Genomic inversions, copy-neutral changes in the orientation of chromosomal segments with respect to the reference, are (also) excellent candidates for being important contributors to the genetic architecture of common diseases. Inversion polymorphisms can alter the function of the including and neighboring genes by multiple mechanisms, disrupting genes, separating their regulatory elements, affecting chromatin structure, and maintaining a strong linkage of functional variants within an interval that escape recombination. Therefore, by putatively affecting multiple genes in numerous ways, inversions are important sources of shared genomic variation underlying different human diseases and traits. Consequently, human inversions show genetic influences in multiple phenotypes. For instance, the common inversion at 8p23.1 has been independently linked to obesity[8], autism (MIM: 209850)[9], neuroticism (MIM: 607834)[10] and several risk behavior traits[11], while inversion at 17q21.31 has been associated with Alzheimer (MIM: 607822)[12] and Parkinson (MIM: 168600) [13] diseases, heart failure[14] and intracranial volume[15]. We previously reported a ~40% of population attributable risk for the co-occurrence of asthma and obesity given by a common inversion polymorphism at 16p11.2[16]. In addition, transcriptional effects have been documented in several tissues for inversions at 17q21.31[13,17], and 16p11.2[16].

It is estimated that each human genome contains about 156 inversions[18]. Therefore, inversions constitute a substantial source of genetic variability. Many of those polymorphic inversions show signatures of positive or balancing selection associated with functional effects[19]. However, the overall impact of polymorphic inversions on human health remains largely unknown because they are difficult to genotype in large cohorts. We overcame this limitation by recently reporting a subset of 20 inversions that can be genotyped with SNP array data as they are old in origin, low or not recurrent and frequent in the population[20]. We have also included an additional inversion in our catalog, 16p11.2, previously validated and genotyped in diverse populations[16]. Three of the inversions are submicroscopic (0.45-4 Mb), flanked by large segmental duplications and contain multiple genes. Five are small (0.7-5 Kb) and intragenic, and 13 are intergenic of variable size (0.7-90 Kb) but highly enriched in pleiotropic genomic regions[21]. While this is clearly not a comprehensive set of inversions, it is probably the largest set that can be genotyped in publicly available datasets.

In this manuscript, we aimed to study the association of 21 common polymorphic inversions in Europeans with highly prevalent co-morbid disorders and related traits. We particularly aimed to decipher the role of inversions in known epidemiological co-occurrences with obesity such as diabetes, hypertension (MIM: 145500), asthma and mental diseases like depression, bipolar disorder (IMIM: 125480) or neuroticism. For significant associations, we investigated whether causal pathways could be established and the most likely underlying mechanisms.

MATERIAL AND METHODS

Discovery Dataset

The UK Biobank (UKB) is a population-based cohort involving 500,000 individuals aged between 37 and 73 years, recruited across UK in the period 2006-2010. Further details on the quality control and genotyping are described in the study design[22]. Phenotypic information is recorded via questionnaires and interviews (e.g., demographics and health status) and SNP genotypes were generated from the Affymetrix Axiom UK Biobank and UKBiLEVE arrays. We based our study on 408,898 individuals from European descent and from whom inversion genotypes were called using SNP array data. Principal components computed by the UK Biobank (data-field 22009) were used in the analyses to control for population stratification.

Replication Datasets

Different public datasets with access grant to the co-authors were used to attempt to replicate our positive findings in the association studies (**Figure 1**). The next sections describe these resources.

Genetic Epidemiology Research on Aging (GERA)

The GERA cohort (dbGaP Study Accession: phs000674.v1.p1) consists on a cohort of over 100,000 adults from the Northern California Region (USA). Only individuals with reported race (variable phv00196837.v2.p2) equal to white were selected for the analyses (n=56,638). The resulting studied cohort is 40% male, 60% female, and ranges in age from 18 to over 100 years old with an average age of 64 years at the time of the survey (2007). Individuals were genotyped with Affymetrix Axiom_KP_UCSF_EUR. After quality control of the inversion genotyping calling process a total of 53,782 individuals with information about sex, age, principal components for genetic ancestry and several diseases including obesity (9,439 cases), diabetes (6,529 cases), hypertension (27,009 cases), asthma (8,716 cases) and depression (6,924 cases) were used in the replication studies.

70KforT2D: diabetes and obesity

The 70KforT2D study (70KT2D)[23] includes 5 datasets publicly available in dbGAP or EGA: NuGENE, FUSION, GENEVA, WTCCC and GERA. Notice that 70KforT2D include cases diagnosed with diabetes and obesity from the GERA cohort. We used information about being diabetic or not as describe elsewhere[23]. The 5 datasets were used to attempt to replicate the significant findings in the UK Biobank data on diabetes. The WTCCC dataset was removed from the obesity and obesity/diabetes analysis since we did not have access to body mass index (BMI) information for that study. The GERA dataset was split in two (GERA1 and GERA2) to speed up the imputation and inversion calling procedure since it is a large dataset. After performing QC on inversion genotypes, a total of 67,299 individuals were used in the replication step (54,801 controls and 12,498 diabetic). Data was accessed from the portal cg.bsc.es/70kfort2d.

The obesity variable was created using the body mass index (BMI) variable. We considered control individuals those having BMI in the interval (18.5–24.9) and obese people those having BMI>30.0. For obesity associations, we excluded individuals with diabetes. As a result, a total of 34,316 individuals (23,818 controls and 10,498 obese) were used for that purpose. The co-occurrence of obesity and diabetes was studied by comparing individuals with no obesity and no diabetes as the reference category with individuals being obese and diabetic simultaneously. This ended up with a total of 23,818 control and 5,715 obese/diabetic individuals. Next, we further describe the studies included in the 70KT2D dataset along with their accession numbers.

Northwestern NUgene Project: Type 2 Diabetes (NUGENE) (dbGaP Study Accession: phs000237.v1.p1) contains data from individuals from the Northwestern University Medical Center (USA). For this study, T2D cases were included if they had been diagnosed of Type 2 Diabetes, they took drugs to treat Type 2 Diabetes or they presented abnormal diabetes-related blood measures. Controls were included if they had not been diagnosed of Type 2 Diabetes, they did not take drugs to treat Type 2 Diabetes, they presented normal diabetes-related blood measures and they did not have any family history of diabetes (either Type 1 or Type 2). In both groups, subjects with Type 1 Diabetes were excluded. These individuals were genotyped with Illumina Human1M-Duov3_B.

The Finland-United States Investigation of NIDDM Genetics - GWAS Study (FUSION) (dbGaP Study Accession: phs000100.v4.p1) aims to investigate the association between genetics and Type 2 Diabetes in Finish families. For this study, cases were included if they had been diagnosed of type 2 diabetes, they took drugs to treat type 2 diabetes or they presented abnormal diabetes-related blood measures. Controls were included if they presented normal diabetes-related blood measures and were frequency matched to the cases by age, sex and birth province. In both groups, individuals with family history of type 1 diabetes were excluded. These individuals were genotyped with Illumina HumanHap300v1.1.

GENEVA Genes and Environment Initiatives in Type 2 Diabetes (Nurses' Health Study/Health Professionals Follow-up Study) (dbGaP Study Accession: phs000091.v2.p1) is a nested case-control (2,720 cases and 3,180 controls) study from two USA female cohorts: the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS) with a mean age of 57 ranging from 40 and 78. These individuals were genotyped with Affymetrix AFFY_6.0.

Geographical variation in Europe

POPRES project (dbGaP Study Accesion: phs000145.v4.p2 access granted to the authors) was used to estimate inversion frequencies in European countries and regions. This project aimed to facilitate exploratory genetic research by assembling a DNA resource from a large number of subjects participating in multiple studies throughout the world. We selected European individuals (variable phv00173964.v2.p2) leading a total of 3,071 samples. A geographic label (North, Center, South) was assigned to each individual using information of variable phv00066613.v2.p2.

Transcriptomic analyses

GTEx Analysis

We associated the 21 chromosomal inversions to changes in gene expression in GTEx project. We determined inversion genotypes on the GTEx v7 genotype calls from dbGAP (dbGaP Study

Accession: phs000424.v7.p2 accession granted to the authors). We only included samples classified as European with a confidence higher than 90% by peddy[24]. Inv3_003 was discarded as the calling was not confident. Gene expression counts from RNA-seq data were downloaded using recount2[25]. We computed the association between gene expression and inversions using voom[26] and limma[27]. The linear model included the inversion coded as additive (0: NN, 1: NI, 2: II) and the same covariates than GTEx (first three genome-wide PCA components, sex and covariates from PEER). In each tissue, we selected those features having more than 10 counts in at least 10% of the samples. We corrected the association results per tissue for multiple comparisons by using a false discovery rate (FDR) adjusted p-value per tissue.

EGCUT Biobank

Estonian Gene Expression Cohort was used to attempt to replicate positive transcriptomic results found in GTEx. The cohort is composed of 1,048 randomly selected samples (mean age 37+/-16.6 years; 50% females) from the cohort of 53,000 samples in the Estonian Genome Center Biobank, University of Tartu. Whole-Genome gene-expression levels from whole blood RNA were obtained by Illumina HT12v3 arrays according to manufacture's protocols. Low quality samples were excluded. All probes with primer polymorphisms were discarded, leaving 34,282 probes. Raw gene expression data was Log-Quantile normalized using MixupMapper software. DNA was genotyped with Human370CNV array.

Pancreatic Islets

We analyzed the transcriptomic effect of inversions 8p23.1 and 16q11.2 on 118 pancreatic human islet samples using RNA-sequencing counts and high-density genotyping data[28]. DNA genotype data (EGA accession number EGAS00001001261) was used to call inversion genotypes using *scoreInvHap*, then the association between gene expression and inversions was assessed using voom[26] and limma[27]. Only genes in the inversion regions were analyzed and un-corrected p-values were reported as a measure of association.

Positional analyses

For the positional analyses, several annotations were gathered from the following sources: TAD boundaries from the Human ES Cell (H1) topological domains[29]; promoters, enhancers, CTCF-peaks and ATAC-seq open-chromatin regions from the human islet regulome annotation[30]; islet-specificity scores were calculated using the gene expression data from[31]; eQTL SNP-gene associations from[28,32]. The chromatin landscape coverage percentage was calculated using a sliding window of 500kb and 1Mb for inversions 8p23.1 and 16p11.2 respectively, using steps of 1% of the window size, and calculating the percentage of covered nucleotides by significant signal in each of the categories. For the islet-specific expression analysis, we calculated the non-islet median expression level and difference between the 75 and 50 quartiles, and considered as islet-specific any gene that was expressed in islet >3 quartiles over the median of non-islet expression. Visualization was done in python3 using the matplotlib graphics library.

Statistical methods

SNP imputation and inversion calling

SNP microarray data was imputed with *imputeInversion* pipeline prior to inversion calling[33]. This pipeline was designed to impute only those SNPs inside the inversion region or closer than 500 Kb to the inversion breakpoints. This step is recommended before performing inversion calling. *imputeInversion* uses shapeITv2.r904 to phase[34], Minimac3[35] to impute and 1000 Genomes as reference haplotypes. Variants with an imputation R2 < 0.3 were discarded. Genotype probabilities were used to call inversions using *scoreInvHap*[20] which is available at Bioconductor. *scoreInvHap* computes a similarity score between an individual's alleles and the reference alleles in each chromosomal status. We used the development version of *scoreInvHap*, which includes references for 21 inversions. These methods were used to perform inversion calling of discovery and replication studies as well as individuals from POPRES.

Inversion frequencies

Inversion frequencies were estimated in UKB and POPRES studies using SNPassoc package[36]. A trend test implemented in the R function *prop.trend.test* was used to assess whether inversion frequencies in European regions from POPRES (North, Center, South) showed a significant cline. Principal component analysis was used to visualize inversion frequencies across European regions of POPRES dataset.

Obesity and obesity co-occurrence traits

Obesity trait was created using body mass index (BMI) information. First, BMI was categorized in 5 categories using World Health Organization (WHO) classification which considers the following categories: underweight (BMI below 18.5), normal weight (BMI between 18.5 and 25), pre-obesity (BMI between 25 and 29.9), obesity class I (BMI between 30 and 34.9) and obesity class II and III (BMI above 35). Obesity was considered as obesity class I, II and III and was compared with normal weight category. The analysis of obesity co-occurrence with diabetes, hypertension, asthma, depression and neuroticism was performed by comparing individuals with normal weight and no presence of the disease with individuals being obese and having the disease of interest.

Inversion association analyses

Each inversion was independently associated with all the traits by using generalized linear models implemented in SNPassoc package[36]. The models were adjusted for gender, age and the first four principal components obtained from GWAS data in order to control for population genetic differences. The inversions were analyzed using an additive model. Multiple comparison problem was addressed by correcting for the total number of inversions and the phenotypes analyzed by considering the effective number of tests (18 independent tests) using Li and Li method[37] that accounts for correlation among traits. This ended up with a corrected p-value equal to 0.00128.

Causal inference

Mediation analysis using *mediation* R package[38] was used to evaluate whether inversion 8p23.1 mediates the association between obesity and diabetes. Additive Bayesian network models using *abn* R package[39] were used to determine optimal Bayesian network models to identify statistical dependencies between inversions 8p23.1, 16p11.2 and 11q13.2 and obesity, diabetes and hypertension in the UKB dataset, and validated in the GERA cohort. The most probable network structure was estimated using exact order-based approach as implemented in the *mostprobable* function of *abn* package.

Data availability

The data used in this work were obtained from publicly available datasets that are accessible through public repositories: UKB study, dbGaP, EGA, GTeX and GEO. The inversion calling of UKB samples will be available through their platform. The inversion calling for the other samples are available upon request. The complete transcriptomic summary statistics of the 21 inversions are also available upon request.

RESULTS

Frequency and stratification of inversions in European populations

Using *scoreInvHap,* we first called the inversion status of individuals from the UK Biobank (UKB) with European ancestry (n=408,898). We confirmed the previously reported frequency in the 1000 Genome project of the 21 inversions analyzed in this work (**Table 1**). As inversion frequencies have a strong demographic effect, we also analyzed 12 European countries from the POPRES study (**Figure S1).** We observed significant clines along north-south latitude for several inversions (**Table 1 and Figure**

S2A) as well as subtle ancestral differences (**Figure S2B**). Thus, population stratification was considered when performing association analyses as explained in the methods section.

Inversions at 8p23.1, 16p11.2 robustly associate with obesity and obesity-related traits

The discovery phase of the study used data from UKB. We performed association analyses between the 21 inversions with obesity and co-morbid diseases and traits (see **Methods**). These include obesity, diabetes, stroke, hypertension, asthma, chronic obstructive pulmonary disease (COPD), depression and bipolar disorder, along with related traits or phenotypes classified as morphometric (4 traits), metabolic (5 traits), lipidic (2 traits), respiratory (3 traits) and behavioral (3 traits) (**Figure 2**). **Table S1** shows the total number of cases and controls used to perform the association analyses on each trait. The significant associations were further validated in the GERA independent dataset that contains information about several diseases. Positive results found in diabetes were validated in the 70KT2D dataset, which includes GERA among others (NUGENE, FUSION, GENEVA and WTCCC) (see **Methods** and **Figure 1** which describes the comprehensive data analysis performed in the different datasets).

The analyses on the UKB revealed several genetic influences of inversions on obesity and related common diseases (**Figure 2**). We observed a total of 74 significant associations after correcting for the number of inversions analyzed and the effective number of tests to consider the multiple analyzed traits (see **Methods**). In general, we observed higher numbers of associations and stronger effects for the largest inversions at 8p23.1, 16p11.2 and 17q21.31, consistent with the fact that they encapsulate more genes. Some smaller inversions such as the ones at 11q13.2 and Xq13.2 also showed notable effects such as shared susceptibility and strength, respectively. We found a prominent inflation of association suggesting common genetic influences of the inversions across multiple phenotypes (**Figure S3A**). Some of the associations found have already been reported such as those at inversion 8p23.1 with obesity[8] and neuroticism[10] and the one with inversion 16p11.2 with obesity[16].

As a summary of the relevant findings, we observed that inversions at 8p23.1, 16p11.2 and 11q13.2 are all strongly associated with several obesity-related diseases (**Figure 2**). Remarkably, the non-inverted (N) allele of inversion 8p23.1 (i.e. the risk allele) is independently associated with diabetes (OR=1.04, p=1.1x10^{-3}), hypertension (OR=1.04, p=7.0x10^{-16}) and asthma (OR=1.03, p=7.0x10^{-5}) (**Table 2**). The association with diabetes was replicated in the 70KT2D study (**Figure 3A**) (OR =1.08, p=1.1x10^{-8}) as well as the association with obesity (OR=1.08, p=5.6x10^{-6}) and the association with hypertension, which was validated in the GERA study (OR=1.03, p=0.0183) (**Table 2**). We also found a significant association between the non-inverted (N) allele of inversion 16p11.2 and obesity (OR=1.05, p=3.9x10^{-24}) that was replicated in GERA study (OR=1.07, p=1.4x10^{-4}). The significant association found in the UKB for the inversion 11q13.2 was not validated in the GERA study (OR=1.03, p= 0.0712). Consistently, the analysis of UKB study also revealed association of inversions at 8p23.1 and 16p11.2 with different obesity-related traits such as body mass index (BMI), waist circumference, high density lipoprotein (HDL) or systolic and diastolic blood pressure, among others (**Figure 2**).

Some interesting associations in the discovery sample included those of inversion 17q21.31 with HDL, waist circumference, waist-hip ratio and systolic and diastolic blood pressure (**Figure 2**). Interestingly, this inversion also showed a significant role in behavioral traits such as mood swing, depression and bipolar disorder, which would need further validation. While we also found significant association of the inversion 6p21.33 with asthma (OR=1.02, p=0.0215) and different respiratory capacity traits (FEV1, p= 3.4x10^{-9} and FVC, p=3.2x10^{-9}) the association with asthma was not replicated in the GERA study. The inversion 7q11.2 was associated with different morphometric traits (BMI, waist circumference and waist-to-hip ratio) and will require further validation studies.

Inversions at 8p23.1, 16p11.2 and 11q13.2 are more strongly associated with the co-occurrence of diseases than with single diseases

Remarkably, the N-allele of the inversion 8p23.1 was significantly associated with the co-occurrence of obesity with diabetes (OR=1.08, p=3.1x10^{-7}), hypertension (OR=1.07, p=1.7x10^{-16}) or asthma (OR=1.08, p=3.0x10^{-11}). These results were validated in the GERA and 70KT2D (**Table 2**). For obesity/diabetes we observed an OR=1.17 (p=1.4x10^{-13}) (**Table 2** and **Figure 3C**) and none of the SNPs located within the inverted region were significantly associated at a genome-wide level (minimum p = 3.8x10^{-5}) (**Figure S4A**). Finally, we also found a significant association of the N-allele of inversion 11q13.2 with the co-occurrence of obesity with diabetes (OR=1.05, p=0.0011) and hypertension (OR=1.03, p=2.9x10^{-5}) (**Figure 2**), which was not validated in the GERA study.

The study of inversion 16p11.2 also revealed some new significant associations between the inversion and the co-occurrence of obesity with several diseases (**Figure 2).** The co-occurrence with diabetes at UKB (OR=1.06, p=7.5x10^{-5}) was independently replicated in the 70KT2D study (OR=1.13, p=1.2x10^{-8}), where none of the SNPs located within the inverted region were significantly associated at a genome-wide level (minimum p: 0.0214) (**Figure S4B**). In addition, the significant co-occurrence with hypertension observed in the UKB study (OR=1.06, 2.7x10^{-14}) was validated in the GERA study (OR=1.05, p=0.0357) further confirming the robustness of these findings (see **Table 2** reporting the effect of the risk allele N).

In order to further illustrate that the association of the inversion is not driven by single variants, we downloaded data from the GWAS catalog and checked whether the GWAS signals for the analyzed traits are associated (i.e. tags) with the inversions. No tag-SNPs for any of these traits were found. In particular, the results for the three inversions associated with the co-occurrence of obesity with other traits showed the following results: the median R^2 between SNPs in the 8p23.1 region and the inversion was 0.36 (IQR: 0.17-0.46), 0.71 (IQR: 0.62-0.89) for the inversion 16p11.2, and all the SNPs are not associated (i.e linkage equilibrium) ($R^2 < 0.06$) for the inversion 11q13.2.

Regulatory region and gene disruption are the mechanisms underlying the effect of inversions on diabetes

To investigate the possible mechanisms underlying the shared genetic influences of the inversions with obesity and its co-morbidities, we analyzed the transcriptional effects of the 21 inversions on different tissues from the GTEx project (see **Methods**). As a result of these analyses, we found that inversion 8p23.1 modulated the transcription in brain, pancreas and adipose tissue of the pseudogene *FAM86B3P* (HGNC: 44371), as well as the genes *MFHAS1* (MIM: 605352)*, IL19* (MIM: 605687)*, HAND2* (MIM: 602407)*, FDFT1* (MIM: 184420)*, FAM167A* (MIM: 610085)*, ERI1* (MIM: 608739)*, CHAC1* (MIM: 614587)*, CCL22* (MIM: 602957)*, CCL19* (MIM: CCL19) *and BLK* (MIM: 191305) in other tissues (**Figure 3D**). Genes *FDFT1* (MIM: 184420)*, C8orf13* (MIM: 610085)*, CLDN23* (MIM: 609203)*, NEIL2* (MIM: 608933), *MTMR9* (MIM: 606260), *MSRA* (MIM: 606260)*, BLK* (MIM: 191305) and were also differentially expressed in blood samples from the validation study we performed in the independent general population cohort belonging to EGCUT Biobank (**Figure 3E**). For the inversion 16p11.2 we found a total of 30 genes differentially expressed at 5% FDR level in blood, brain, pancreas or adipose tissue including *TUFM* (MIM: 602389)*, SULT1A2* (MIM: 601292)*, SPNS1* (MIM: 612583)*, EIF3CL* (MIM: 603916)*, FOXO1* (MIM: 136533) among many others (**Table S2**). These results were also observed in the blood samples of the validation cohort from EGCUT Biobank (**Figure S5**). The genes affected by the other inversions and the different tissues can be found in **Table S2**.

Inversions 8p23.1 and 16p11.2 affect key genes associated with diabetes in pancreatic islets

We conducted a more detailed analysis of gene expression on a relevant tissue to support the association on diabetic/obese individuals. We first genotyped the inversions and analyzed RNA sequencing in human pancreatic islets from 89 deceased donors (see **Methods**). This revealed a significant association between inversion 8p23.1 and the expression levels of *CLDN23* ($p=1.3x10^{-3}$) and *ERI1* (p=0.0356). We observed a nominally significant interaction of inversion 8p23.1 with obese/diabetic status associated with the expression of lncRNA *FAM66A* (HGNC: 30444) (p= 0.0254), where individuals carrying the risk allele for obesity and diabetes also present *FAM66A* downregulation (**Figure 3F**). In addition, results with inversion at 16p11.2 also revealed a significant interaction between the inversion and obese/diabetic status for the expression of *NUPR1* (MIM: 614812) (p= 0.0116) and *ATXN2L* (MIM: 607931) (p= 0.0167) (**Figure S6**).

Cis-regulatory SNPs are disrupted by breakpoints of inversions 8p23.1 and 16p11.2

We also investigated whether the positional effects of the inversions could be associated with diabetes (see **Methods**). **Figure 4A** shows the chromatin landscape of the region of the inversion 8p23.1 as well as the location of all genes having a significant alteration of expression, including those that are islet-specifically expressed. A cluster of islet-specific genes is located outside the rightmost boundary of the inversion but inside the inversion's topologically associated domains (TAD). Therefore, it is likely that the regulatory regions of these genes lie across the inversion's boundary, and thus their *cis*-regulatory SNPs being separated from their target genes by the right breakpoint of the inversion 8p23.1 in the case of genes *FAM66A* and *FAM66D* (HGNC: 24159) (**Figure 4A**). Similarly, the analysis of the inversion 16p11.2 also revealed four eQTLs in which the *cis*-regulatory SNPs were separated from their target genes by the inversion breakpoints: *TUFM*, *SULT1A1* (MIM: 171150), *EIF3C* (MIM: 603916), *EIF3CL* (**Figure 4B**). The *EIF3CL* gene is disrupted by the inversion breakpoint providing a different mechanism of action for that gene (**Figure 4B**).

Obesity mediates the association of inversions with diabetes and hypertension

We first aimed to disentangle the shared genetic influence of the inversion 8p23.1 in obesity and diabetes. To this end, a Bayesian network analysis was performed on the discovery study (see **Methods**). Based on the BIC, the most likely model was for the sequence inv8p23.1 -> obesity -> diabetes, suggesting a mediatory effect of obesity on the association between the inversion and diabetes (**Figure 5A**). The same network was obtained in the GERA cohort. This was consistent with mediation analyses showing that 38.7% (CI95%: 25.2-59.0%) of the diabetes risk variance explained by the inversion 8p23.1 was mediated by obesity ($p<10^{-16}$). Then, we also investigated whether inversion 8p23.1, 16p11.2 and 11q13.2 act jointly or not on obesity, diabetes and hypertension. The Bayesian network analysis including the three inversions in the model revealed that the inversions 8p23.1 and 16p11.2 independently associated with diabetes and hypertension being mediated by obesity (**Figure 5B**).

DISCUSSION

Epidemiological studies largely support the co-occurrence of obesity with numerous traits and diseases such as diabetes, hypertension, asthma and psychiatric disorders among others[40,41]. The extent to which obesity is a cause, a consequence or shares common causes with these traits is subject of intense research[42–44]. Here, we show that at least two common polymorphic inversions at 8p23.1 and 16q11.2 offer a genetic substrate to some widely observed co-morbidities of obesity, such as those with diabetes, hypertension, asthma and depression.

The analysis of UKB dataset validated the estimated inversion allele frequencies in European populations reported in our recent analyses[20]. The observed differences of some inversion allele frequencies among major populations could explain part of the existing geographic variability in disease incidence[45]. In particular, the reported cline of the inversion at 8p23.1 and 16p11.2 could capture a proportion of the observed North-South European differences in obesity[46], diabetes and hypertension[47] incidence.

The analysis of our discovery sample also confirmed previous reported associations of inversions with phenotypes, such as neuroticisms for the inversions 17q21.31 and 8p23.1[10], obesity for inversion 8p23.1[8], and the co-occurrence of asthma and obesity with the inversion 16p11.2[16]. In addition, we discovered and robustly validated new associations of the inversion 8p23.1 with diabetes and hypertension as well as the co-occurrences of obesity with diabetes, hypertension and asthma. These results suggest a relevant role of the inversion 8p23.1 in this metabolic syndrome[48].

Our data suggest a causal path in which obesity mediates the observed association between inversions and several complex diseases. In particular, obesity mediates the independent effect of inversions at 8p23.1 and 16p11.2 on diabetes. Transcriptome analyses from general population has revealed candidate genes to mediate this effect, such as *BLK,* involved in pancreatic β-cell insulin metabolism whose rare mutations are associated with young age of onset diabetes[49], or *FDFT1,* linked to C-reactive protein (CRP) and lipids levels[50] and one of the strong candidates for obesity in gene expression networks derived from mouse intercrosses[51]. A more specific analysis of transcriptome and eQTLs on pancreatic islets leads to another interesting gene: *FAM66A*. *FAM66* is a multiple copy non-coding gene located in the flanking segmental duplications of the 8p23.1 inversion breakpoint highly expressed in brain and with low-level expression in pancreas. Diabetic individuals carrying the N-allele have lower gene expression, while no differential expression across inversion genotypes is observed in control individuals. Consistently, allele-specific expression analysis of this gene shows clear differences in expression in pancreatic cells of already symptomatic diabetic subjects. Remarkably, a copy-number gain variant including *FAM66* genes has been associated with increased risk of diabetes[52]. Our positional analyses also pointed out at *FAM66D* (8p23.1) as a candidate since the gene body was split in two by the inversion breakpoint.

We have also shown that inversion at 16p11.2 affects the joint effect of obesity with diabetes and hypertension and that this effect is independent of the effect found for inversion 8p23.1. Also, the odds ratios found for these associations are stronger than those observe when analyse those diseases independently. The functional consequences of this inversion were previously reported to be mediated by deregulation of *TUFM*, *SULT1A1*, *SULT1A2*, *SH2B1* (MIM: 608937), *APOB48R* (MIM: 605220), and *EIF3C* in blood[16]. Position transcriptional analysis in pancreatic islets revealed that *TUFM* and *EIF3C* have their lead eQTL SNPs separated in the inverted allele. Remarkably, the eQTL SNP rs42861 of *TUFM* does not seem to be causal in the centiSNP database[53] suggesting that it is in LD with the causal variant. This SNP is located into the promoter region that is closer to *TUFM* in the inverted haplotypes. This supports the hypothesis that the positional changes made by the inversion can affect

TUFM gene expression and subsequently have an effect in obesity/diabetes increased risk. Positional analyses also pointed out *EIF3CL,* a gene also split in two by the inversion breakpoint, and with some isoforms preferentially expressed in human pancreatic islets[31].

The inversions at 8p23.1 and 16p11.2 were also associated with the joint occurrence of obesity with behavioral traits, in particular with depression. These data further support our hypothesis that polymorphic inversions are strong candidates for the joint genetic susceptibility to co-occurring diseases by simultaneously affecting multiple genes. The observation that some SNPs located in both inversion regions are not or weakly associated with the analyzed traits, while inversion haplotypes are associated even at genome-wide significant level for GWAS, and the strongest association found in people having more than one disease, also confirm that inversions are main contributors to the shared genetic susceptibility of co-occurring diseases. The fact that inverted alleles do not recombine preserving haplotypes in strong linkage disequilibrium highly suggest that the underlying evolutionary genetic event that has maintained or selected functional eQTLs *in cis* in these haplotypes is the inversion. Functional analyses in the appropriate tissue in cases and controls, as the one we performed for obesity and diabetes, will shed light into the genes and mechanisms involved in behavioral or psychiatric traits.

Our hypothesis that inversions underlie the shared genetic susceptibility to common diseases is particularly supported by our findings in large inversions. These inversions encapsulate multiple genes and their associations with phenotypes were highly significant and could be replicated. Smaller inversions showed significant effects for numerous traits in the discovery study but only one result could be confirmed, namely the correlation of inversion at 11q13.2 with obesity and related traits and also with the co-occurrence of obesity, hypertension and diabetes. Similarly, this study opens the door to further association studies of these and other inversions with traits and disorders not studied in this work. Additionally, the large number of significant genes associated with different tissues as well as the significant associations found for some traits also provides good candidate genes for some human diseases that are likely under the influence of inversions. These include, among others, Autism, Alzheimer and Parkinson disease.

In conclusion, we report the largest association study of genomic inversions and human traits that represents a breakthrough for genomic association of comorbid disorders, in which polymorphic inversions were often previously disregarded. Our results underscore the role of some inversions as major genetic contribution to the joint susceptibility to common diseases. The results in obesity and diabetes reveal a mechanism in which cis-regulatory SNPs are separated from their target genes by inversion breakpoints. Our findings set a new framework for future studies which are now accessible to the research community thanks to inversion genotyping tools such as our scoreInvHap method[20].

SUPPLEMENTARY DATA DESCRIPTION

The file **inversions_supplementary_material.pdf** contains supplementary figures S1-S6 and supplementary table S1.

The file **Table_S2.xlxs** contains a the supplementary table S2 with the gene expression data analysis of GTEx and inversions (one chromosome per tab).

ACKNOWLEDGEMENTS

This research has received funding from Ministerio de Ciencia, Innovación y Universidades (MICIU), Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional, UE (RTI2018-100789-B-I00) also through the “Centro de Excelencia Severo Ochoa 2019-2023” Program (CEX2018-000806-S); and the Catalan Government through the CERCA Program and projects SGR2017/801 and #016FI_B 00272 to CR-A. JG is funded by the European Commission (H2020-ERC-2014-CoG-647900) and the MINECO/AEI/FEDER, EU (BFU2017-82937-P). LAPJ lab was funded by the Spanish Ministry of Science and Innovation (ISCIII-FEDER P13/02481), the Catalan

Department of Economy and Knowledge (SGR2014/1468, SGR2017/1974 and ICREA Acadèmia), and also acknowledges support from the Spanish Ministry of Economy and Competiveness “Programa de Excelencia María de Maeztu” (MDM-2014-0370). This research was conducted using the UK Biobank Resource under Application Number 43983. 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.

DECLARATIONS OF INTERESTS

LAP-J is a founding partner and scientific advisor of qGenomics Laboratory. All other authors declare no conflict of interest.

WEB RESOURCES SECTION

Worldwide health organization (WHO): http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi)

Online Mendelian Inheritance in Man (OMIM): http://www.omim.org

GWAS catalog: https://www.ebi.ac.uk/gwas/

REFERENCES

1. Collaborators, T.G. 2015 O. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N. Engl. J. Med. *377*, 13–27.
2. Dixon, J.B. (2010). The effect of obesity on health outcomes. Mol. Cell. Endocrinol. *316*, 104–108.
3. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.
4. Serra-Juhé, C., Martos-Moreno, G.Á., Bou de Pieri, F., Flores, R., González, J.R., Rodríguez-Santiago, B., Argente, J., and Pérez-Jurado, L.A. (2017). Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. PLOS Genet. *13*, e1006657.
5. Kaminsky, E.B., Kaul, V., Paschall, J., Church, D.M., Bunke, B., Kunig, D., Moreno-De-Luca, D., Moreno-De-Luca, A., Mulle, J.G., Warren, S.T., et al. (2011). An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental

disabilities. Genet. Med. *13*, 777–784.
6. Selvanayagam, T., Walker, S., Gazzellone, M.J., Kellam, B., Cytrynbaum, C., Stavropoulos, D.J., Li, P., Birken, C.S., Hamilton, J., Weksberg, R., et al. (2018). Genome-wide copy number variation analysis identifies novel candidate loci associated with pediatric obesity. Eur. J. Hum. Genet. *26*, 1588–1596.
7. Vuillaume, M.L., Naudion, S., Banneau, G., Diene, G., Cartault, A., Cailley, D., Bouron, J., Toutain, J., Bourrouillou, G., Vigouroux, A., et al. (2014). New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. Am. J. Med. Genet. Part A *164*, 1965–1975.
8. Cáceres, A., and González, J.R. (2015). Following the footprints of polymorphic inversions on SNP data: from detection to association tests. Nucleic Acids Res. 1–11.
9. Gutiérrez Arumi, A. (2015). Ancestral genomic submicroscopic inversions of human genome and their relation with multifactorial human diseases. Univ. Pompeu Fabra.
10. Okbay, A., Baselmans, B.M.L.B.M.L., De Neve, J.-E.J.-E., Turley, P., Nivard, M.G.M.G., Fontana, M.A.M.A., Meddens, S.F.W.F.W., Linnér, R.K.R.K., Rietveld, C.A.C.A., Derringer, J., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. *48*, 624–633.
11. Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S.F.W., Wedow, R., Fontana, M.A., Lebreton, M., Tino, S.P., Abdellaoui, A., Hammerschlag, A.R., et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat. Genet. *51*, 245–257.
12. Laws, S.M., Friedrich, P., Diehl-Schmid, J., Müller, J., Eisele, T., Bäuml, J., Förstl, H., Kurz, A., and Riemenschneider, M. (2007). Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer’s disease. Mol. Psychiatry *12*, 510–517.
13. Zabetian, C.P., Hutter, C.M., Factor, S.A., Nutt, J.G., Higgins, D.S., Griffith, A., Roberts, J.W., Leis, B.C., Kay, D.M., Yearout, D., et al. (2007). Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson’s disease. Ann. Neurol. *62*, 137–144.
14. Pilbrow, A.P., Lewis, K.A., Perrin, M.H., Sweet, W.E., Moravec, C.S., Tang, W.H.W., Huising, M.O., Troughton, R.W., and Cameron, V.A. (2016). Cardiac CRFR1 Expression Is Elevated in Human Heart Failure and Modulated by Genetic Variation and Alternative Splicing. Endocrinology *157*, 4865–4874.
15. Ikram, M.A., Fornage, M., Smith, A. V, Seshadri, S., Schmidt, R., Debette, S., Vrooman, H.A., Sigurdsson, S., Ropele, S., Taal, H.R., et al. (2012). Common variants at 6q22 and 17q21 are associated with intracranial volume. Nat. Genet. *44*, 539–544.
16. González, J.R., Cáceres, A., Esko, T., Cuscó, I., Puig, M., Esnaola, M., Reina, J., Siroux, V., Bouzigon, E., Nadif, R., et al. (2014). A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. Am. J. Hum. Genet. *94*,.
17. de Jong, S., Chepelev, I., Janson, E., Strengman, E., van den Berg, L.H., Veldink, J.H., and Ophoff, R.A. (2012). Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. BMC Genomics *13*, 458.
18. Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*, 1784.
19. Giner-Delgado, C., Villatoro, S., Lerga-Jaso, J., Gayà-Vidal, M., Oliva, M., Castellano, D., Pantano, L., Bitarello, B.D., Izquierdo, D., Noguera, I., et al. (2019). Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat. Commun. *10*, 4222.
20. Ruiz-Arenas, C., Cáceres, A., López-Sánchez, M., Tolosana, I., Pérez-Jurado, L., and González, J.R. (2019). scoreInvHap: Inversion genotyping for genome-wide association studies. PLOS Genet. *15*, e1008203.
21. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. *48*, 709–717.
22. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. *12*, e1001779.
23. Bonàs-Guarch, S., Guindo-Martínez, M., Miguel-Escalada, I., Grarup, N., Sebastian, D.,

Rodriguez-Fos, E., Sánchez, F., Planas-Fèlix, M., Cortes-Sánchez, P., González, S., et al. (2018). Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat. Commun. *9*, 321.
24. Pedersen, B.S., and Quinlan, A.R. (2017). Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am. J. Hum. Genet. *100*, 406–413.
25. Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E., Langmead, B., and Leek, J.T. (2017). Reproducible RNA-seq analysis using recount2. Nat. Biotechnol. *35*, 319–321.
26. Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.
27. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.
28. van de Bunt, M., Manning Fox, J.E., Dai, X., Barrett, A., Grey, C., Li, L., Bennett, A.J., Johnson, P.R., Rajotte, R. V, Gaulton, K.J., et al. (2015). Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. PLoS Genet. *11*, e1005694.
29. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
30. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., Bunt, M. van de, et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk–associated variants. Nat. Genet. *46*, 136.
31. Miguel-Escalada, I., Bonàs-Guarch, S., Cebola, I., Ponsa-Cobas, J., Mendieta-Esteban, J., Atla, G., Javierre, B.M., Rolando, D.M.Y., Farabella, I., Morgan, C.C., et al. (2019). Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat. Genet. *51*, 1137–1148.
32. Fadista, J., Vikman, P., Laakso, E.O., Mollet, I.G., Esguerra, J. Lou, Taneera, J., Storm, P., Osmark, P., Ladenvall, C., Prasad, R.B., et al. (2014). Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc. Natl. Acad. Sci. U. S. A. *111*, 13924–13929.
33. Tolosana, I., Ruiz-Arenas, C., and González, J.R. (2018). imputeInversion.
34. Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
35. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
36. González, J.R., Armengol, L., Solé, X., Guinó, E., Mercader, J.M., Estivill, X., and Moreno, V. (2007). SNPassoc: an R package to perform whole genome association studies. Bioinformatics *23*, 644–645.
37. Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb). *95*, 221–227.
38. Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). **mediation** : *R* Package for Causal Mediation Analysis. J. Stat. Softw. *59*, 1–38.
39. Lewis, F.I., and Ward, M.P. (2013). Improving epidemiologic data analyses through multivariate regression modelling. Emerg. Themes Epidemiol. *10*, 4.
40. Banks, J., Marmot, M., Oldfield, Z., and Smith, J.P. (2006). Disease and Disadvantage in the United States and in England. JAMA *295*, 2037.
41. Stunkard, A.J., Faith, M.S., and Allison, K.C. (2003). Depression and obesity. Biol. Psychiatry *54*, 330–337.
42. Martins-Silva, T., Vaz, J. dos S., Hutz, M.H., Salatino-Oliveira, A., Genro, J.P., Hartwig, F.P., Moreira-Maia, C.R., Rohde, L.A., Borges, M.C., and Tovo-Rodrigues, L. (2019). Assessing causality in the association between attention-deficit/hyperactivity disorder and obesity: a Mendelian

randomization study. Int. J. Obes. 1.
43. Xu, S., Gilliland, F.D., and Conti, D. V (2019). Elucidation of causal direction between asthma and obesity: a bi-directional Mendelian randomization study. Int. J. Epidemiol.
44. Millard, L.A.C., Davies, N.M., Tilling, K., Gaunt, T.R., and Davey Smith, G. (2019). Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization. PLOS Genet. *15*, e1007951.
45. Puig, M., Casillas, S., Villatoro, S., and Cáceres, M. (2015). Human inversions and their functional consequences. Brief. Funct. Genomics *14*, 369–379.
46. Berghöfer, A., Pischon, T., Reinhold, T., Apovian, C.M., Sharma, A.M., and Willich, S.N. (2008). Obesity prevalence from a European perspective: a systematic review. BMC Public Health *8*, 200.
47. Wolf-Maier, K., Cooper, R.S., Banegas, J.R., Giampaoli, S., Hense, H.-W., Joffres, M., Kastarinen, M., Poulter, N., Primatesta, P., Rodríguez-Artalejo, F., et al. (2003). Hypertension Prevalence and Blood Pressure Levels in 6 European Countries, Canada, and the United States. JAMA *289*, 2363.
48. Povel, C.M., Boer, J.M.A., Reiling, E., and Feskens, E.J.M. (2011). Genetic variants and the metabolic syndrome: a systematic review. Obes. Rev. *12*, 952–967.
49. Borowiec, M., Liew, C.W., Thompson, R., Boonyasrisawat, W., Hu, J., Mlynarski, W.M., El Khattabi, I., Kim, S.-H., Marselli, L., Rich, S.S., et al. (2009). Mutations at the BLK locus linked to maturity onset diabetes of the young and -cell dysfunction. Proc. Natl. Acad. Sci. *106*, 14460–14465.
50. Ligthart, S., Vaez, A., Hsu, Y.-H., Stolk, R., Uitterlinden, A.G., Hofman, A., Alizadeh, B.Z., Franco, O.H., Dehghan, A., Alizadeh, B.Z., et al. (2016). Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC Genomics *17*, 443.
51. Logsdon, B.A., Hoffman, G.E., and Mezey, J.G. (2012). Mouse obesity network reconstruction with a variational Bayes algorithm to employ aggressive false positive control. BMC Bioinformatics *13*, 53.
52. Bailey, J.N.C., Lu, L., Chou, J.W., Xu, J., McWilliams, D.R., Howard, T.D., Freedman, B.I., Bowden, D.W., Langefeld, C.D., and Palmer, N.D. (2013). The Role of Copy Number Variation in African Americans with Type 2 Diabetes-Associated End Stage Renal Disease. J. Mol. Genet. Med. *7*, 61.
53. Moyerbrailean, G.A., Kalita, C.A., Harvey, C.T., Wen, X., Luca, F., and Pique-Regi, R. (2016). Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding? PLOS Genet. *12*, e1005875.

FIGURE TITLE AND LEGENDS

Figure 1. Discovery and validation datasets. The flow chart shows the discovery sample and the validation datasets as well as the datasets used for post-genomic data analyses. Sample size (n) used from each dataset after performing quality control are also shown.

Figure 2. **Association analyses between 21 inversions and 8 diseases (in bold) and 17 traits and the co-occurrence of obesity with 6 other complex diseases.** Circles represent the direction (color) and the strength (size) of the association for different groups of traits (morphometric, metabolic, lipidic, respiratory and behavioral) and the epidemiological well-established co-occurrence of obesity-related diseases. Inversions are grouped by size and features: 1) submicroscopic are large (0.4-4Mb) encompassing multiple genes and flanked by segmental duplications; 2) intragenic are located within a gene, either intronic or containing one exon; and 3) intergenic are enriched in pleitropic regions

Figure 3. **Validation of positive associations between the inversion 8p23.1 with diabetes, obesity and their co-occurrence in the 70KT2D dataset and transcriptional allelic effects in samples from EGCUT Biobank and GTEx tissues. A**: Meta-analysis of datasets belonging to 70KT2D for the association of inversion 8p23.1 with diabetes. **B**: Meta-analysis of datasets belonging to 70KT2D for the association of inversion 8p23.1 with obesity. **C**: Meta-analysis of datasets belonging to 70KT2D for the association of inversion 8p23.1 with obese and diabetic individuals. **D:** Differential expressed genes at inversion genotypes (at 5% FDR) in different tissues from GTEx. **E:** Differentially expressed genes at inversion genotypes (at 5% FDR) in blood samples from EGCUT Biobank. **F:** *FAM66A* gene expression interaction between diabetic status and inversion 8p23.1 in pancreatic islets samples (p=0.0254).

Figure 4. **Mechanisms underlying the inversion association with diabetes. Panel A** shows the islet specific expression of inversion 8p23.1 genes. We observed a cluster of islet-specific genes, mainly lncRNAs, next to the distal inversion breakpoint that could be separated from regulatory elements located inside the inverted region**.** The bottom panel depicts an eQTLs (rs1478898) of *FAM66A* gene disrupted by the inversion distal breakpoint (van de Bunt et al, 2015)**.** *FAM66D* has its gene body split in two by the inversion, and would also have its promoter separated from its eQLT SNP (rs140730217) by the inversion. This could be the most likely causal candidate. **Panel B** shows the same information for the inversion 16p11.2**.** *TUFM* and *EIF3C* have their lead eQTL SNP separated by the inversion breakpoint. There is no evidence in the centiSNP database[53] for SNP rs42861 to be causal, suggesting that it should be in LD with the causal variant. This promoter region SNP is located in a segmental duplication block that is closer to *TUFM* in the inverted haplotypes. Therefore, positional changes made by the inversion can affect *TUFM* gene expression by separating the gene from regulatory sequences and subsequently increasing obesity risk.

Figure 5. **Mediation effect of obesity in the causal link between inversions and diabetes and hypertension. Figure A** shows mediation analysis of obesity in the association between inversion 8p23.1 which is the Best Bayesian Network when analyzing these three variables. **Figure B** shows the Best Bayesian Network based on AIC obtained after including obesity, hypertension, diabetes and inversions 8p23.1, 16p11.2 and 11q13.2. Results are obtained from UKB data.

TABLE TITLES AND LEGENDS

Table 1. **Characteristics of the 21 genomic inversions.** The table shows the coordinates, SNP content, size, and inversion frequency obtained from 1000 Genomes as described in Ruiz-Arenas et al[20], the UKB and European regions (north, center and south) using the regions described in the POPRES dataset (see **Methods**). The p-value correspond to a trend test to assess north-south linear association (in bold those significant at 5% level).

						European Populations (POPRES)			
Chr. band	**Coordinates**	**Num. SNPs**	**Length (Kb)**	**Inv. Freq.[20]**	**UKB**	**North**	**Center**	**South**	**trend p-value**
1p22.1	chr1:92,131,841-92,132,615	6	0.77	11.23	10.1	**8.9**	**9.1**	**14.4**	**0.0057**
1q31.3	chr1:197,756,784-197,757,982	5	1.2	19.68	20.2	19.4	21.7	19.1	0.8781
2p22.3	chr2:33,764,554-33,765,272	6	0.72	15.11	15.5	13.8	13.5	11.7	0.3199
2q22.1	chr2:139,004,949-139,009,203	13	4.25	71.47	75.3	**76.6**	**71.9**	**66.4**	**0.0003**
3q26.1	chr3:162,545,362-162,547,641	6	2.28	56.16	51.1	**53.4**	**55.2**	**61.1**	**0.0140**
6p21.33	chr6:31,009,222-31,010,095	5	0.87	63.12	62	**61.3**	**65.0**	**72.8**	**0.0001**
6q23.1	chr6:130,848,198-130,852,318	12	4.12	6.56	7.6	7.3	8.7	8.1	0.6070
7p14.3	chr7:31,586,765-31,592,019	11	5.25	23.56	23.5	22.6	23.3	26.5	0.1605
7p11.2	chr7:54,302,450-54,376,389	180	73.9	50.39	51	52.1	51.2	54.4	0.4715
7q11.22	chr7:70,426,185-70,438,879	10	12.7	63.52	61.8	61.0	61.8	62.4	0.6196
7q36.1	chr7:151,010,030-151,012,107	5	2.08	19.88	20.7	20.1	24.0	24.7	0.0775
8p23.1	chr8:8,055,789-11,980,649	13,411	3,925	57.95	55.6	56.5	54.9	53.6	0.3424
11p12	chr11:41,162,296-41,167,044	7	4.75	15.81	15.4	14.3	13.9	14.6	0.8479
11q13.2	chr11:66,018,563-66,019,946	5	1.38	34.39	28.5	32.4	31.3	30.5	0.5287
12q13.11	chr12:47,290,470-47,309,756	43	19.3	7.46	6.6	**6.2**	**7.9**	**10.9**	**0.0085**
12q21.2	chr12:71,532,784-71,533,816	4	1.03	36.98	38.8	37.4	36.5	33.3	0.1647
14q23.3	chr14:65,842,304-65,843,165	4	0.86	29.42	25.5	26.5	26.9	26.4	0.9823
16p11.2	chr16:28,424,774-28,788,943	361	364.17	ND	40.5	**39.3**	**32.0**	**29.1**	**0.0007**
17q21.31	chr17:43,661,775-44,372,665	3637	711	23.96	22.6	**15.1**	**19.4**	**22.1**	**0.0035**
21q21.3	chr21:28,020,653-28,021,711	11	1.06	51.29	49.2	51.6	52.1	57.4	0.0651
Xq13.2	chrX:72,215,927-72,306,774	135	90.8	13.3	13.9	**12.4**	**12.1**	**8.5**	**0.0400**

Table 2: Association between inversions 8p23.1 and 16p1.2 and different obesity-related traits in UKB and replication data sets. The table shows the odds ratios (OR) and their confidence intervals at 95% (CI95%) for the inverted allele and different diseases and the joint co-occurrence with obesity at UKB and replication datasets. The p corresponds to the best genetic model depict in the first column of each inversion.

	Inversion 8p23.1 (effect of risk-Haplotype: N-allele)					**Inversion 16p11.2 (effect of risk-Haplotype: N-allele)**			
	UKB		**Replication**			**UKB**		**Replication**	
Disease	**OR CI95%**	**p-value**	**OR CI95%**	**p-value**		**OR CI95%**	**p-value**	**OR CI95%**	**p-value**
Obesity	1.04 (1.03-1.05)	2.4x10^{-13}	1.08 (1.04-1.11)	5.6x10^{-6}		1.05 (1.04-1.06)	3.9x10^{-24}	1.07 (1.03-1.10)	1.4x10^{-4}
Diabetes	1.04 (1.01-1.06)	1.1x10^{-3}	1.08 (1.05-1.11)	1.1x10^{-8}		1.02 (0.99-1.04)	0.1450	1.07 (1.04- 1.11)	1.2x10^{-6}
Hypertension	1.04 (1.03-1.05)	7.0x10^{-16}	1.03 (1.00-1.05)	0.0183		1.01 (1.00-1.02)	0.0184	1.02 (0.99-1.05)	0.2127
Asthma	1.03 (1.01-1.04)	7.0x10^{-5}	1.02 (0.90-1.05)	0.2225		1.00 (0.99-1.01)	0.9529	1.00 (0.97-1.04)	0.8074
Depression	0.98 (0.97-0.99)	0.0119	1.01 (0.97-1.05)	0.6630		0.98 (0.96-1.00)	0.0184	1.01 (0.98-1.05)	0.5384
Joint occurrence of Obesity with:									
Diabetes	1.08 (1.05-1.11)	3.1x10^{-7}	1.17 (1.12-1.22)	1.4 x10^{-13}		1.06 (1.03-1.08)	7.5x10^{-5}	1.13 (1.08- 1.17)	1.2x10^{-8}
Hypertension	1.07 (1.05-1.08)	1.7x10^{-16}	1.06 (1.02-1.11)	6.9x10^{-3}		1.06 (1.05-1.07)	2.7x10^{-14}	1.05 (1.00-1.10)	0.0357
Asthma	1.08 (1.06-1.10)	3.0x10^{-11}	1.09 (1.02-1.16)	9.7x10^{-3}		1.05 (1.03-1.07)	7.4x10^{-6}	1.08 (1.01-1.15)	0.0287
Depression	1.04 (1-02-1.07)	1.4x10^{-3}	1.12 (1.04-1.20)	3.8x10^{-3}		1.06 (1.03-1.08)	1.4x10^{-6}	1.03 (0.95-1.11)	0.5241