SUPPLEMENTARY INFORMATION

A genome-wide association meta-analysis identifies new childhood obesity loci

Jonathan P. Bradfield[1,69], H. Rob Taal[2,3,4,69], Nicholas J. Timpson[5], André Scherag[6], Cecile Lecoeur[7], Nicole M. Warrington[8], Elina Hypponen[9], Claus Holst[10], Beatriz Valcarcel[11], Elisabeth Thiering[12], Rany M. Salem[13,14,15,16,17,18], Fredrick R. Schumacher[19], Diana L. Cousminer[20], Patrick M.A. Sleiman[1,21,22], Jianhua Zhao[21], Robert I. Berkowitz[23,24,25], Karani S. Vimaleswaran[9], Ivonne Jarick[26], Craig E. Pennell[8], David M. Evans[5], Beate St. Pourcain[5], Diane J. Berry[9], Dennis O Mook-Kanamori[2,3,4,27], Albert Hofman[3], Fernando Rivadeinera[3,28], André G. Uitterlinden[3,28], Cornelia M. van Duijn[3], Ralf J.P. van der Valk[2,29,30], Johan C. de Jongste[2, 30], Dirkje S. Postma[29], Dorret I. Boomsma[31], William J. Gauderman[19], Mohamed T. Hassanein[19], Cecilia M. Lindgren[32], Reedik Mägi[32,33], Colin A.G. Boreham[34], Charlotte E. Neville[35], Luis A. Moreno[36,37], Paul Elliott[11], Anneli Pouta[38], Anna-Liisa Hartikainen[39], Mingyao Li[40], Olli Raitakari[41], Terho Lehtimäki[42], Johan G. Eriksson[43,44,45,46], Aarno Palotie[18,20,47,48], Jean Dallongeville[49], Shikta Das[11], Panos Deloukas[47], George McMahon[50], Susan M. Ring[50], John P. Kemp[50], Jessica L. Buxton[51], Alexandra I.F. Blakemore[51], Mariona Bustamante[52,53,54], Mònica Guxens[52,53], Joel N. Hirschhorn[13,14,15,16,17,18], Matthew W. Gillman[55], Eskil Kreiner-Møller[56], Hans Bisgaard[56], Frank D. Gilliland[19], Joachim Heinrich[12], Eleanor Wheeler[47], Inês Barroso[47,57,58], Stephen O'Rahilly[57,58], Aline Meirhaeghe[49], Thorkild I.A. Sørensen[10,59], Chris Power[9], Lyle J. Palmer[60], Anke Hinney[61], Elisabeth Widen[20], I. Sadaf Farooqi[57,58], Mark I. McCarthy[62,63,64], Philippe Froguel[7,51], David Meyre[7,65], Johannes Hebebrand[61], Marjo-Riitta Jarvelin[11,38,66,67], Vincent W.V. Jaddoe[2,3,4], George Davey Smith[5], Hakon Hakonarson[1,21,22,68] and Struan F.A. Grant[1,21,22,68**] for the Early Growth Genetics (EGG) Consortium

Affiliations

[1]Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; [2]The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands; [3]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; [4]Department of Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands; [5]The Medical Research Council Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Oakfield House, Bristol, UK; [6]Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany; [7]Centre National de la Recherche Scientifique (CNRS) 8199, Lille North of France University, Pasteur Institute, Lille, France; [8]School of Women's and Infants' Health, University of Western Australia, Perth, Western Australia, Australia; [9]Centre for Paediatric Epidemiology and Biostatistics, MRC Centre of Epidemiology for Child Health, University College of London Institute of Child Health, London, UK; [10]Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark; [11]Department of Epidemiology and Biostatistics, MRC-HPA Centre for Environment and Health, School of Public Health, Faculty of Medicine, Imperial College London, London UK; [12]Helmholtz Zentrum Muenchen, German Research Centre for Environmental Health, Institute of Epidemiology I, Neuherberg, Germany; [13]Divisions of Genetics and [14]Endocrinology and [15]Program in Genomics, Children's Hospital, Boston, Massachusetts; USA; [16]Metabolism Program and [17]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; [18]Harvard Medical School, Boston, Massachusetts, USA; [19]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; [20]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [21]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; [22]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [23]Behavioral Health Center and [24]Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia PA, USA; [25]Center for Weight and Eating Disorders, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA; [26]Institute of Medical Biometry and Epidemiology, Philipps-University of Marburg, Marburg, Germany; [27]Department of Physiology and Biophysics, Weill Cornell Medical College - Qatar, Doha, Qatar; [28]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; [29]Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands; [30]Division of Respiratory Medicine, Department of Pediatrics, Erasmus University Medical Center - Sophia Children's Hospital, Rotterdam, The Netherlands; [31]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [32]Genetic and Genomic Epidemiology Unit, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [33]Estonian Genome Center, University of Tartu, Tartu, Estonia; [34]Institute for Sport and Health, University College Dublin, Ireland; [35]Centre for Public Health, Queen's University Belfast, Institute of Clinical Science B, Belfast, UK; [36]GENUD (Growth, Exercise, NUtrition and Development) Research Group and [37]School of Health Sciences (EUCS), Universidad de Zaragoza, Zaragoza, Spain; [38]National Institute for Health and Welfare, Oulu, Finland; [39]Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, Finland; [40]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; [41]Department of Clinical Physiology, Turku University Hospital and University of Turku, Finland; [42]Department of Clinical Chemistry, Tampere University Hospital and University of Tampere, Finland; [43]National Institute for Health and Welfare, Helsinki, Finland; [44]Department of General Practice and Primary Health Care, University of Helsinki, Finland; [45]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland; [46]Folkhälsan Research Center, Helsinki, Finland; [47]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK; [48]Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland; [49]INSERM, U744; Institut Pasteur de Lille, Univ Lille Nord de France, Lille, France; [50]School of Social and Community Medicine, University of Bristol, UK; [51]Department of Genomics of Common Disease, Imperial College London, London, UK; [52]Center for Research in Environmental Epidemiology (CREAL) - Hospital del Mar Research Institute (IMIM), Barcelona, Catalonia, Spain; [53]CIBER Epidemiologia y Salud Pública (CIBERESP), Spain; [54]Genes and Disease Program, Center for Genomic Regulation (CRG-UPF), Barcelona, Catalonia, Spain; [55]Obesity Prevention Program, Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Boston, MA; [56]Copenhagen Prospective Studies on Asthma in Childhood, Health Sciences, University of Copenhagen, Copenhagen University Hospital, Gentofte, Denmark; [57]University of Cambridge Metabolic Research Laboratories and [58]NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom; [59]Novo Nordisk Foundation Center for Basic Metabolic Research, Section on Metabolic Genetics, University of Copenhagen, Copenhagen, Denmark; [60]Ontario Institute for Cancer Research and Samuel Lunenfeld Research Institute, Toronto, Canada; [61]Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen, Germany; [62]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK; [63]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; [64]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford, UK; [65]Department of Clinical Epidemiology & Biostatistics, McMaster University, Hamilton, Canada; [66]Institute of Health Sciences, University of Oulu. Oulu, Finland; [67]Biocenter Oulu, University of Oulu; Oulu, Finland; [68]Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [69]These authors contributed equally to this work.

Supplementary Note

Individual Study Cohort Descriptions

Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC provided data at both the discovery and replication stages of this study. ALSPAC is a prospective study, which recruited pregnant women with expected delivery dates between April 1991 and December 1992 from Bristol UK. We used the first four clinical exams - 6wks, 9mnths, 18mnths and pre - school (these are reported to health visitors - heights and weights) then we have clinic visits for the measurement of height and weight that follow at 7,8,9,10,11,~12 and ~14 yrs. Height was measured to the last complete mm using the Harpenden Stadiometer. Children were positioned with their feet flat and heels together, standing straight so that their heels, calves, buttocks and shoulders came into contact with the vertical backboard of the stadiometer. The headboard was lowered down the backboard until it touched the child's head and a 1kg weight was placed on the headboard to ensure head contact and to minimise the effect of hair thickness. Any problems with measuring were noted.

Weight and bioelectrical impedance were measured using the Tanita Body Fat Analyser (Model TBF 305). The child was encouraged to pass urine (see F7MS029) and undress to their underclothes. 'Female Standard' was entered into the machine for all children and their height was entered to the nearest cm. The child stepped onto the measuring platform which had been wiped with disinfecting alcohol and positioned so that both feet were located in parallel with the toe and heel in contact with their respective electrodes. Measurement was completed when the weight and fat ratio readings were fixed and the buzzer beeped. Weight was measured to the nearest 50g. Any problems were noted.

ALSPAC has its own Law and ethics Committee that reviews all proposals for new data collection and approves policies for data handling and analysis. Proposals for new data collection are also approved by the Local Research Ethics Committees (LRECs). Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the Local Research Ethics Committees.

3,714 participants (discovery sample) were genotyped using either the Illumina 317K or 610K genome-wide SNP genotyping platforms by the Wellcome Trust Sanger Institute, Cambridge, UK and the Centre National de Génotypage, Evry, France. A common set of SNPs (present in both genotyping platforms) were extracted and the resulting raw genome-wide data was subjected to standard quality control methods. Individuals were excluded on the basis of having incorrect sex assignments; minimal (0.34) or excessive (0.36) heterozygosity; disproportionate levels of individual missingness (>3%) and evidence of cryptic relatedness (PI HAT >0.11). The remaining individuals were assessed for evidence of population stratification by multidimensional scaling (MDS) analysis, using CEU, YRI, JPT and CHB individuals from the HapMap as reference ethnic groups. Only those clustering with the CEU individuals were included. The underlying population stratification was thereafter controlled for by using the first two EIGENSTRAT derived ancestry informative covariates (EIG-PC1 and EIG-PC2) as well as any additional EIGENSTRAT derived ancestry informative covariates that were correlated with the phenotype (EIG-PC3 for g-men; EIG PC9 for exr.XY; EIG-PC4 for prn-alL; EIG-PC10 for ls-pg; EIG-PC5 for PC3; EIG-PC5 for PC11; EIG-PC9 for PC9). SNPs with a minor allele frequency of <0.5% and call rate of <97% were removed. Furthermore, only SNPs which passed an exact test of Hardy-Weinberg equilibrium (P>5x10-7) were considered for analysis. The resulting dataset consisted of 3,233 individuals and 285,531 SNPs. We then conducted imputation with MACH 1.0 Markov Chain Haplotyping software17, using CEU individuals from phase 2 of the HapMap project as a reference set (release 22). The final imputed dataset consisted of 3,233 subjects, each with 2,543,887 imputed autosomal markers.

Additional genotyping (replication sample) was carried out at two different centres (The Wellcome Trust Sanger Centre, Cambridge, UK and Laboratory Corporation of America, Burlington, NC, US) using the Illumina Human550 quad array (Illumina, Inc., San Diego, CA). Individuals were excluded on the basis of the following: sex mismatches, minimal or excessive heterozygosity, disproportionate levels of individual missingness (>3%), cryptic relatedness measured as proportion of identity by descent (IBD > 0.1), and insufficient sample replication (IBD < 0.8). The remaining individuals were assessed for evidence of population stratification by multidimensional scaling analysis and compared with Hapmap II (release 22)

European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed. Hidden population-stratification was thereafter controlled for by using EIGENSTRAT (Alkes et al. 2006). SNPs with a minor allele frequency of < 1%, a call rate of < 95% or evidence for violations of Hardy-Weinberg equilibrium (P < 5E-7) were removed. Genotypic data were subsequently imputed using Markov Chain Haplotyping software (MACH v.1.0.) and phased haplotype data from CEU individuals (Hapmap release 22, Phase II NCBI B36, dbSNP 126) based on a cleaned dataset of 8,365 individuals and 464,311 autosomal SNPs..

Genotypic dosages used for main modeling and analyzed in MACH2DAT according to the models specified and where samples had not been previously run in the discovery collection.

Northern Finnish Birth Cohort Study 1966

The Northern Finland Birth Cohort study 1966 (NFBC1966) (http://kelo.oulu.fi/NFBC/) includes 12,058 live born individuals, of European descent, with expected dates of birth during 1966 in the two northernmost provinces of Finland,Oulu and Lapland[1]. The University of Oulu Ethics Committee and the Ethical Committee of Northern Ostrobothnia Hospital District have approved the study. The data on all cohort members were prospectively collected since pregnancy and supplemented at the ages of 1, 14 and 31 years. Growth measurements were obtained from communal child health clinics[2]. All those living in northern Finland or in the capital area were invited to a clinical examination and blood sampling at age 31 years[3]. BMI used in this analysis was calculated from the measurements of height and weight obtained throughout the childhood. For this analysis a total of 52,486 BMI measures obtained at ages 2-18 years were used. Blood samples were drawn and DNA was extracted when individuals were 31 years old. Participants provided written informed consent.

DNA was extracted from 5753 individuals and GWAS data for current analyses was available for 1221 individuals. Illumina's HumanCNV370-Duo DNA Analysis BeadChip was used to obtain genome-wide data. It contains an informative set of tag SNPs derived from the HapMap European-derived (CEU)

sample. Imputation was performed on 328,007 SNPs using IMPUTE software version 0.3.1, applying information threshold of >0.4 and MAF threshold of >1%.

Northern Finnish Birth Cohort Study 1986

The NFBC1986 includes 9,432 live born children with expected dates of birth between 1st July 1985 and 30th June 1986 in the two northernmost provinces of Finland, Oulu and Lapland. The University of Oulu Ethics Committee and the Ethical Committee of Northern Ostrobothnia Hospital District have approved the study. Cohort has been followed up since early pregnancy until adolescence. Growth measurements were obtained from communal child health clinics. All those alive with known address were invited to a clinical examination at the age of 15 to 16 years. BMI was calculated from the measurements height and weight obtained at ages 2-18 years. A total of 65,034 BMI measures were used in this analysis. All those alive with known address were invited to a clinical examination at the age of 15 to 16 years. At this point, blood samples were drawn and DNA was extracted for 6,266 subjects using standard methods Participants provided written informed consent. DNA was extracted for 6,266 participants.

For the current analyses, data was available for 1503 individuals. Genotyping data for three of the eight SNPs selected for replication was obtained following genotyping of 6208 individuals at the Broad Institute, using a custom Illumina chip (Cardio-MetaboChip). The other five SNPs were genotyped by Kbiosciences (http://www.kbioscience.co.uk), and genotypes called using their Klustercaller software. Genotyping success rate for all SNPs was >95.0% and none of the SNPs deviated from Hardy-Weinberg equilibrium (all *P*-values>0.0001).

British 1958 Birth Cohorts (BC58)

The 1958 British Birth Cohort (1958BC) is a population based prospective study of all born during one week in March 1958 in England, Scotland, and Wales (n=17,638)[4]. The cohort has been followed-up during childhood, at ages 7, 11, and 16 years and into adulthood, with measurements of height and weight obtained at all childhood contacts. BMI used in this analysis was calculated from the measurements of

height and weight obtained at ages 7, 11 and 16 years and centiles were defined from within the study. At age 45 years, 11,971 cohort members who had not died or emigrated, were invited to a biomedical assessment. In total 9,377 participants provided data, including several clinical assessments and DNA collection. Ethical approval for the 45y survey was obtained from South East Multi-centre Research Ethics Committee (ref. 01/1/44). Genome-wide data for the 1958BC has been obtained through two sub-studies, both using the 1958BC members as a control population. First, 3000 DNA samples were randomly selected as part of the Welcome Trust Case Control Consortium (WTCCCII) and genotyped on the Affymetrix SNP 6.0 platform[5]. Secondly, 2,592 DNA samples from the 1958BC were used as controls for a Type 1 diabetes case-control study. Samples were genotyped through the JDRF/WT Diabetes and Inflammation Laboratory (DIL) using the Illumina Infinium 550K chip[6]. Imputation was done in IMPUTE after quality control. For B58C-WTCCC quality control included SNP exclusions (MAF < 0.01, HWE<$1x10^{-20}$, call rate < 0.98, genotype plate association <$1 x10^{-5}$), and sample exclusions (heterozgygocity, call rate, relatedness, non-European ancestry and sex discrepancy, with outliers determined by Bayesian clustering approaches)[7,8]. For B58C-T1DGC, criteria for SNP exclusions were MAF < 0.01, HWE<$1x10^{-7}$, or SNP call rate < 0.95, and for sample exclusions call rate < 0.97, heterozygosity <0.29 or >0.34, non-European ancestry or sex discrepancy.

FRENCH YOUNG

We studied 670 French obese children, recruited through a multi-media campaign by the CNRS UMR8199 unit in Lille, in the Department of Pediatric Endocrinology of Jeanne de Flandres Hospital or in the Toulouse Children's Hospital[9]. Obese children, recruited by the CNRS UMR8199 unit and by the Jeanne de Flandres hospital between 1997 and 2007, are issued from pedigrees having at least one obese child with a BMI ≥ 97^{th} percentile for sex and age and both parents. Additional obese children were recruited in Toulouse Children's Hospital between 1997 and 2001 during pediatric obesity consultation and all harbor a BMI ≥ 97^{th} percentile for sex and age. The 349 lean French children were selected from the STANISLAS family study[10] or from the Fleurbaix Laventie Ville Santé II study, both being family-based

recruitments [11]. The Fleurbaix Laventie Ville Santé II Study includes 224 nuclear families recruited in 1999 and representative of the Northern France general population. The STANISLAS cohort includes 1006 families consisting of the two biological parents with at least two children, recruited in the Nancy area between September 1993 and August 1995 and representative of the Eastern France general population. Cases and controls were genotyped using the Illumina Human CNV370-Duo array. IMPUTE v0.5.0 based on the CEU HapMap reference panel (NCBI build 36). Prior to imputation, a quality control filter on genotyped SNPs was applied: i) call rate >= 95%, ii) *P*-value of the Hardy-Weinberg equilibrium test > 0.0001.

Lifestyle – Immune System – Allergy Study and German Infant Study on the influence of Nutrition Intervention (LISA+GINI)

The influence of Life-style factors on the development of the Immune System and Allergies in East and West Germany PLUS the influence of traffic emissions and genetics (LISAplus) Study is a population based birth cohort study. A total of 3097 healthy, mature (gestational age over 37 weeks) neonates with a birth weight over 2500g were recruited between 1997 and 1999 in Munich, Leipzig, Wesel and Bad Honnef.

A total of 5991 mothers and their newborns were recruited into the German Infant study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development (GINIplus) between September 1995 and June 1998 in Munich and Wesel. Detailed descriptions of the LISAplus and GINIplus studies have been published elsewhere [12,13].

Weight and height were measured in both studies at age 2, 4, and 5 years by the family physician and at age 10 years by the physician of the study or by the parents. BMI >95th percentile was defined based on sex and age standardized German reference values (K. Kromeyer-Hauschild, et al.: Monatsschr. Kinderheilk. 149(2001)). For both studies, approval by the local Ethics Committees and written consent from participant's families were obtained.

In the discovery analysis, 277 children from the LISAplus study from Munich with genome-wide data at least two BMI measurements were included. DNA was analyzed using the Affymetrix Human SNP Array 5.0 for each individual. Genome-wide data was called using BRLMM-P algorithm and imputed after quality control (MAF>1%, HWE>0.01, call rate per SNP and person >95%) in IMPUTE. Genome-wide association analysis of BMI was carried out in SNPTEST V2.

For replication, a independent subset of 497 children from Munich from both studies with genome-wide data and at least two BMI measurements was included (371 (74.6%) children from the GINIplus study and 126 (25.4%) children from the LISAplus study)). 435 individuals were analyzed using the Affymetrix Human SNP Array 5.0 and 62 individuals from the LISAplus study were analyzed using Affymetrix Human SNP Array 6.0. Genotypes were called using BRLMM-P algorithm (5.0), respectively BIRDSEED V2 algorithm (6.0), imputed after quality control (MAF>1%, HWE>1e-05, heterocygocity 4SD, call rate per SNP and person >95%) in IMPUTE2 and genome-wide association analysis of BMI was carried out in SNPTEST V2.

Western Australian Pregnancy Cohort study (RAINE)

The Western Australian Pregnancy Cohort study (Raine) was started as a randomized controlled trial to evaluate the effects of repeated ultrasound in pregnant women in Perth, Western Australia. In total, 2,900 pregnant women were recruited between 1989 and 1991 prior to 18‐weeks gestation at the King Edward Memorial Hospital (Perth, Western Australia). Women were randomized to repeated ultrasound measurements at 18, 24, 28, 34 and 38 weeks gestation or to a single ultrasound assessment at 18‐weeks. Children have been assessed at average ages of 1, 2, 3, 5, 8, 10, 14 and 17 and both height and weight were collected at each assessment. The obesity phenotype for this study was defined as greater than the 95th percentile on the Centre for Disease Control (CDC) growth charts at any measured time throughout childhood (n=232). Controls were defined as consistently below or equal to the 50th percentile on the CDC growth charts at all measurement times (n=125).

DNA was collected at the year 14 and 17 follow-ups. The study was conducted with appropriate institutional ethics approval, and written informed consent was obtained from mothers at all follow‐ups and participants at the year 17 follow-up. DNA was collected using standardized procedures from 74% of all adolescents who attended the 14 year follow-up on and a further 5% at the 17 year follow-up measurements. We performed high throughput genome-wide SNP genotyping using the genome‐wide Illumina 660 Quad Array for each individual. Genotype data were imputed against Hapmap phase2 build 36 release 22 using MACH v1.0.16 after quality control (MAF>1%, HWE>$5x10^{-7}$, call rate per SNP and person >95%). Genome-wide association analysis of the obesity phenotype was carried out in mach2dat.

Children's Hospital of Philadelphia (CHOP)

All subjects were consecutively recruited from the Greater Philadelphia area from 2006 to 2010 at the Children's Hospital of Philadelphia. Our study cohort consisted of 1,445 obese children and 2,802 lean children of European ancestry. All of these participants had their blood drawn in to an 8ml EDTA blood collection tube and were subsequently DNA extracted for genotyping. All subjects were biologically unrelated and were aged between 2 and 18 years old. This study was approved by the Institutional Review Board of the Children's Hospital of Philadelphia. Parental informed consent was given for each study participant for both the blood collection and subsequent genotyping. Self-reported ethnicity was confirmed by multidimensional scaling methodologies. BMI ≥ 95th percentile was defined using the Center for Disease control (CDC) z-score=1.645 (http://www.cdc.gov/nchs/about/major/nhanes/growthcharts/datafiles.htm). We performed high throughput genome-wide SNP genotyping, using the Illumina Infinium™ II HumanHap550 BeadChip technology (Illumina, San Diego), at the Center for Applied Genomics at CHOP. We used 750ng of genomic DNA to genotype each sample, according to the manufacturer's guidelines.

Samples were genotyped on a combination of the HumanHap 550 version 1, HumanHap 550 Version 2 and 610 Quad SNP chips. The 535,752 SNPs in common with the 3 different chip versions used

were the basis for all further analyses. 326 Individuals were removed for missing more than 2 percent of their genotypes and 417 individuals were removed for having more than 15 percent of their genome identical by state. 19, 462 SNPs were excluded for having a Hardy-Weinberg equilibrium p-value less than 10-6. 4,172 SNPs were thrown out for missing more than 5 percent of genotypes and 20,136 SNPs were thrown out for having a minor allele frequency less than 1 percent. Mach 1.0 was used to impute ~2.54 million SNPs using the CEU HapMap (release 22, build 36) haplotypes. Post imputation quality control included excluding SNPs with a r2hat less than 0.30 and minor allele frequency less than 1 percent.

Essen Obesity Study (ESSEN) / Essen trios

The Essen Obesity Study (Essen Case – Control) is a case-control study of 397 extremely obese children and adolescents ('cases'; 56.7% female; mean age 13.6 ± 2.5 years; age range 5-18 years; mean BMI 32.7 ± 6.3 kg/m^2; mean BMI-SDS 4.5 ± 2.1) who were recruited in hospitals specialized for the inpatient treatment of extreme obesity while 435 healthy lean individuals ('controls'; 60.7% female; mean age 26.1 ± 5.8 years; age range 17-58 years; mean BMI 18.3 ± 1.1 kg/m^2; mean BMI-SDS -1.4 ± 0.4) were ascertained at the University of Marburg (for details see[14]; note that - due to different in- and exclusion criteria - less cases are reported here). According to self-reports 77% of controls had a body weight equal or below average at their age of 15. Thus, our control group mainly comprises individuals who presumably also had a lower body weight during adolescence. The Essen Obesity Study (Essen Obesity Trios) is a nuclear family study with 705 obese offspring (54.9% female; mean age 13.4 ± 3.0 years; age range 3-24 years; mean BMI 32.0 ± 5.8 kg/m^2; mean BMI-SDS 4.2 ± 2.0) who were ascertained like the cases in the case-control study; in addition the 2x705 biological parents of the offspring were ascertained[15]. To derive BMI percentiles we used the German reference values[16]. Only cases with a measured BMI above the 95th age- and sex-specific percentile were included in this analysis, whereas all controls had a BMI below the 15th percentile. Note that most of the cases (94.5%) and offspring (30.6%) had more extreme BMIs above the 97th percentile. Genotyping for the Essen Obesity

Study (Essen Case – Control and Essen Obesity Trios) was performed on the Genome-Wide Human SNP Array 6.0. Genotypes were called using the 'birdseed V2 calling'-algorithm. We excluded SNPs with a call rate ≤ 95%, those with evidence for departure from Hardy-Weinberg equilibrium (exact two-sided p ≤ .001 in either the control group or the parents) and those with a minor allele frequency < 1 % in each of the total samples. In addition, all families were Mendelian consistent for ≥ 95% of all SNPs. Afterwards, imputation was performed with MACH 1.0 based on the filtered HapMap (phase II, release 23) dataset (60 CEU founder individuals, ~ 2.3 million autosomal SNPs), applying the quality control filters: information threshold of > 0.3 and minor allele frequency threshold of > 1 %. In the Essen Obesity trios sample, Mendelian inconsistent imputation results were additionally excluded from subsequent family-bases association testing. Written informed consent was given by all participants and in case of minors by their parents. The study was approved by the Ethics Committees of the Universities of Marburg and Essen and conducted in accordance with *The Declaration of Helsinki*.

Helsinki Birth Cohort Study (HBCS)

The Helsinki Birth Cohort Study (HBCS) is a population-based prospective cohort of singletons born at the Helsinki University Central Hospital between the years 1934 and 1944. Birth records on 4630 men and 4130 women who lived in Finland in 1971 were taken, including measurements of weight and length. Serial measurements of height and weight were extracted from child welfare clinic and school health clinic records, with an average of 10 measurements between birth and 2 years, and 8 measurements between 2 and 11 years of age. Between 2000 and 2002, a representative subset of 928 males and 1075 females returned for clinical examinations. At this visit, blood was taken for DNA extraction. Genotyping was performed on a custom Illumina 670 Quad platform at the Wellcome Trust Sanger Centre. Quality control was performed before imputation (excluding genotypes with call rate <0.95, MAF <0.01, HWE p <10e-06) with MACH. The current analysis includes 260 cases with a BMI above the 95th percentile at any point in childhood and 405 controls with BMI below or equal to the 50th percentile consistently throughout

childhood. Informed consent was collected from all study participants. The study design was approved by the local ethics committee.

Cardiovascular Risk in Young Finns Study (YF)

Cardiovascular Risk in Young Finns (YF) is an ongoing collaborative study between five Finnish university medical schools (Helsinki, Turku, Oulu, Kuopio, and Tampere) of risk factors for atherosclerosis. The baseline cross-sectional study was carried out in 1980, including 3596 subjects at ages 3, 6, 9, 12, 15, and 18. Between 1980 and 1992, these subjects were followed up at 3-year intervals, and then as adults in 2001 and 2007. Genotyping was performed with the Illumina 670K SNP chip at the Wellcome Trust Sanger Centre. Quality control measures were taken prior to imputation with MACH (excluding genotypes with call rate <0.95, MAF <0.01, HWE p <10e-06). The present study includes 166 cases with BMI above the 95th percentile at any point in childhood and 537 controls with BMI below or equal to the 50th percentile throughout childhood. The classification of cases and controls was typically based on 3 measurements. Informed consent was collected from all study participants. The study design was approved by the local ethics committee.

Copenhagen Study on Asthma in Childhood (COPSAC)

The COPSAC birth cohort study is a prospective clinical study of a birth cohort of 411 infants born to mothers with a history of asthma. The newborns were enrolled at the age of 1 month, the recruitment of which was previously described in detail [17-19]. The study was approved by the Ethics Committee for Copenhagen (KF 01-289/96) and The Danish Data Protection Agency (2008-41-1754) and informed consent was obtained from both parents. The families used doctors employed at the clinical research unit, and not the family practitioner, for diagnosis and treatment of any respiratory or skin-related symptoms. Participants were assessed at the COPSAC clinical research unit at six monthly intervals; additional visits were arranged immediately upon the onset of symptoms. All growth parameters were measured and obtained by the COPSAC physicians at each scheduled six monthly visit until age 7 and history was

obtained using structured questions and closed response categories. At every visit weight was measured using calibrated digital weight scales and length by infantometer, Kiddimetre® (Raven Equipment Limited, Essex England). From 2.5 years, height was measured using a stadiometer (Harpenden; Holtain Ltd, Crymych, Dyfed, Wales)[17].

High throughput genome-wide SNP genotyping were performed using the Illumina Infinium™ II HumanHap550 v1, v3 or quad BeadChip platform (Illumina, San Diego), at the Children's Hospital of Philadelphia's Center for Applied Genomics, as described previously [20] and available in 161 subjects. Imputation was performed using IMPUTE v2 and Hapmap release 22 as reference panel with quality control prior to imputation excluding genotypes with call rate <0.95, MAF <0.01 and HWE P <10^{-4}. Statistical analysis was carried out using SNPTEST, assuming an additive model and taking genotype uncertainty into account.

CM-GOYA study (CM-GOYA)

The cohort was derived from a draft board examination cohort for men, constituted by young Danish adults with negligible admixture of other ethnicities. These men also had measurements of height and weight in school which was used to define cases and controls in this current study. The study was approved by the regional scientific ethics committee and by the Danish Data Protection Board. Genome-wide genotyping on the Illumina610 quad BeadChip was carried out at the Centre National de Génotypage (CNG), Evry, France. We carried out imputation to HapMap release 22 (CEU individuals) using Mach 1.0, Markov Chain Haplotyping[21]. BMI was calculated from yearly measurements of height and weight at age 7 to 13 years. Cases are defined as having at least one BMI above the age specific 95% percentile and controls as having all BMI below the 50th percentile.

Generation R Study (GENERATIONR)

The Generation R Study is a population-based prospective cohort study from fetal life until young adulthood. All children were born between April 2002 and January 2006. This study is designed to identify

early environmental and genetic determinants of growth, development and health from fetal life until young adulthood and has been described previously in detail [22,23]. Detailed measurements were performed using ultrasound and physical examinations, biological samples and advanced imaging techniques. The study has been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam. Written informed consent was obtained from all participants or their parent(s).

Analysis were restricted to Caucasian individuals with genome-wide data and at least one BMI measurement available between 2 and 6 years of age (n=916, number of boys 445 (48.6%)). Length was measured to the nearest millimeter and weight was recorded by well-trained staff in community health centers using standardized procedures[23]. Visits to the community health centers were based on the national routine health care program. Sex- and age-adjusted standard deviation scores (SDS) were constructed using Growth Analyser 3.0 (http://www.growthanalyser.org; Dutch Growth Research Foundation, Rotterdam, the Netherlands). The reference curve for body mass index in the Netherlands, 1997 was used. According to the case definition, 192 cases and 724 controls were included in this analysis[24].

Cord blood for DNA isolation was available in 59% of all live-born participating children. Sex-mismatch rate between genome based sex and midwife-record based sex was low (<0.5%), indicating that possible contamination of maternal DNA was extremely low. Missing cord blood samples were mainly due to logistical constraints at the delivery. Individual genotype data were extracted from the genome-wide Illumina 610 Quad Array[25]. Prior to imputation, SNPs with a Hardy-Weinberg equilibrium *P*-value $<1.0\times10^{-6}$, a call rate <0.95 and a minor allele frequency <0.01 were excluded. MACH was used to impute the dataset to ~2.5 million SNPs using the HapMap Phase II, release 22 as reference set.

Healthy Lifestyle in Europe by Nutrition in Adolescence study (HELENA)

The recruitment and phenotyping of the adolescents participating in the HELENA cross-sectional study ("Healthy Lifestyle in Europe by Nutrition in Adolescence," www.helenastudy.com) have been described previously[26]. Briefly, a total of 3865 adolescents, aged 12–18 years old, were recruited between

2006 and 2007. Data were collected in 10 centers from 9 European countries: Athens (Greece), Dortmund (Germany), Ghent (Belgium), Heraklion (Greece), Lille (France), Pecs (Hungary), Rome (Italy), Stockholm (Sweden), Vienna (Austria) and Zaragoza (Spain). Adolescents were randomly selected from schools by using a proportional cluster sampling method and taking age into account. One-third of the classes were randomly selected for blood collection, resulting in a total of 1155 blood samples for the subsequent clinical biochemistry assays and genetic analyses. Data were collected on a detailed case report form in accordance with standardized procedures. In each centre, trained researchers carried out comprehensive physical examinations, including weight, height, and blood pressure measurements. The protocol was approved by the appropriate investigational review board for each investigating centre. Written informed consent was obtained from each adolescent and both of his or her parents or legal representatives. Participation in the study was voluntary.

DNA was extracted from white blood cells with the Puregene kit (QIAGEN, Courtaboeuf, France) and stored at –20°C. The harmonized, standardized anthropometric measurements were strictly monitored. Participants were barefoot and in underwear, and anthropometric measurements were taken by trained researchers. Weight was measured with an electronic set of scales (Type SECA 861; precision 0.05 kg) and height was measured in the Frankfort plane with a height gauge (Type SECA 225; precision 1 mm). The BMI was calculated. SNPs were genotyped using the KASPAR technology (KBioscience).

Young Hearts studies

The Young Hearts project is a prospective study investigating the development of biological and behavioral risk factors for cardiovascular disease in an adolescent population in Northern Ireland. Details of the study design and sampling procedure have been presented elsewhere[27]. Briefly, in 1989-1990, a 2 % representative sample of school children aged 12 and 15 years in Northern Ireland (YH3, n=1015) was collected. The original 12-year old population was followed up in 1992-1993 (YH2) with complete data collected on 225 males and 230 females (90% response rate). Between 1997 and 1999 all original YH

participants were invited to participate in the third screening phase (YH3: age 21-25 yrs, n=489) and a blood sample for DNA extraction was taken at that time.

A further cross-sectional survey, the Young Hearts 2000 (YH2000), was carried out in 2000. Approximately 2000 boys and girls aged 12 and 15 years (500 in each of the four age-sex groups) were recruited through post-primary schools. Details of the study design have been presented elsewhere[28].

Ethical approval was obtained from the Research Ethics Committee of the Queen's University of Belfast, and written informed consent for participation was obtained from all participants, and from each participant's parent or guardian. SNPs were genotyped using the KASPAR technology (KBioscience).

CHS

The Children's Health Study (CHS) is an ongoing cohort study in Southern California and with the primary aim of investigating the genetic and environmental risk factors related to respiratory health in children. Information regarding the study subjects, design, recruitment, and health assessment are reported elsewhere[29]. Briefly, recruitment of children took place between 1993 and 2002 across sixteen Southern California communities. At baseline, parents/guardians of children provided written informed consent as well as a completed questionnaire assessing health and environmental exposures. Attempts were made to collect buccal cell samples from all subjects for genetic analysis. In the spring of each subsequent year, anthropometric measures were taken by a trained technician, and an updated questionnaire was completed for each child by the parent/guardian.

We utilized weight and height measurements taken at baseline and at each annual follow-up time point to compute BMI (kg of weight/height in m^2). Our total sample of 311 cases and 330 controls represent subjects that consistently remained a case or control throughout the entire follow-up period. Given our subjects were initially identified based on their asthma status, we adjusted for this variable to ensure no confounding was occurring in our analysis. We excluded all subjects who were not self-reported as non-Hispanic White.

Buccal cells were collected from study subjects and genotyping was performed at the USC Epigenome Center using the Illumina HumanHap550, HumanHap550-Duo and Human610-Quad BeadChip microarrays. The program IMPUTE was used for imputation based on the HapMap phase 2 release 22 as the reference sample[30,31]. Quality scores for the replicated SNPs ranged from 1.0- 0.96 for the SNPs imputed. Genetic ancestry estimates were computed using the program STRUCTURE on 557 unlinked ancestry informative markers designed to distinguish between populations of European, African, Native American, and East Asian ancestry and loading factors were used to account for population stratification[32-34].

INfancia y Medio Ambiente [Environment and Childhood] Project (INMA)

Population-based birth cohorts were established as part of the INMA – INfancia y Medio Ambiente [Environment and Childhood] Project in several regions of Spain following a common protocol. This analysis uses the INMA cohorts of Valencia and Sabadell (Catalonia) established between 2003 and 2006. This project aims to study the associations between pre- and postnatal environmental exposures and growth, health, and development from early foetal life until adolescence and has been described previously in detail [35]. Pregnant women were enrolled during the 1st trimester of pregnancy at public primary health care centers or public hospitals. Detailed measurements were performed using ultrasound and physical examinations and biological samples. Informed consent was obtained from all participants and the study was approved by the Hospital Ethics Committees in each participating region.

Analyses were restricted to Caucasian individuals with genome-wide data and a BMI measurement available between 3.5 and 5 years of age. Weight and height were measured with calibrated scales and a wall-mounted stadiometer, respectively, with the children standing in light clothing and barefoot, by trained field workers. BMI was calculated as weight in kilograms divided by the square of height in meters. Sex- and age-adjusted standard deviation scores (SDS) were constructed using Growth Analyser 3.0 (http://www.growthanalyser.org; Dutch Growth Research

Foundation, Rotterdam, the Netherlands). The reference curve for body mass index in the Netherlands, 1997 was used[24].

Cord blood for DNA isolation was available for 909/1409 (64.5%) of all live-born participating children, and 752 samples of these samples were selected for the genome-wide genotyping using the HumanOmni1-Quad Beadchip (Illumina). Sex-mismatch rate between genome based sex and midwive-record based sex was 0, indicating that possible contamination of maternal DNA was extremely low.

Project Viva

Project Viva is an ongoing pre-birth cohort study in eastern Massachusetts, USA. Its primary goal is to examine pre- and peri-natal determinants of common childhood conditions and precursors to adult health outcomes, including growth, obesity, cardio-metabolic risk factors, asthma, allergy, cognition, and behavior. In 1999-2002, Project Viva staff recruited 2,128 women in the first trimester from 8 urban and suburban obstetric offices who went on to deliver live singleton infants at the two study hospitals in Boston. In-person examinations have occurred at 2 d, 6m, 3y, and 7y, at which staff measured length/height and weight. Data from electronic medical records, biosamples including umbilical cord blood, and annual questionnaires augment the in-person examinations. Approximately 30% of participants are non-white, i.e., non–European ancestry, and were excluded from this analysis. For this analysis, from white participants we chose 48 cases of obesity defined as BMI exceeding the 95th percentile for age and sex from the US Centers for Disease Control and Prevention 2000 growth charts, and 184 controls with BMI less than the 50th percentile. Analysis was performed using SAS 9.2 (Cary,NC), adjusted for age and gender.

DNA was isolated from cord blood, augmented with blood collection at the 3y visit. Genotyping for Project Viva samples was done at Children's Hospital, Boston using the Sequenom MassARRAY platform and the iPLEX genotyping protocol (Sequenom, Inc., San Diego, CA, USA). Oligos were purchased from Integrated DNA Technologies (Integrated DNA Technologies, Inc, Coralville, Iowa, USA). The Project

Viva study was conducted with approval from the local institutional review board, and informed consent was obtained from parents/guardians of all study participants.

Prevention and incidence of asthma and mite allergy birth cohort study (PIAMA)

PIAMA is a birth cohort study consisting of two parts: a placebo controlled intervention study in which the effect of mite impermeable mattress covers on the development of asthma and allergy was studied and a natural history study in which no intervention took place. Details of the study design have been published previously[36]. Recruitment took place in 1996-1997 through prenatal clinics. A screening questionnaire was distributed to pregnant women visiting one of 52 prenatal clinics at three regions in the Netherlands. A total of 10,232 pregnant women completed a validated screening questionnaire. Mothers reporting a history of asthma, current hay fever or allergy to pets or house dust mite were defined as allergic. Based on this screening, 7,862 women were invited to participate, of whom 4,146 women (1,327 allergic and 2,819 non-allergic) gave written informed consent. Follow-up of the children took place at 3 months of age and yearly from 1 to 8 years of age. The Medical Ethical Committees of the participating institutes approved the study, and all participants gave written informed consent. DNA was collected from 2,162 children, and height and weight measures were obtained in 2,440 children during medical examinations at age 4 and/or 8 years. Genome-wide genotyping was performed in two phases. The first phase was performed within the framework of the GABRIEL Consortium using an Illumina Human 610K quad array[37]. Genotypes were available from 172 children with asthma and from 187 controls after quality control. A second group of 268 children who were more extensively examined during follow up was genotyped with an Illumina HumanOmniExpress array. The current replication analysis was restricted to Caucasian individuals with genome-wide data and phenotype information available (n=153; cases=68 (44.4%)).

SCOOP-UK (Severe Childhood Onset Obesity Project United Kingdom)

The Severe Childhood Onset Obesity Project UK (SCOOP-UK) comprises 1500 UK Caucasian (self-reported) subjects with severe early onset obesity of unknown aetiology. This cohort has emerged out of the Genetics of Obesity Study (GOOS). The entry criteria for the GOOS cohort comprise a BMI > 3 SDS and an onset of obesity before the age of 10 years[38]. Several monogenic obesity syndromes have previously been identified from the GOOS cohort. SCOOP-UK represents a subgroup of GOOS patients of UK Caucasian ancestry in whom all the known monogenic obesity syndromes (including *MC4R* mutation carriers) have been excluded by direct nucleotide sequencing.

Genome-wide genotyping on patients from the SCOOP-UK cohort was carried out using the Affymetrix Human SNP Array 6.0 chip at the Wellcome Trust Sanger Institute, Cambridge, UK. Genotypes on the same platform were also available for additional obesity patients with developmental delay as reported elsewhere[39]. Publicly available genotype data on the same platform was available for population controls recruited from the Wellcome Trust Case Control Consortium 2 (WTCCC2 -http://www.wtccc.org.uk/ccc2). We leveraged the 3,000 dataset from the UK Blood Service Collection.

After sample and SNP quality control, data was available for 1509 patients (333 with developmental delay) and 2,674 controls. These were imputed to the HapMap Phase 2 reference panel in IMPUTE v1.0.0. Case control analyses under a log-additive model were performed using SNPTEST v1.1.5.

References

1. Rantakallio, P. Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr Scand* **193**, 1–71 (1966).

2. Jarvelin, M.R. et al. Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension* **44**, 838-46 (2004).

3. Sovio, U. et al. Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet* **5**, e1000409 (2009).

4. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* **35**, 34-41 (2006).

5. Craddock, N. et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713-20 (2010).

6. Todd, J.A. et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* **39**, 857-64 (2007).

7. Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-9 (2011).

8. Barrett, J.C. et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* **41**, 1330-4 (2009).

9. Meyre, D. et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* **41**, 157-9 (2009).

10. Visvikis-Siest, S. & Siest, G. The STANISLAS Cohort: a 10-year follow-up of supposed healthy families. Gene-environment interactions, reference values and evaluation of biomarkers in prevention of cardiovascular diseases. *Clin Chem Lab Med* **46**, 733-47 (2008).

11. Lafay, L. et al. Determinants and nature of dietary underreporting in a free-living population: the Fleurbaix Laventie Ville Sante (FLVS) Study. *Int J Obes Relat Metab Disord* **21**, 567-73 (1997).

12. Zutavern, A. et al. Timing of solid food introduction in relation to atopic dermatitis and atopic sensitization: results from a prospective birth cohort study. *Pediatrics* **117**, 401-11 (2006).

13. Rzehak, P. et al. Associations between BMI and the FTO gene are age dependent: results from the GINI and LISA birth cohort studies up to age 6 years. *Obes Facts* **3**, 173-80.

14. Scherag, A. et al. Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genet* **6**, e1000916 (2010).

15. Jarick, I. et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet* **20**, 840-52 (2011).

16. Hebebrand, J., Heseker, H., Himmelmann, G.W., Schäfer, H. & Remschmidt, H. Altersperzentilen fuer den Body Mass Index aus Daten der Nationalen Verzehrstudie einschliesslich einer Übersicht zu relevanten Einflussfaktoren. *Aktuelle Ernährungsmedizin* **19**, 259-265 (1994).

17. Bisgaard, H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann Allergy Asthma Immunol* **93**, 381-9 (2004).

18. Bisgaard, H., Hermansen, M.N., Loland, L., Halkjaer, L.B. & Buchvald, F. Intermittent inhaled corticosteroids in infants with episodic wheezing. *N Engl J Med* **354**, 1998-2005 (2006).

19. Bisgaard, H. et al. Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med* **357**, 1487-95 (2007).

20. Hakonarson, H. et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448**, 591-4 (2007).

21. Paternoster, L. et al. Genome-Wide Population-Based Association Study of Extremely Overweight Young Adults - The GOYA Study. *PLoS One* **6**, e24303 (2011).

22. Jaddoe, V.W. et al. The Generation R Study Biobank: a resource for epidemiological studies in children and their parents. *Eur J Epidemiol* **22**, 917-23 (2007).

23. Jaddoe, V.W. et al. The Generation R Study: design and cohort update 2010. *Eur J Epidemiol* **25**, 823-41 (2010).

24. Fredriks, A.M. et al. Continuing positive secular growth change in The Netherlands 1955-1997. *Pediatr Res* **47**, 316-23 (2000).

25. Estrada, K. et al. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* **25**, 2750-2 (2009).

26. Moreno, L.A. et al. Design and implementation of the Healthy Lifestyle in Europe by Nutrition in Adolescence Cross-Sectional Study. *Int J Obes (Lond)* **32 Suppl 5**, S4-11 (2008).

27. Boreham, C., Savage, J.M., Primrose, D., Cran, G. & Strain, J. Coronary risk factors in schoolchildren. *Arch Dis Child* **68**, 182-6 (1993).

28. Watkins, D.C. et al. Ten-year trends for fatness in Northern Irish adolescents: the Young Hearts Projects--repeat cross-sectional study. *Int J Obes (Lond)* **29**, 579-85 (2005).

29. Navidi, W., Thomas, D., Stram, D. & Peters, J. Design and analysis of multilevel analytic studies with applications to a study of air pollution. *Environ Health Perspect* **102 Suppl 8**, 25-32 (1994).

30. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).

31. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

32. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-87 (2003).

33. Falush, D., Stephens, M. & Pritchard, J.K. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* **7**, 574-578 (2007).

34. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).

35. Guxens, M. et al. Cohort Profile: The INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. *Int J Epidemiol* (2011).

36. Brunekreef, B. et al. The prevention and incidence of asthma and mite allergy (PIAMA) birth cohort study: design and first results. *Pediatr Allergy Immunol* **13 Suppl 15**, 55-60 (2002).

37. Moffatt, M.F. et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).

38. Cole, T.J., Freeman, J.V. & Preece, M.A. Body mass index reference curves for the UK, 1990. *Arch Dis Child* **73**, 25-9 (1995).

39. Bochukova, E.G. et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666-70 (2010).

40. Speliotes, E.K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).

Early Growth Genetics Consortium (EGG) Membership and Affiliations

Linda S. Adair[1], Wei Ang[2], Mustafa Atalay[3], Toos van Beijsterveldt[4], Nienke Bergen[5,6], Kelly Benke[2], Diane Berry[7], Jonathan Bradfield[8], Mariona Bustamente[9,10,11,12], Pimphen Charoen[13,14], Lachlan Coin[13], Diana Cousminer[15], Shikta Das[13], Oliver S.P. Davis[16], Paul Elliott[17], Dave M. Evans[18], Bjarke Feenstra[19], Claudia Flexeder[20], Tim Frayling[21], Rachel Freathy[18,21], Romy Gaillard[5,6], Frank Geller[19], Maria Groen-Blokhuis[4], Liang-Kee Goh[22,23], Mònica Guxens[9,10,11], Claire M.A. Haworth[16], Dexter Hadley[8], Johannes Hedebrand[24], Anke Hinney[24], Joel N. Hirschhorn[25,26,27], John W. Holloway[28,29], Claus Holst[30], Jouke Jan Hottenga[4], Momoko Horikoshi[31,32], Ville Huikari[33], Elina Hypponen[7,34], Carmen Iñiguez[10,35], Marika Kaakinen[33,36], Tuomas O. Kilpeläinen[37], Mirna Kirin[38], Mattew Kowgier[2], Hanna-Maaria Lakka[39], Leslie A. Lange[40], Debbie A. Lawlor[18], Terho Lehtimäki[41,42], Alex Lewin[13], Cecilia Lindgren[43], Virpi Lindi[3], Reedik Mägi[43,44], Julie Marsh[2], Christel Middeldorp[4], Iona Millwood[13,45], Dennis O. Mook-Kanamori[5,6,46,47], Jeffrey C. Murray[48], Michel Nivard[4], Ellen Aagaard Nohr[30], Ioanna Ntalla[49], Emily Oken [25,26,27], Paul O'Reilly[13], Lyle Palmer[50,51], Kalliope Panoutsopoulou[52], Jennifer Pararajasingham[16], Inga Prokopenko[31,32], Alina Rodriguez[13,16,53], Rany M. Salem[25,26,27], Sylvain Sebert[13], Niina Siitonen[54], Ulla Sovio[13,55], Beate St Pourcain[18], David P. Strachan[56], Jordi Sunyer[9,10,11,57], H. Rob Taal[5,6,46], Yik-Ying Teo[23], Elisabeth Thiering[20], Carla Tiesler[20,58], Andre G. Uitterlinden[6,59], Beatriz Valcárcel[13], Nicole Warrington[2,50], Scott White [2], Gonneke Willemsen[4], Hanieh Yaghootkar[21], Eleftheria Zeggini[52], Dorret I. Boomsma[4], Cyrus Cooper[60], Xavier Estivill[10,57,61], Matthew Gillman[62], Struan F.A. Grant[8,63,64], Hakon Hakonarson[8,63,64], Andrew T. Hattersley[65], Joachim Heinrich[20], Berthold Hocher[66,67], Vincent W.V. Jaddoe[5,6,46], Marjo-Riitta Jarvelin[17,33,36,68], Timo A. Lakka[3], Mark I. McCarthy[31,32,69], Mads Melbye[19], Karen L. Mohlke[40], George V. Dedoussis[49], Ken K. Ong[70], Ewan R. Pearson[71], Craig E. Pennell[2], Thomas S. Price[16] , Chris Power[7], Olli T. Raitakari[54,72], Seang-Mei Saw[22,23,73], Andre Scherag[74], Olli Simell[54,75], Thorkild I.A. Sørensen[30,76], Jordi Sunyer[9,10,11,77], Nicholas J. Timpson[18], Elisabeth Widen[15], James F. Wilson[38,78]

1. Department of Nutrition, University of North Carolina, Chapel Hill, NC.
2. School of Women's and Infants' Health, The University of Western Australia, Perth, Australia
3. Department of Physiology, Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland.
4. Department of Biological Psychology, VU University, Amsterdam, The Netherlands.
5. The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands.
6. Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands.
7. Centre For Paediatric Epidemiolgy and Biostatistics/MRC Centre of Epidemiology for Child Health, University College of London Institute of Child Health, London, UK.
8. Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
9. Hospital del Mar Research Institute (IMIM), Barcelona, Catalonia, Spain.
10. CIBER Epidemiologia y Salud Pública (CIBERESP), Barcelona, Catalonia, Spain.
11. Center for Research in Environmental Epidemiology (CREAL), Barcelona, Catalonia, Spain.
12. Genes and Disease Program, Centre for Genomic Regulation (CRG) and UPF, Barcelona, Catalonia, Spain.
13. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK.
14. Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.
15. Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.
16. MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK.
17. MRC-HPA Center, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.
18. MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, UK.
19. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.
20. Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
21. Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Magdalen Road, Exeter, EX1 2LU, UK.

22. Duke-NUS Graduate Medical School, Singapore.
23. Saw Swee Hock School of Public Health, National University of Singapore.
24. Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen, Germany.
25. Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts 02115, USA.
26. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
27. Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA.
28. Human Genetics and Medical Genomics, Human Development & Health, Faculty of Medicine, University of Southampton.
29. Clinical & Experimental Sciences, Faculty of Medicine, University of Southampton.
30. Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark.
31. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK.
32. Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.
33. Institute of Health Sciences, University of Oulu, Finland
34. Department of Genomics of Common Disease, School of Public Health, Imperial College London.
35. Division of Environment and Health, Center for Public Health Research-CSISP, Valencia, Spain.
36. Biocenter Oulu, University of Oulu, Finland.
37. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.
38. Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland.
39. Department of Public Health, Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio Campus, Finland.
40. Department of Genetics, University of North Carolina, Chapel Hill, NC.
41. Department of Clinical Chemistry, Tampere University Hospital, Tampere, Finland.
42. Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland
43. Genetic and Genomic Epidemiology Unit, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
44. Estonian Genome Center, University of Tartu, Tartu, Estonia.
45. Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, UK.
46. Department of Paediatrics, Erasmus Medical Center, Rotterdam, the Netherlands.
47. Department of Physiology and Biophysics, Weill Cornell Medical College - Qatar, Doha, Qatar.
48. Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA.
49. Department of Dietetics - Nutrition, Harokopio University of Athens, Athens, Greece.
50. Samuel Lunenfeld Research Institute, University of Toronto, Toronto, Canada
51. Genetic Epidemiology and Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario, Canada
52. Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.
53. Department of Psychology, Mid Sweden University, Sweden.
54. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
55. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom.
56. Division of Population Health Sciences and Education, St George's, University of London
57. Pompeu Fabra University (UPF), Barcelona, Catalonia, Spain.
58. Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany.
59. Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands.
60. MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom.
61. Genes and Disease Program, Center for Genomic Regulation (CRG-UPF), Barcelona, Catalonia, Spain.
62. Obesity Prevention Program, Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Obesity Prevention Program, Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, Boston, MA 02215 USA.
63. Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
64. Department of Pediatrics, University of Pennsylvania, Philadelphia PA 19104, USA.
65. Peninsula NIHR Clinical Research Facility, Peninsula College of Medicine and Dentistry, University of Exeter, Barrack Road, Exeter, EX2 5DW, UK.

66. Institute of Nutritional Science, University of Potsdam, D-14558 Nuthetal Potsdam, Germany.
67. Center for Cardiovascular Research/Institute of Pharmacology, Charité, Berlin, Germany.
68. National Institute of Health and Welfare, Oulu, Finland.
69. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK.
70. MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, CB2 0QQ, United Kingdom.
71. Biomedical Research Institute, University of Dundee.
72. Department of Clinical Physiology, University of Turku and Turku University Hospital, Turku, Finland.
73. Singapore Eye Research Institute.
74. Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany.
75. Department of Pediatrics, University of Turku and Turku University Hospital, Turku, Finland.
76. The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copnahgen.
77. Pompeu Fabra University (UPF), Barcelona, Catalonia, Spain.
78. MRC Institute of Genetics and Molecular Medicine at the University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, Scotland

Supplementary Table 1: Descriptive Table of the cohorts used in the discovery and comparable replication stages.

Study cohort	Year(s) of birth	Age range of BMI measurements (years)	Time points at which BMI was measured (years)	Median number of BMI measurements (95% range)	Total N (with GWAS data and BMI available at any timepoint)	N Cases (% of total cohort)	N Controls (% of total cohort)	% Male cases/controls	Among cases				
									Median number of measurements >95th percentile (95% range)	% dropping <95th percentile at later measurements	% dropping <95th percentile later and meeting >95th percentile after that	% persistent >95th percentile at later measurements	% with no measurement available after meeting case definition
DISCOVERY													
ALSPAC*	1991-1992	0.1 – 12.5	11 measures	6.5 (6-11)	3,233	976 (30.2%)	1,244 (38.5%)	54.4%/51.5%	5 (1-6)	78%	13%	7%	2%
NFBC1966	1966	2.0-18.9	2.0-18.9	12 (5-21)	3,596	700 (19.5%)	521 (14.5%)	52.4%/48.3%	2 (1-12)	58.1%	17.5%	18.4%	6.1%
B58C-T1DGC	1958	7-16	7,11,16 years	3 (1-3)	2,446	192 (7.9%)	367 (15.0%)	50.5%/48.2%	1 (1-3)	35.8%	1.6%	31.6%	31.0%
B58C-WTCCC	1958	7-16	7,11,16 years	3 (1-3)	2,621	188 (7.2%)	428 (16.3%)	50.0%/51.4%	1 (1-3)	38.4%	1.1%	29.7%	30.8%
FRENCH YOUNG	1981-2002	6-17	At inclusion	1	1,019	670 (30.1%)	349 (15.5%)	44.8%/49.9%	1	NA	NA	NA	100%
LISA	1998-1999	2-10	2,4,5,10 years	3 (2-4)	396	27 (6.8%)	250 (63.1%)	66.7%/54.4%	1 (1-3)	40.7%	3.7%	25.9%	29.6%
RAINE	1989-1991	2-18	2, 3, 5, 8, 10, 14, 17	6 (5-6)	1,196	232 (19.4%)	125 (10.5%)	58.2%/54.4%	2 (1-3)	40.1%	7.3%	35.8%	16.8%
CHOP	1988-Present	2-18	At inclusion	1	9,159	1611 (17.6%)	2760 (30.1%)	57.0%/54.9%	1	NA	NA	NA	100%
ESSEN**	cases: 1977-1997, controls: 1944-1988	Cases: 5-18	At inclusion	1	Cases: 487 controls: 442	397 (82%)	435 (98%)	43.3%/39.3%	1	NA	NA	NA	100%
HBCS	1934-1944	2-12	2,3,4,5,6,7,8, 9,10,11,12	10 (7-11)	1,566	260 (16.6%)	405 (25.9%)	45.4% / 44.7%	2(1-8)	64.1%	14.6%	7.2%	14.1%
YF	1962-1977	3-18	3,6,9,12, 15,18	3 (1-3)	2,443	166 (6.8%)	537 (22.0%)	47.6% / 46.9%	1(1-3)	23.8%	11.3%	31.0%	33.9%

COPSAC	1998-2001	9	9	1	382	62 (16.2%)	99 (25.9%)	40.0%/48.6%	1	NA	NA	NA	100%
CM-GOYA	1937-1969	7-13	7,8,9,10, 11,12,13	7 (6-7)	206	21 (10.2%)	34 (16.5%)	100%/100%	2 (1-7)	42.9%	23.8%	23.8%	9.5%
GENERATIONR	2002-2006	2-6	2,3,4,6	3 (2-3)	2,326	192 (8.3%)	724 (31.1%)	45.8%/49.3%	1 (1-3)	52.6%	4.7%	17.2%	25.0%
REPLICATION													
HELENA	1988-1994	11.5 – 17.8	11.5 – 17.8	1	1,128	56 (5.0%)	563 (49.9%)	62.5%/48.9%	1	NA	NA	NA	100%
Young Hearts studies	1974-1977 1985-1990	14.7 - 16.2	14.7 - 16.2	1	901	44 (4.9%)	450 (50.0%)	27.3%/60.2%	1	NA	NA	NA	100%
LISA+GINI	1995-1999	2-10	2,4,5,10	3 (2-4)	761	40 (5.3%)	457 (60.1%)	47.5%/51.0%	1 (1-3)	47.5%	7.5%	20.0%	25.0%
CHS	1997-2006	5-14	Annual measurements for up to 8 years	4 (2-8)	1,729	311 (18.0%)	330 (19.1%)	62.7%/53.0%	3 (1-9)	41.2%	13.8%	49.5%	9.6%
ALSPAC*	1991-1992	0.1-15.5	11 measures	5.5 (2-11)	6,180	1,452 (22.8%)	1,045 (16.43%)	51.7%/ 50.1%	5 (1-5)	75%	12%	4%	8%
INMA	2003-2006	3.5-5	4	1	603	55 (9.1%)	213 (35.3%)	54.6%/51.2%	1	NA	NA	NA	100%
VIVA	1999-2002	2.8-6.0	3	1	582	48 (8.2%)	184 (31.6%)	60.4%/50.5%	1	NA	NA	NA	100%
PIAMA	1996-1997	4-8	4 and 8	2 (1-2)	666	68 (10.2%)	85 (12.8%)	55.9%/58.8%	2 (1-2)	25.0%	NA	22.1%	52.9%
NFBC1986	1986	2.0-18.9	2.0-18.9	12 (6-17)	4,453	744 (16.7%)	749 (16.8%)	49.1%/51.2%	2(1-12)	60.8%	13.3%	20.9%	5.0%

**ALSPAC was the only cohort to factor in subjects less than 2 years old in their consideration of trait definition. While the other studies mostly did not have weight and height measures available before the age of 2 years, except for birth weight, ALSPAC had height and weight hence considering it a metric of early adiposity.

** The ESSEN study is not a (birth) cohort study and information on year of birth is sometimes lacking; the study is a cross-sectional study with cases that are on average more extreme than the ≥95th percentile; further details can be found in the supplementary material text, in Hinney et al. (2007) and Scherag et al. (2010)

Supplementary Table 2: Top signals at each locus (all known) that reach genome wide significance ($P < 5x10^{-8}$) sorted by *P*-value, in the discovery stage (5,530 cases and 8,318 controls).

SNP	Chr	Pos	Allele1	Allele2	Freq1	OR [95% C.I.]	P-value	Direction	Nearest Gene
rs9941349	16	52382989	T	C	0.442	1.219 [1.158, 1.282]	$1.99x10^{-14}$	++-++-+++++++-	*FTO*
rs4854344	2	628144	T	G	0.804	1.277 [1.194, 1.365]	$7.36x10^{-13}$	++++++++++++-+	*TMEM18*
rs6752378	2	25003620	A	C	0.458	1.185 [1.127, 1.245]	$2.81x10^{-11}$	+++++++++++++++	*POMC*
rs663129	18	55989381	A	G	0.245	1.220 [1.151, 1.295]	$3.55x10^{-11}$	++++++++++-+-	*MC4R*
rs7138803	12	48533735	A	G	0.369	1.182 [1.123, 1.245]	$2.09x10^{-10}$	+++++++++++++++	*FAIM2*
rs1040070	1	74750458	C	G	0.537	0.861 [0.818, 0.907]	$1.07x10^{-8}$	----------+--	*TNNI3K*
rs10913469	1	176180142	T	C	0.792	0.838 [0.786, 0.892]	$3.18x10^{-8}$	----------+++	*SEC16B*

1. Avon Longitudinal Study of Parents And Children (ALSPAC)
2: Northern Finland 1966 Birth Cohort (NFBC1966)
3. British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium subset (B58C-T1DGC)
4. British 1958 Birth Cohort – Wellcome Trust Case Control Consortium Subset (B58C-WTCCC)
5. French Young study (FRENCH YOUNG) PCA adjusted
6. Lifestyle Immune System Allergy Study (LISA)
7: Western Australian Pregnancy Cohort study (RAINE)
8. Children's Hospital of Philadelphia (CHOP) PCA adjusted
9. Essen Obesity Study (ESSEN) PCA adjusted
10. Helsinki Birth Cohort Study (HBCS)
11. Cardiovascular Risk in Young Finns Study (YF)
12. Copenhagen Study on Asthma in Childhood (COPSAC)
13. CM-GOYA study (CM-GOYA)
14. Generation R Study (GENERATIONR)

Supplementary Table 3: Using genomic control on the overall meta-analysis results: Top signals at each locus (all known) that reach genome wide significance in Supplementary Table 2 sorted by *P*-value, in the discovery stage (5,530 cases and 8,318 controls)

SNP	Chr	Pos	Allele1	Allele2	Freq1	OR [95% C.I.]	P-value	Direction	Nearest Gene
rs9941349	16	52382989	T	C	0.442	1.219 [1.157, 1.284]	$1.16x10^{-13}$	++-++-+++++++-	*FTO*
rs4854344	2	628144	T	G	0.804	1.277 [1.192, 1.368]	$3.22x10^{-12}$	++++++++++++-+	*TMEM18*
rs6752378	2	25003620	A	C	0.458	1.185 [1.125, 1.247]	$1.05x10^{-10}$	+++++++++++++++	*POMC*
rs663129	18	55989381	A	G	0.245	1.220 [1.148, 1.297]	$1.27x10^{-10}$	++++++++++-+-	*MC4R*
rs7138803	12	48533735	A	G	0.369	1.182 [1.121, 1.246]	$6.50x10^{-10}$	+++++++++++++++	*FAIM2*
rs1040070	1	74750458	C	G	0.537	0.861 [0.817, 0.908]	$2.78x10^{-8}$	----------+--	*TNNI3K*
rs10913469	1	176180142	T	C	0.792	0.838 [0.785, 0.893]	$7.99x10^{-8}$	----------+++	*SEC16B*

1. Avon Longitudinal Study of Parents And Children (ALSPAC)
2: Northern Finland 1966 Birth Cohort (NFBC1966)
3. British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium subset (B58C-T1DGC)
4. British 1958 Birth Cohort – Wellcome Trust Case Control Consortium Subset (B58C-WTCCC)
5. French Young study (FRENCH YOUNG) PCA adjusted
6. Lifestyle Immune System Allergy Study (LISA)
7: Western Australian Pregnancy Cohort study (RAINE)
8. Children's Hospital of Philadelphia (CHOP) PCA adjusted
9. Essen Obesity Study (ESSEN) PCA adjusted
10. Helsinki Birth Cohort Study (HBCS)
11. Cardiovascular Risk in Young Finns Study (YF)
12. Copenhagen Study on Asthma in Childhood (COPSAC)
13. CM-GOYA study (CM-GOYA)
14. Generation R Study (GENERATIONR)

Supplementary Table 4: Novel signals at each locus that did not reach genome wide significance but yielded $P < 5x10^{-6}$, sorted by chromosomal location, in the discovery stage (5,530 cases and 8,318 controls). The outcome of the replication effort of the eight loci taken forward in to nine comparable independent cohorts (n = 2,818 cases and 4,083 controls) is also indicated. Separate discovery and replication data plus combined data are shown; those signals achieving genome wide significance are indicated in red bold. The 'Het *P*-value' is also indicated as a test of heterogeneity in the Discovery cohort, of which none were significant.

	Locus	SNP	Allele1/2	Nearest Gene	Direction	OR [95% C.I.]	*P*-value	Freq Allele 1	Het *P*-value
Discovery									
	1p36	rs2300095	A/G	*MTOR-ANGPTL7*	++++--++++++++	1.140 [1.079, 1.206]	$3.69x10^{-6}$	0.318	0.716
	4q25	rs4833407	A/C	*ALPK1*	+++++++++++++++	1.130 [1.075, 1.189]	$2.07x10^{-6}$	0.420	0.431
	4q28	rs4864201	T/C	*BC041448*	0++++++++++-++	1.145 [1.085, 1.208]	$7.03x10^{-7}$	0.366	0.244
	5q12	rs28636	T/C	*MAST4*	----+------++-	0.863 [0.813, 0.916]	$1.58x10^{-6}$	0.240	0.492
	10p11	rs1290002	A/G	*PARD3*	++++++++++++-	1.137 [1.077, 1.200]	$3.03x10^{-6}$	0.310	0.384
	13q14	rs9568856	A/G	*OLFM4*	++++++++++++-+	1.210 [1.123, 1.305]	$6.58x10^{-7}$	0.158	0.599
	17q21	rs9299	T/C	*HOXB5*	++-++-++++++++	1.144 [1.084, 1.207]	$9.12x10^{-7}$	0.652	0.483
	18q12	rs17697518	T/C	*KC6*	+++++++++++++++	1.204 [1.118, 1.296]	$9.03x10^{-7}$	0.140	0.346
Replication									
	1p36	rs2300095	A/G	*MTOR-ANGPTL7*	++++-+---	1.020 [0.937, 1.110]	0.645		
	4q25	rs4833407	A/C	*ALPK1*	+++---+++	1.069 [0.989, 1.155]	0.0913		
	4q28	rs4864201	T/C	*BC041448*	+++-++-++	1.081 [0.999, 1.169]	0.0519		
	5q12	rs28636	T/C	*MAST4*	-+-+-+--+	0.995 [0.909, 1.087]	0.904		
	10p11	rs1290002	A/G	*PARD3*	??++-?--+	1.019 [0.936, 1.109]	0.669		
	13q14	rs9568856	A/G	*OLFM4*	+-++++-++	1.225 [1.089, 1.378]	$7.13x10^{-4}$		
	17q21	rs9299	T/C	*HOXB5*	++-++++++	1.145 [1.056, 1.242]	0.00104		
	18q12	rs17697518	T/C	*KC6*	+-+++-+-+	1.092 [0.980, 1.217]	0.112		
Combined									
	1p36	rs2300095	A/G	*MTOR-ANGPTL7*		1.103 [1.052, 1.155]	$3.80x10^{-5}$		
	4q25	rs4833407	A/C	*ALPK1*		1.111 [1.065, 1.160]	$9.79x10^{-7}$		
	4q28	rs4864201	T/C	*BC041448*		1.124 [1.076, 1.175]	$2.07x10^{-7}$		
	5q12	rs28636	T/C	*MAST4*		0.902 [0.858, 0.948]	$5.10x10^{-5}$		
	10p11	rs1290002	A/G	*PARD3*		1.102 [1.053, 1.153]	$3.08x10^{-5}$		
	13q14	**rs9568856**	**A/G**	***OLFM4***		**1.215 [1.140, 1.294]**	$\mathbf{1.82x10^{-9}}$		
	17q21	**rs9299**	**T/C**	***HOXB5***		**1.144 [1.094, 1.196]**	$\mathbf{3.54x10^{-9}}$		
	18q12	rs17697518	T/C	*KC6*		1.167 [1.098, 1.241]	$7.32x10^{-7}$		

Discovery cohorts

1. Avon Longitudinal Study of Parents and Children (ALSPAC)
2: Northern Finland 1966 Birth Cohort (NFBC1966)
3. British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium subset (B58C-T1DGC)
4. British 1958 Birth Cohort – Wellcome Trust Case Control Consortium Subset (B58C-WTCCC)
5. French Young study (FRENCH YOUNG) PCA adjusted
6. Lifestyle Immune System Allergy Study (LISA)
7: Western Australian Pregnancy Cohort study (RAINE)
8. Children's Hospital of Philadelphia (CHOP) PCA adjusted
9. Essen Obesity Study (ESSEN) PCA adjusted
10. Helsinki Birth Cohort Study (HBCS)
11. Cardiovascular Risk in Young Finns Study (YF)
12. Copenhagen Study on Asthma in Childhood (COPSAC)
13. CM-GOYA study (CM-GOYA)
14. Generation R Study (GENERATIONR)

Comparable replication cohorts

1. Healthy Lifestyle in Europe by Nutrition in Adolescence study (HELENA)
2. Young Hearts studies
3. Lifestyle – Immune System – Allergy Study plus German Infant Study on the influence of Nutrition Intervention (LISA+GINI)
4. Children's Health Study (CHS)
5. INfancia y Medio Ambiente [Environment and Childhood] Project (INMA) (2 proxies used - rs965013 and rs11099020 for rs4833407)
6. Project Viva (VIVA)
7. Prevention and incidence of asthma and mite allergy birth cohort study (PIAMA)
8. Northern Finland 1986 Birth Cohort (NFBC1986) (2 proxies used - rs10779751 for rs230095; rs4883723 for rs9568856)
9. Avon Longitudinal Study of Parents and Children (ALSPAC)

Supplementary Table 5: Using genomic control on the overall meta-analysis results: Signals at each locus that did not reach genome wide significance but yielded *P* < 5x10^{-6} in Table 1, sorted by *P*-value, in the discovery stage (5,530 cases and 8,318 controls).

SNP	Chr	Pos	Allele1	Allele2	Freq1	OR [95% C.I.]	P-value	Direction	Nearest Gene
rs13130484	4	44870448	T	C	0.451	1.154 [1.094, 1.217]	1.30x10^{-7}	++-++++++++++	*GNPDA2**
rs9568856	13	52962982	A	G	0.158	1.210 [1.120, 1.308]	1.36x10^{-6}	++++++++++++-+	*OLFM4*
rs4864201	4	130950734	T	C	0.366	1.145 [1.084, 1.210]	1.41x10^{-6}	0++++++++++-++	*BC041448*
rs17697518	18	37019657	T	C	0.140	1.204 [1.115, 1.299]	1.85x10^{-6}	+++++++++++++	*KC6*
rs9299	17	44024429	T	C	0.652	1.144 [1.082, 1.209]	1.91x10^{-6}	++-++-++++++++	*HOXB5*
rs28636	5	66184869	T	C	0.240	0.863 [0.811, 0.918]	3.08x10^{-6}	----+------++-	*MAST4*
rs4833407	4	113531239	A	C	0.420	1.130 [1.073, 1.191]	3.88x10^{-6}	+++++++++++++	*ALPK1*
rs1290002	10	34506662	A	G	0.310	1.137 [1.076, 1.202]	5.70x10^{-6}	++++++++++++-	*PARD3*
rs2300095	1	11188304	A	G	0.318	1.140 [1.077, 1.207]	7.03x10^{-6}	++++--++++++++	*MTOR-ANGPTL7*

*Known locus

1. Avon Longitudinal Study of Parents and Children (ALSPAC)
2: Northern Finland 1966 Birth Cohort (NFBC1966)
3. British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium subset (B58C-T1DGC)
4. British 1958 Birth Cohort – Wellcome Trust Case Control Consortium Subset (B58C-WTCCC)
5. French Young study (FRENCH YOUNG) PCA adjusted
6. Lifestyle Immune System Allergy Study (LISA)
7: Western Australian Pregnancy Cohort study (RAINE)
8. Children's Hospital of Philadelphia (CHOP) PCA adjusted
9. Essen Obesity Study (ESSEN) PCA adjusted
10. Helsinki Birth Cohort Study (HBCS)
11. Cardiovascular Risk in Young Finns Study (YF)
12. Copenhagen Study on Asthma in Childhood (COPSAC)
13. CM-GOYA study (CM-GOYA)
14. Generation R Study (GENERATIONR)

Supplementary Table 6: Exploration of the eight loci in two extreme independent cohorts (705 trios; 1,509 cases and 2,674 controls), sorted by chromosomal location. Separate replication data and combined data are shown along with an overall assessment of all data used in the study. Those signals achieving genome wide significance are indicated in red bold.

				Extreme obesity replication samples			Combined with Discovery		Meta-analysis of all discovery and replication cohorts	
SNP	**Chr**	**Pos**	**Nearest Gene**	**OR [95% C.I.]**	**P-value**	**Direction**	**OR [95% C.I.]**	**P-value**	**OR [95% C.I.]**	**P-value**
rs2300095	1	11188304	*MTOR-ANGPTL7*	1.014 [0.931, 1.105]	0.745	+-	1.102 [1.051, 1.154]	4.88×10^{-5}	1.082 [1.039, 1.127]	1.58×10^{-4}
rs4833407	4	113531239	*ALPK1*	1.097 [1.014, 1.187]	0.0211	++	1.121 [1.074, 1.169]	1.61×10^{-7}	1.108 [1.068, 1.150]	6.51×10^{-8}
rs4864201	4	130950734	*BC041448*	1.055 [0.974, 1.143]	0.187	++	1.117 [1.068, 1.167]	1.20×10^{-6}	1.108 [1.066, 1.151]	2.19×10^{-7}
rs28636	5	66184869	*MAST4*	0.990 [0.902, 1.086]	0.826	+-	0.899 [0.854, 0.945]	3.31×10^{-5}	0.921 [0.881, 0.962]	2.39×10^{-4}
rs1290002	10	34506662	*PARD3*	1.013 [0.931, 1.102]	0.764	-+	1.100 [1.051, 1.151]	4.16×10^{-5}	1.081 [1.039, 1.125]	1.38×10^{-4}
rs9568856	**13**	**52962982**	***OLFM4***	**1.072 [0.958, 1.200]**	**0.223**	**-+**	**1.166 [1.095, 1.241]**	$\mathbf{1.51 \times 10^{-6}}$	**1.179 [1.115, 1.246]**	$\mathbf{5.33 \times 10^{-9}}$
rs9299	**17**	**44024429**	***HOXB5***	**1.003 [0.866, 1.161]**	**0.970**	**+?**	**1.126 [1.071, 1.184]**	$\mathbf{3.78 \times 10^{-6}}$	**1.131 [1.084, 1.181]**	$\mathbf{1.54 \times 10^{-8}}$
rs17697518	18	37019657	*KC6*	0.986 [0.878, 1.107]	0.810	-+	1.136 [1.068, 1.209]	6.02×10^{-5}	1.125 [1.066, 1.188]	1.96×10^{-5}

Supplementary Table 7. The key signals when restricting the cohorts to defining cases and controls over the age of 2 years old (4,829 cases and 8,294 controls).

SNP	Chr	Pos	Allele1	Allele2	Freq1	OR [95% C.I.]	P-value	Direction	Nearest Gene
rs2300095	1	11188304	A	G	0.320	1.129 [1.065, 1.197]	4.70×10^{-5}	++++--++++++++	*MTOR-ANGPTL7*
rs4833407	4	113531239	A	C	0.420	1.149 [1.090, 1.212]	3.05×10^{-7}	+++++++++++++++	*ALPK1*
rs4864201	4	130950734	T	C	0.370	1.161 [1.098, 1.228]	1.63×10^{-7}	++++++++++-++	*BC041448*
rs28636	5	66184869	T	C	0.242	0.863 [0.810, 0.920]	6.26×10^{-6}	----+------++-	*MAST4*
rs1290002	10	34506662	A	G	0.312	1.121 [1.059, 1.187]	8.05×10^{-5}	-++++++++++-	*PARD3*
rs9568856	13	52962982	A	G	0.160	1.240 [1.146, 1.341]	9.19×10^{-8}	++++++++++-+	*OLFM4*
rs9299	17	44024429	T	C	0.651	1.140 [1.078, 1.207]	5.38×10^{-6}	++-++-++++++++	*HOXB5*
rs17697518	18	37019657	T	C	0.141	1.217 [1.126, 1.315]	7.28×10^{-7}	+++++++++++++	*KC6*

1. Avon Longitudinal Study of Parents and Children (ALSPAC)
2: Northern Finland 1966 Birth Cohort (NFBC1966)
3. British 1958 Birth Cohort – Type 1 Diabetes Genetics Consortium subset (B58C-T1DGC)
4. British 1958 Birth Cohort – Wellcome Trust Case Control Consortium Subset (B58C-WTCCC)
5. French Young study (FRENCH YOUNG) PCA adjusted
6. Lifestyle Immune System Allergy Study (LISA)
7: Western Australian Pregnancy Cohort study (RAINE)
8. Children's Hospital of Philadelphia (CHOP) PCA adjusted
9. Essen Obesity Study (ESSEN) PCA adjusted
10. Helsinki Birth Cohort Study (HBCS)
11. Cardiovascular Risk in Young Finns Study (YF)
12. Copenhagen Study on Asthma in Childhood (COPSAC)
13. CM-GOYA study (CM-GOYA)
14. Generation R Study (GENERATIONR)

Supplementary Table 8. Replication effort of the eight loci taken forward in to nine comparable independent cohorts (n = 1,750 cases and 4,017 controls) when restricting the cohorts to defining cases and controls over the age of 2 years old, sorted by chromosomal location. Both separate replication data and combined data are shown. Those signals achieving genome wide significance are indicated in red bold.

				Comparable replication samples			Combined with Discovery	
SNP	**Chr**	**Pos**	**Nearest Gene**	**OR [95% C.I.]**	**P-value**	**Direction**	**OR [95% C.I.]**	**P-value**
rs2300095	1	11188304	*MTOR-ANGPTL7*	1.047 [0.949, 1.155]	0.361	++++-+--+	1.107 [1.053, 1.164]	7.33×10^{-5}
rs4833407	4	113531239	*ALPK1*	1.082 [0.990, 1.182]	0.0813	+++---+++	1.131 [1.081, 1.184]	1.25×10^{-7}
rs4864201	**4**	**130950734**	***BC041448***	**1.108 [1.013, 1.212]**	**0.0251**	**+++-++-++**	**1.146 [1.093, 1.202]**	**1.81×10^{-8}**
rs28636	5	66184869	*MAST4*	0.963 [0.868, 1.068]	0.473	-+-+-+--+	0.889 [0.842, 0.939]	2.39×10^{-5}
rs1290002	10	34506662	*PARD3*	0.992 [0.899, 1.094]	0.872	??++-?---	1.087 [1.035, 1.142]	8.77×10^{-4}
rs9568856	**13**	**52962982**	***OLFM4***	**1.224 [1.072, 1.397]**	**0.00275**	**+-++++-++**	**1.236 [1.155, 1.322]**	**9.26×10^{-10}**
rs9299	17	44024429	*HOXB5*	1.130 [1.028, 1.242]	0.0113	++-++++++	1.138 [1.084, 1.194]	1.94×10^{-7}
rs17697518	18	37019657	*KC6*	1.137 [1.007, 1.285]	0.0383	+-+++-+-+	1.193 [1.118, 1.274]	1.22×10^{-7}

1. Healthy Lifestyle in Europe by Nutrition in Adolescence study (HELENA)
2. Young Hearts studies
3. Lifestyle – Immune System – Allergy Study plus German Infant Study on the influence of Nutrition Intervention (LISA+GINI)
4. Children's Health Study (CHS)
5. INfancia y Medio Ambiente [Environment and Childhood] Project (INMA) (2 proxies used - rs965013 and rs11099020 for rs4833407)
6. Project Viva (VIVA)
7. Prevention and incidence of asthma and mite allergy birth cohort study (PIAMA)
8. Northern Finland 1986 Birth Cohort (NFBC1986) (2 proxies used - rs10779751 for rs230095; rs4883723 for rs9568856)
9. Avon Longitudinal Study of Parents and Children (ALSPAC)

Supplementary Table 9. Exploration of the eight loci in two extreme independent cohorts (705 trios; 1,509 cases and 2,674 controls) when restricting the cohorts to defining cases and controls over the age of 2 years old, sorted by chromosomal location. Separate replication data and combined data are shown along with an overall assessment of all data used in the study. Those signals achieving genome wide significance are indicated in red bold.

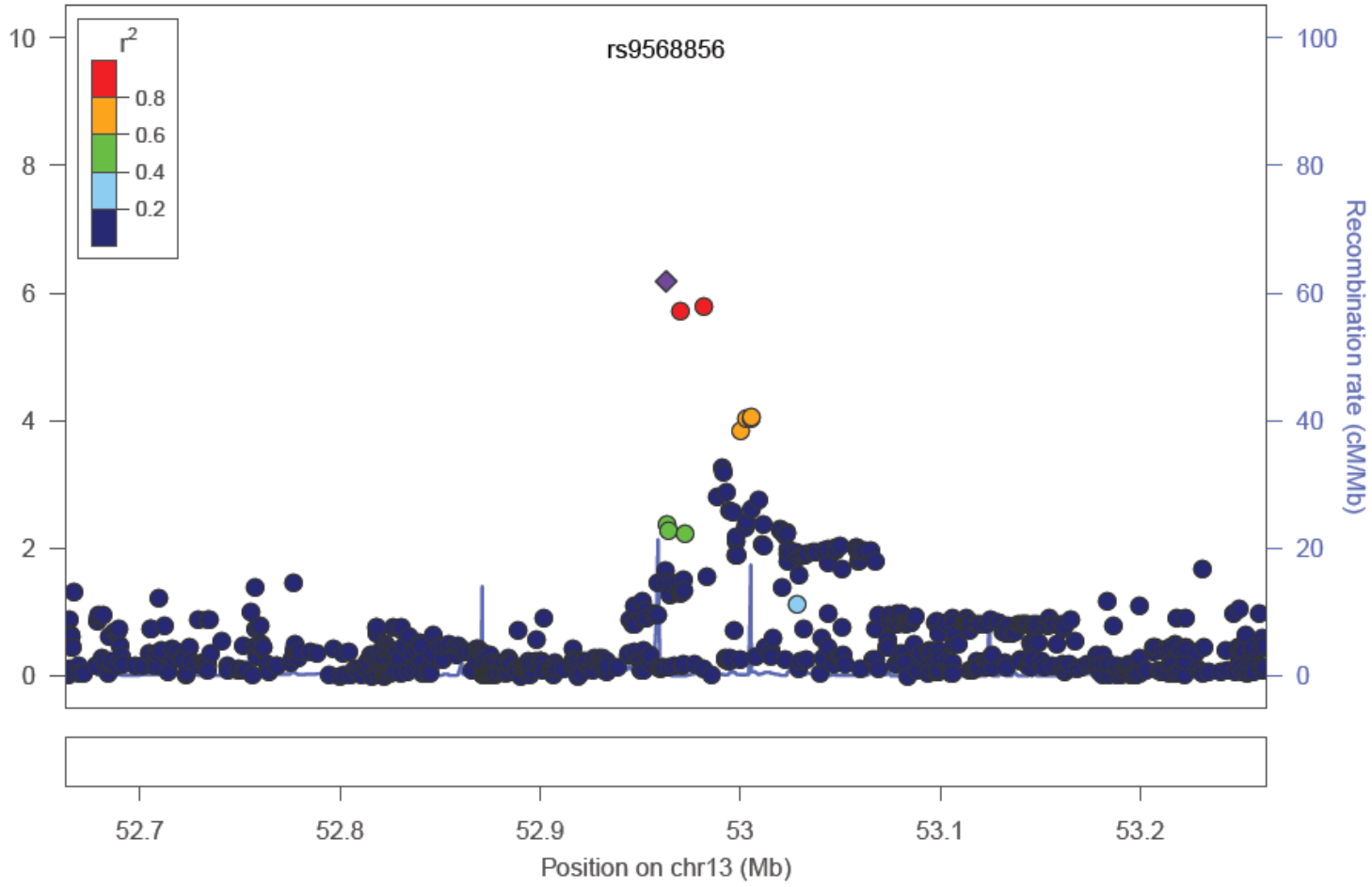
				Extreme obesity replication samples			Combined with Discovery		Meta-analysis of all discovery and replication cohorts	
SNP	**Chr**	**Pos**	**Nearest Gene**	**OR [95% C.I.]**	**P-value**	**Direction**	**OR [95% C.I.]**	**P-value**	**OR [95% C.I.]**	**P-value**
rs2300095	1	11188304	*MTOR-ANGPTL7*	1.014 [0.931, 1.105]	0.745	+-	1.091 [1.040, 1.146]	3.92×10^{-4}	1.083 [1.037, 1.131]	3.37×10^{-4}
rs4833407	**4**	**113531239**	***ALPK1***	**1.097 [1.014, 1.187]**	**0.0211**	**++**	**1.133 [1.084, 1.184]**	$\mathbf{3.15 \times 10^{-8}}$	**1.122 [1.079, 1.168]**	$\mathbf{1.00 \times 10^{-8}}$
rs4864201	**4**	**130950734**	***BC041448***	**1.055 [0.974, 1.143]**	**0.187**	**++**	**1.125 [1.075, 1.178]**	$\mathbf{4.46 \times 10^{-7}}$	**1.122 [1.077, 1.168]**	$\mathbf{3.50 \times 10^{-8}}$
rs28636	5	66184869	*MAST4*	0.990 [0.902, 1.086]	0.826	+-	0.902 [0.856, 0.950]	1.18×10^{-4}	0.914 [0.872, 0.958]	1.70×10^{-4}
rs1290002	10	34506662	*PARD3*	1.013 [0.931, 1.102]	0.764	-+	1.086 [1.036, 1.139]	5.84×10^{-4}	1.068 [1.023, 1.114]	0.00247
rs9568856	**13**	**52962982**	***OLFM4***	**1.072 [0.958, 1.200]**	**0.223**	**-+**	**1.182 [1.108, 1.260]**	$\mathbf{3.93 \times 10^{-7}}$	**1.190 [1.123, 1.261]**	$\mathbf{4.32 \times 10^{-9}}$
rs9299	17	44024429	*HOXB5*	1.003 [0.866, 1.161]	0.970	+?	1.121 [1.064, 1.182]	2.07×10^{-5}	1.123 [1.073, 1.176]	7.33×10^{-7}
rs17697518	18	37019657	*KC6*	0.986 [0.878, 1.107]	0.810	-+	1.140 [1.069, 1.216]	6.82×10^{-5}	1.140 [1.076, 1.207]	7.16×10^{-6}

1. Essen trios (4 proxies used - rs1010447 for rs230095; rs30731 for rs28636; rs613367 for rs1290002; rs4793944 for rs9299)
2. SCOOP-UK (1 proxy used - rs4845856 for rs230095)

Supplementary Table 10: Replication effort of the eight loci taken forward in to the adult GIANT[40] cohort. Seven of eight loci show a consistent direction of effect with observations in our childhood obesity analyses (only rs1290002 was inconsistent).

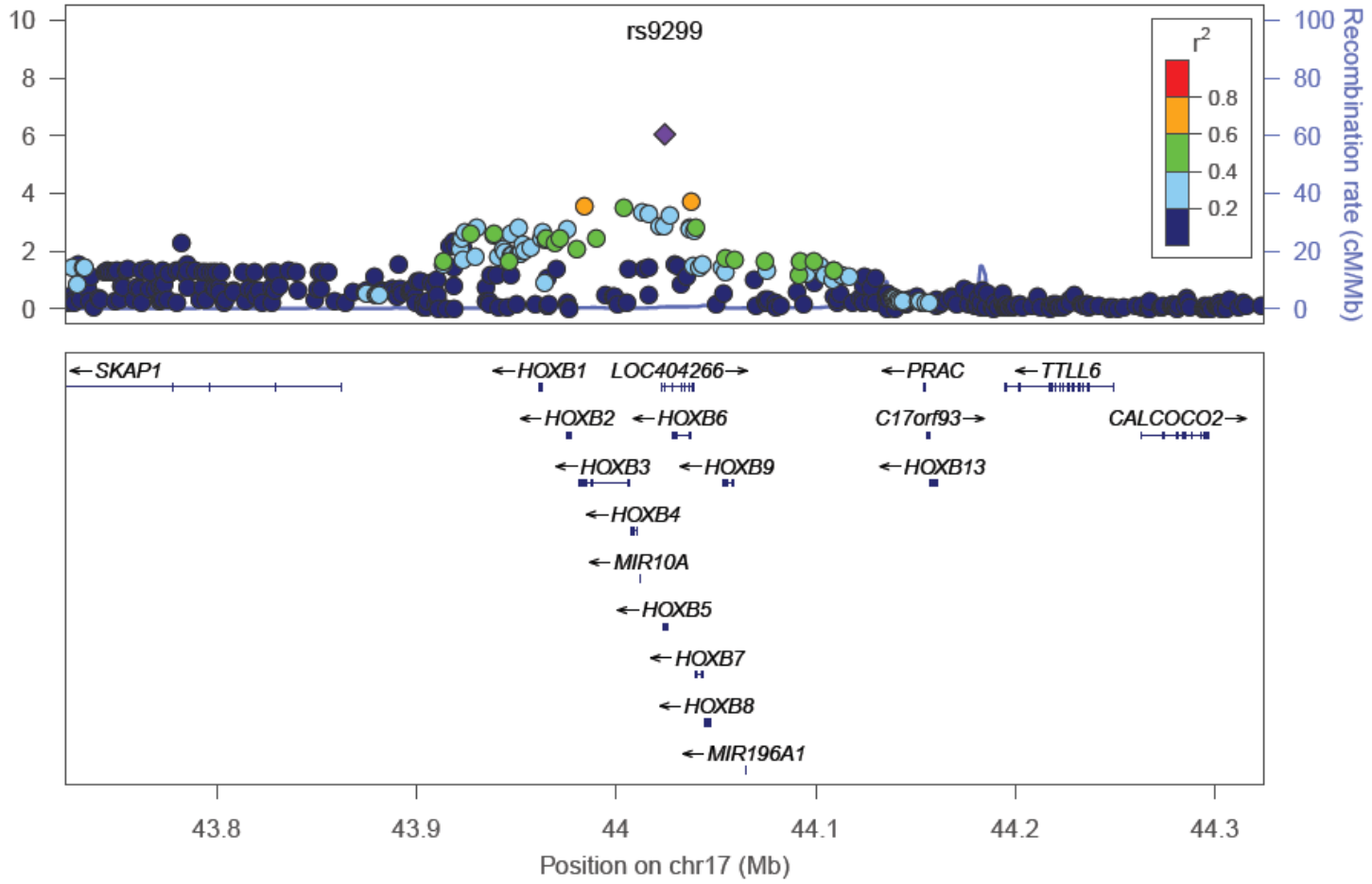
SNP	Chr	Pos	Trait-increasing allele	Nearest Gene	N	*P*-value
rs2300095	1	11188304	A	*MTOR-ANGPTL7*	123861	0.00255
rs4833407	4	113531239	A	*ALPK1*	123844	0.0132
rs4864201	4	130950734	T	*BC041448*	123862	9.66×10^{-4}
rs28636	5	66184869	C	*MAST4*	123860	0.677
rs1290002	10	34506662	G	*PARD3*	123864	0.818
rs9568856	13	52962982	A	*OLFM4*	122771	7.75×10^{-5}
rs9299	17	44024429	T	*HOXB5*	123844	0.0152
rs17697518	18	37019657	T	*KC6*	123846	0.664

Supplementary Figure 1. Regional plot of the *OLFM4* associated region. –log10(*P*-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.



Nature Genetics: doi:10.1038/2247

Supplementary Figure 2. Regional plot of the *HOXB5* associated region. –log10(*P*-values) are shown for all SNPs in the region and color of circles indicates degree of LD with the most associated SNP in the region. Recombination rate is overlaid on the figure and the position with respect to genes is shown at the bottom.



Nature Genetics: doi:10.1038/2247