LETTERS

A new highly penetrant form of obesity due to deletions on chromosome 16p11.2

A list of authors and their affiliations appears at the end of the paper

Obesity has become a major worldwide challenge to public health, owing to an interaction between the Western 'obesogenic' environment and a strong genetic contribution[1]. Recent extensive genome-wide association studies (GWASs) have identified numerous single nucleotide polymorphisms associated with obesity, but these loci together account for only a small fraction of the known heritable component[1]. Thus, the 'common disease, common variant' hypothesis is increasingly coming under challenge[2]. Here we report a highly penetrant form of obesity, initially observed in 31 subjects who were heterozygous for deletions of at least 593 kilobases at 16p11.2 and whose ascertainment included cognitive deficits. Nineteen similar deletions were identified from GWAS data in 16,053 individuals from eight European cohorts. These deletions were absent from healthy non-obese controls and accounted for 0.7% of our morbid obesity cases (body mass index (BMI) $\geq 40\,kg\,m^{-2}$ or BMI standard deviation score ≥ 4; $P = 6.4 \times 10^{-8}$, odds ratio 43.0), demonstrating the potential importance in common disease of rare variants with strong effects. This highlights a promising strategy for identifying missing heritability in obesity and other complex traits: cohorts with extreme phenotypes are likely to be enriched for rare variants, thereby improving power for their discovery. Subsequent analysis of the loci so identified may well reveal additional rare variants that further contribute to the missing heritability, as recently reported for *SIM1* (ref. 3). The most productive approach may therefore be to combine the 'power of the extreme'[4] in small, well-phenotyped cohorts, with targeted follow-up in case-control and population cohorts.

The extent to which copy-number variants (CNVs) might contribute to the missing heritability of common disorders is currently under debate[2]. Because most common simple CNVs are well tagged by single nucleotide polymorphisms (SNPs), it has recently been suggested that common CNVs are unlikely to contribute substantially to the missing heritability[5]. However, rare variants or recurring CNVs that have arisen on multiple independent occasions are unlikely to be captured by SNP tagging, and their identification will require alternative approaches.

We have previously proposed that cohorts with extreme phenotypes that include obesity may be enriched for rare but very potent risk variants[4,6]. Here we investigate 312 subjects, from three centres in the UK and France, presenting with congenital malformations and/or developmental delay in addition to obesity as defined previously[6,7] (see Methods). Known syndromes (for example, Prader–Willi and fragile X) were excluded. A combination of array comparative genomic hybridization (aCGH), genotyping arrays, quantitative PCR (qPCR) and multiplex ligation-dependent probe amplification (MLPA) was used to identify and confirm the presence of a heterozygous deletion on 16p11.2 in nine individuals (2.9%). These deletions, estimated to be a total of 740 kilobases (kb) in size (one copy of a segmental duplication plus 593 kb of unique sequences; Fig. 1a), have previously been associated to varying extents with autism, schizophrenia and developmental delay[8–11]; however, the observed

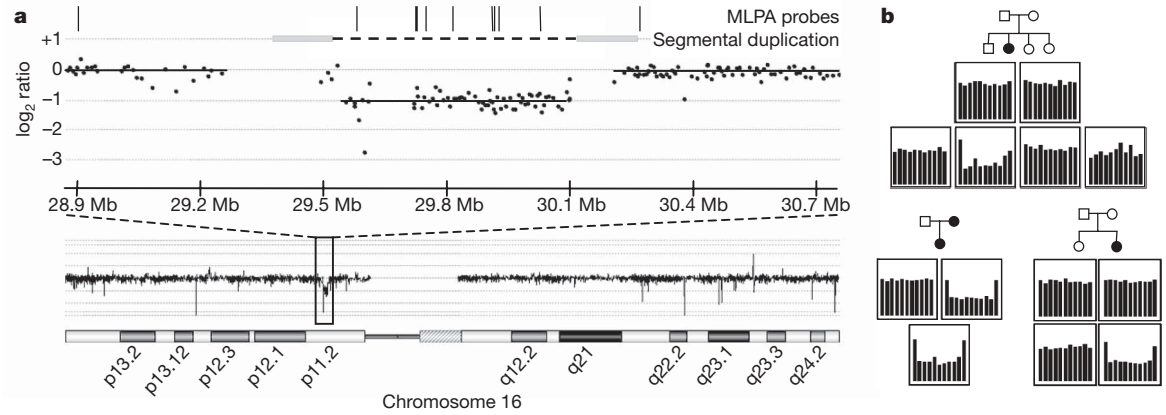

Figure 1 | Identification and validation of deletions at 16p11.2. a, aCGH data showing the location of the 16p11.2 deletion. The data show the $\log_2$ intensity ratio for a deletion carrier compared with that for an undeleted control sample. Grey bars connected by a broken line denote the segmental duplication flanking the deletion region. Vertical bars indicate the positions of the probe pairs used for MLPA validation. Note that CGH and genotyping array probes targeted against segmental duplications may not accurately report copy number as a result of the increased number of homologous sequences in the diploid state. Genome coordinates are in accordance with the hg18 build of the reference genome. **b**, MLPA validation of 16p11.2 deletions. Representative MLPA results are shown, illustrating one instance of maternal transmission and two instances of *de novo* deletions. Genotyping data excluded the possibility of non-paternity. Full results for MLPA validation and inheritance analysis are shown in Supplementary Fig. 1. Each panel shows the relative magnitude of the normalized, integrated signal at each probe location, in order of chromosomal position of the MLPA probe pairs as indicated in **a**. Each panel corresponds to its respective position on the associated pedigree, as shown.

frequency of deletions in our cohort is appreciably higher than the reported frequencies in the cohorts from the previous studies (less than 1%), which did not include obesity as an inclusion criterion.

A parallel, independent survey of aCGH and SNP-CGH data from eight cytogenetic centres in France, Switzerland and Estonia, involving 3,947 patients with developmental delay and/or malformations but this time without selection for obesity, revealed 22 unrelated cases with similar deletions (0.6%). This is a frequency consistent with those found in the previous studies[8–11], but is significantly lower than for the above cohort, which included only obese subjects ($P = 2.2 \times 10^{-4}$, Fisher's exact test).

Analysis of the available clinical data for these 22 new carriers indicated that, in addition to the ascertained cognitive deficits or behavioural abnormalities (including hyperphagia, specifically identified in at least nine cases; see Supplementary Table 1), a 16p11.2 deletion gave rise to a strongly expressed obesity phenotype in adults, with a more variable phenotype in childhood. All four teenagers and adults carrying a deletion were obese, whereas child carriers were also frequently either obese (4 of 15) or overweight (2 of 15), a tendency that has previously been noted[11]; the very young (under 2 years old) were of normal weight. This age-dependent penetrance was observed in all instances of deletions for which phenotypic data were available, whether from this study or from previously published reports[10–15], and regardless of ascertainment (Fig. 2; see Supplementary Tables 2 and 3).

Taken together, the data from these parallel studies suggest a possible direct association of deletions at 16p11.2 with obesity, distinct from their cognitive phenotype. Also identified in these cohorts were instances of the reciprocal duplication, which has also been implicated in neurodevelopmental disorders, but with a variable phenotype and lower penetrance[9,10,12]. The frequency of the duplication in the two cohorts (12 of 4,183 (0.3%)) was consistent with previous reports for patients with cognitive deficits (0.3–0.7%)[10,12]. Carriers of the duplication neither were obese nor had reported hyperphagia.

To investigate further the association of 16p11.2 deletions with obesity, and to estimate the extent to which it is observed independently of ascertainment for neurodevelopmental symptoms, we performed algorithmic and statistical analyses of genome-wide SNP genotyping data (see Table 1) from Swiss (CoLaus[16]), Finnish (NFBC1966 (ref. 17)) and Estonian (EGPUT[18]) general population cohorts (11,856 subjects in total), from child obesity and adult morbid obesity case-control cohorts[6,19,20] (1,224 and 1,548 subjects, respectively), from an extreme early-onset obesity cohort (SCOOP, 931 subjects) and from 141 patients undergoing bariatric weight-loss surgery (see Methods); in total, we identified 17 instances of deletions (and four duplications) with no significant gender bias (Table 1). In addition, we identified two further unrelated carriers of a deletion from 353 members of 149 families with sibling pairs discordant for obesity (SOS Sib Pair Study[21]). When DNA was available for further analysis (15 of 19 samples), the presence of a deletion was validated by using MLPA (Fig. 1b) or qPCR; the remaining deletions were validated by applying a second independent algorithm to the data. With the exception of a single individual who is apparently diabetic (fasting blood glucose more than 7 mM), all adult carriers of such deletions were obese, the majority being morbidly obese; similarly, each of the seven child or adolescent carriers had a BMI in the top 0.1% of the population range for their age and gender. None of the individuals ascertained on the basis of their obesity had any reported developmental delay or cognitive deficit; four subjects were reported as having hyperphagia.

To enable sufficient statistical power to give robust conclusions, we combined data from the population and obesity cohorts in an overall case-control association analysis (the samples from sib-pair families were excluded to avoid complications due to their relatedness). In comparison with lean or normal weight subjects (see Table 1 and Methods), 16p11.2 deletions were associated with obesity ($P = 5.8 \times 10^{-7}$, Fisher's exact test; odds ratio 29.8, 95% confidence limits 3.9 and 225) and morbid obesity ($P = 6.4 \times 10^{-8}$; odds ratio 43.0, 95% confidence limits 5.6 and 329) at or near genome-wide levels of significance. Expanding the control group to include all non-obese individuals increased the significance to $P = 4.2 \times 10^{-9}$ (obese) and $P = 6.1 \times 10^{-10}$ (morbidly obese).

Previous reports have indicated that these deletions are frequently not inherited from either parent but arise *de novo*, possibly by non-allelic homologous recombination between the more than 99% sequence-identical segmental duplications flanking the deleted region[11,14]. Therefore, where possible we investigated the parents of carriers of deletions, identifying 11 cases of maternal transmission and 4 of paternal transmission. The available data showed that all first-degree relatives carrying a deletion were also obese (Supplementary Table 1). In ten instances the deletion was apparently *de novo* (see Fig. 1b). Extrapolation to our full data set indicates that about 0.4% of all morbidly obese cases are due to an inherited 16p11.2 deletion. The frequency of *de novo* events is consistent with a previous report, in which ascertainment was for developmental delay and/or congenital anomalies[11]; by contrast, deletions are reported to be almost exclusively *de novo* in autistic subjects[8–10].

Although they may be heterogeneous in nature, these deletions are highly likely to be the causal variants, representing the second most frequent genetic cause of obesity after point mutations in *MC4R*[22,23]. Their repeated *de novo* occurrence is likely to result in a lack of linkage disequilibrium with any other flanking variant—no consistent haplotype has been identified by analysis of the available surrounding genotypes. To assess the effect of a deletion on the expression of nearby genes (for example, the obesity GWAS-associated *SH2B1* locus 800 kb distant[24]), we analysed available transcript data for subcutaneous adipose tissue samples from the discordant sibling cohort. Comparisons of the two subjects carrying a deletion with their corresponding non-obese siblings, and with other obese and non-obese subjects (Supplementary Fig. 4 and Supplementary Tables 4 and 5), showed that many, although not all, transcripts from within the deletion had a markedly decreased abundance (0.4–0.7-fold). In contrast, no clear evidence was found for consistent *cis* effects of the deletion on the abundance of messenger RNAs encoded by genes

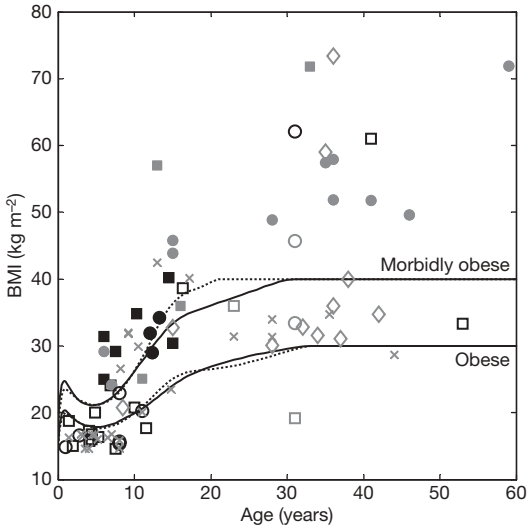

Figure 2 | Dependence of BMI on age in subjects having a deletion at 16p11.2. Data are shown for all individuals carrying a deletion for whom phenotypic data were available. Similar data from this study only are shown in Supplementary Figs 2 and 3. Lines denote the thresholds corrected for age and gender (solid, male; broken, female) for obesity and morbid obesity. Squares, male; circles, female; black, ascertained for developmental delay; grey, not ascertained for developmental delay; filled, ascertained for obesity; open, not ascertained for obesity; diamonds, first-degree relative of proband; crosses, previously published data[10–15]. The 31-year-old male with a BMI of about 20 kg m^{-2} was diabetic, as determined by a fasting blood glucose of more than 7 mM.

Table 1 | Frequency of detected 16p11.2 deletions in multiple cohorts

Cohort	Deletions/total					Technology
	Lean/normal	Overweight	Obese	Morbidly obese	Total	
Ascertained for cognitive deficits/malformations and obesity						
Lille/Strasbourg*					8/279	qPCR, aCGH
London*					1/33	aCGH, MLPA
Ascertained for cognitive deficits/malformations						
French–Swiss cytogenetic clinical diagnostic group*					21/3,870	aCGH, QMPSF, qPCR, FISH
Estonian cases of cognitive deficit*					1/77	Illumina CNV370-Duo, qPCR
Ascertained for obesity						
Swedish families with discordant siblings†§	0/140	0/54	0/115	2/44	2/353	Illumina 610K-Quad, MLPA
French adult case-control†	0/669	0/174	–	4/705	4/1,548	Illumina CNV370-Duo, MLPA
French child case-control‡	0/530	0/51	1/260	3/383	4/1,224	Illumina CNV370-Duo, MLPA
British extreme early-onset obesity (SCOOP)‡	–	–	–	3/931	3/931	Affymetrix 6.0, MLPA
French bariatric weight-loss surgery†	–	–	0/15	2/126	2/141	Illumina 1M-duo, MLPA
Population cohorts (origin)						
NFBC1966 (Finnish)†	1/3,148	0/1,622	1/434	1/42	3/5,246	Illumina CNV370-Duo
CoLaus (Swiss)†	0/2,675	0/2,049	0/830	0/58	0/5,612	Affymetrix 500K
EGPUT (Estonian)†	0/412	0/358	1/213	0/15	1/998	Illumina CNV370-Duo, qPCR
Total without ascertainment for cognitive deficits/malformations§	1/7,434	0/4,254	3/1,752	13/2,260		

For each cohort, 16p11.2 deletions were identified and validated with the indicated technologies. Where full phenotypic data were available, members of cohorts were categorized in accordance with the appropriate obesity criteria (see Supplementary Information): *not categorised, complete phenotypic data not available; †BMI thresholds for overweight, obese and morbidly obese were at least 25 kg m^{-2}, at least 30 kg m^{-2} and at least 40 kg m^{-2}, respectively; ‡BMI thresholds for overweight, obese and morbidly obese were the 90th centile, 97th centile and 4 standard deviations above the mean, respectively, corrected for age and gender. §Discordant siblings were not included in totals because of relatedness. QMPSF, quantitative multiplex PCR of short fluorescent fragments; FISH, fluorescence *in situ* hybridization.

flanking the deletion. In addition, global analysis of this data set has not identified any *trans*-acting expression quantitative trait loci either within or nearby the deletion.

Thus, although we cannot completely exclude the possibility that a 16p11.2 deletion affects the expression of nearby genes (for instance, its impact may be different in other tissues), the expression analysis described strongly indicates that the observed phenotypes are likely to be due to haploinsufficiency of one or more of the about 30 genes within the deleted region. Indeed, rather than being due to a single haploinsufficiency, the phenotype may well result from the deletion of multiple genes with an impact on pathways central to the development of obesity (see Supplementary Table 5). Functional network analysis of the deleted genes has led to the suggestion of a similar multigene effect for the cognitive phenotype[8]. The extent to which there is overlap between the genes involved in the obesity and cognitive phenotypes remains to be elucidated.

There is a strong correlation between developmental and cognitive disabilities and the prevalence of obesity: patients with autism or who have learning disabilities have a greatly increased risk of obesity[25], and the severely obese exhibit significant cognitive impairment[26]. Possible explanations include a direct causal relationship between obesity and developmental delay, the involvement of the same or related regulatory pathways, or different outcomes of the same set of behavioural disorders with complex pleiotropic effects and variable ages of onset and expressivities. The higher frequency of 16p11.2 deletions in the cohort ascertained for both phenotypes (2.9%), compared with cohorts ascertained for either phenotype alone (0.4% and 0.6%, respectively), confirms their impact on both obesity and developmental delay, adding to the evidence that these two phenotypes may be fundamentally interrelated.

METHODS SUMMARY

Obesity. Definitions for overweight, obesity and morbid obesity were based on previous studies[6,7]: for adults, BMI $\geq$ 25, 30 and 40 kg m^{-2} respectively; for children, BMI respectively above the 90th, 97th centiles and at least four standard deviations above the mean, calculated according to their age and gender from a French reference population[27,28].

Statistics. All reported statistical tests used Fisher's exact test[29], performed on contingency tables constructed for the number of subjects carrying or lacking a 16p11.2 deletion versus the obesity status or ascertainment of the individual. Because no homozygous deletions were observed, it was unnecessary to make a prior distinction between recessive, additive and dominant models of disease risk. Odds ratios and 95% confidence limits were calculated as described[30].

CNV discovery. Subjects ascertained for cognitive deficit/malformations with or without obesity were selected from those clinically referred for genetic testing; 16p11.2 deletions were identified in these individuals by standard clinical diagnostic procedures. Algorithmic analyses of GWAS data were performed variously using the cnvHap algorithm, a moving-window average-intensity procedure, a Gaussian mixture model, QuantiSNP, PennCNV, BeadStudio GT module, and Birdseed. When experimental validation was not possible, at least two independent algorithms were used for each data set.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 November; accepted 1 December 2009.

1. Walley, A. J., Asher, J. E. & Froguel, P. The genetic contribution to non-syndromic human obesity. *Nature Rev. Genet.* **10,** 431–442 (2009).
2. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).
3. Stutzmann, F. *et al.* Loss-of-function mutations in SIM1 cause a specific form of Prader–Willi-like syndrome. *Diabetologia* **52,** S104 (2009).
4. Froguel, P. & Blakemore, A. I. F. The power of the extreme in elucidating obesity. *N. Engl. J. Med.* **359,** 891–893 (2008).
5. Conrad, D. F. *et al.* Origins and functional effect of copy number variation in the human genome. *Nature* advance online publication doi:10.1038/nature08516 (7 October 2009).
6. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature Genet.* **41,** 157–159 (2009).
7. Farooqi, I. S. & O'Rahilly, S. Recent advances in the genetics of severe childhood obesity. *Arch. Dis. Child.* **83,** 31–34 (2000).
8. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17,** 628–638 (2008).
9. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
10. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
11. Bijlsma, E. K. *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur. J. Med. Genet.* **52,** 77–87 (2009).
12. McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41,** 1223–1227 (2009).
13. Ghebranious, N., Giampietro, P. F., Wesbrook, F. P. & Rezkalla, S. H. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am. J. Med. Genet. A.* **143A,** 1462–1471 (2007).
14. Fernandez, B. A. *et al.* Phenotypic spectrum associated with *de novo* and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *J. Med. Genet.*, doi:10.1136/jmg.2009.069369 (in the press).
15. Shimojima, K., Inoue, T., Fujii, Y. & Ohno, K. Yamamoto, T. A familial 593 kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae. *Eur. J. Med. Genet.* **52,** 433–435 (2009).

16. Firmann, M. *et al.* Prevalence of obesity and abdominal obesity in the Lausanne population. *BMC Cardiovasc. Disord.* **8,** 330, doi:10.1186/1471-2261-8-6 (2008).
17. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genet.* **41,** 35–46 (2009).
18. Nelis, M. *et al.* Genetic structure of Europeans: a view from the north-east. *PLoS One* **4,** e5472, doi:10.1371/journal.pone.0005472 (2009).
19. Balkau, B., Eschwege, E., Tichet, J. & Marre, M. Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). *Diabetes Metab.* **23,** 428–434 (1997).
20. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445,** 881–885 (2007).
21. Jernås, M. *et al.* Regulation of carboxylesterase 1 (CES1) in human adipose tissue. *Biochem. Biophys. Res. Commun.* **383,** 63–67 (2009).
22. Yeo, G. S. *et al.* A frameshift mutation in *MC4R* associated with dominantly inherited human obesity. *Nature Genet.* **20,** 111–112 (1998).
23. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human *MC4R* is associated with a dominant form of obesity. *Nature Genet.* **20,** 113–114 (1998).
24. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41,** 25–34 (2009).
25. Chen, A. Y., Kim, S. E., Houtrow, A. J. & Newacheck, P. W. Prevalence of obesity among children with chronic conditions. *Obesity (Silver Spring),* **18,** 210–213 (2010).
26. Boeka, A. G. & Lokken, K. L. Neuropsychological performance of a clinical sample of extremely obese individuals. *Arch. Clin. Neuropsychol.* **23,** 467–474 (2008).
27. Poskitt, E. M.; European Childhood Obesity Group. Defining childhood obesity: the relative body mass index (BMI). *Acta Paediatr.* **84,** 961–963 (1995).
28. Rolland-Cachera, M. F. *et al.* Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45,** 13–21 (1991).
29. Fisher, R. A. On the interpretation of χ^2 from contingency tables, and the calculation of *P. J. R. Stat. Soc. A* **85,** 87–94 (1922).
30. Woolf, B. On estimating the relation between blood group and disease. *Ann. Hum. Genet.* **19,** 251–253 (1955).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.J.W., A.I.F.B. and P.F. are supported by grants from the Wellcome Trust and the Medical Research Council (MRC). J.S.B. is supported by a grant from the Swiss National Foundation (310000-112552). L.J.M.C. is supported by an RCUK Fellowship. S.J. is funded by Swiss National Fund 320030_122674 and the Synapsis Foundation, University of Lausanne. A.V. is funded by the Ludwig Institute for Cancer Research. S.B. is supported by the Swiss Institute of Bioinformatics. I.S.F. and M.E.H. are funded by the Wellcome Trust and the MRC. We thank the DHOS (Direction de l'Hospitalisation et de l'Organisation des Soins) from the French Ministry of Health for their support in the development of several array CGH platforms in France. We thank 'le Conseil Regional Nord Pas de Calais/FEDER' for their financial support. Part of the CoLaus computation was performed at the Vital-IT center for high-performance computing of the Swiss Institute of Bioinformatics. The CoLaus authors thank Y. Barreau, M. Firmann, V. Mayor, A.-L. Bastian, B. Ramic, M. Moranville, M. Baumer, M. Sagette, J. Ecoffey and S. Mermoud for data collection. The CoLaus study was supported by grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne and by the Swiss National Foundation (33CSCO-122661). K.M., A.K., T.E., M.N. and A.M. received support from targeted financing from Estonian Government SF0180142s08, and P.P. from SF0180026s09; and from the EU through the European Regional Development Fund. T.E., M.N. and A.M. received support from FP7 grants (201413 ENGAGE, 212111 BBMRI, ECOGENE (no. 205419, EBC)). The genotyping of the Estonian Genome Project samples were performed in the Estonian Biocentre Genotyping Core Facility. The EGPUT authors thank V. Soo for technical help in genotyping. The Northwick Park authors acknowledge support from the NIHR Biomedical Research Centre Scheme and the Hammersmith Hospital Charity Trustees. Genome Canada and Genome Quebec funded the genotyping of DESIR subjects. Work on the SOS sib pair cohort was supported by grants from the Swedish Research Council (K2008-65X-20753-01-4, K2007-55X-11285-13, 529-2002-6671), the Swedish Foundation for Strategic Research to Sahlgrenska Center for Cardiovascular and Metabolic Research, the Swedish Diabetes Foundation, the Åke Wiberg Foundation, Foundations of the National Board of Health and Welfare, the Jeansson Foundations, the Magn Bergvall Foundation, the Tore Nilson Foundation, the Royal Physiographic Society (Nilsson-Ehle Foundation), VINNOVA-VINNMER, and the Swedish federal government under the LUA/ALF agreement. The DESIR study has been supported by INSERM, CNAMTS, Lilly, Novartis Pharma and Sanofi-Aventis, by INSERM (Réseaux en Santé Publique, Interactions entre les determinants de la santé), by the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon. Northern Finland Birth Cohort 1966 (NFBC1966) was supported by the Academy of Finland (project grants 104781, 120315 and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), the ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the MRC (studentship grant G0500539). The NFBC authors thank P. Rantakallio for the launch of NFBC1966 and initial data collection, S. Vaara for data collection, T. Ylitalo for administration, M. Koiranen for data management, and O. Tornwall and M. Jussila for DNA biobanking.

Author Contributions A.I.F.B., P.F., J.S.B. and L.J.M.C. designed and supervised the study. F.C., D.M., S.J., J.A. and S.B. coordinated and managed patient databases. R.G.W., A.V., A.J.d.S., C.L., F.S., F.C., J.-C.C., J.L.B., S.L., N.H. and J.S.E.-S.M. performed data analysis. A.J.d.S. conducted the MLPA analysis. J.A., M.F. and A.J.W. analysed expression data. A.-E.A., A.B., A.D., A.F., A.G., A.G., A.L., A.P., B.B., B.D., B.I., B.L., C.V.-D., C.L.C., D.C., D.M., D.S., F.F., G.M., G.P., J.-L.M., J.-M.C., J.A., J.C., K.M., K.D.M., K.O., M.M.v.H., M.-P.C., M.-P.L., M.P., M.B.-D., M.H.-E., M.M., N.C., O.B., P.J., R.C., R.E., R.F.K., R.T., S.D.-G., S.J., V.G. and V.M. supervised patient recruitment and performed cytogenetic analysis. A.-L.H., A.K., A.M., A.R., B.B., D.M., D.W., E.G.B., E.H., F.P., G.W., I.S.F., J.A., J.K., L.C., L.P., L.S., M.E.H., M.I.M., M.N., M.-R.J., N.H., P.E., P.J., P.P., P.V., R.S., S.B., S.O'R., T.E., V.M. and V.V. supervised cohort recruitment and/or conducted genotyping. R.G.W., S.J., A.V., A.J.d.S., L.J.M.C., A.I.F.B., P.F. and J.S.B. wrote and edited the manuscript and prepared figures. P.F. and J.S.B. contributed equally. All authors commented on and approved the manuscript.

Author Information The expression microarray data for carriers of 16p11.2 deletions is deposited in Gene Expression Omnibus under accession number GSE19238. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.F. (p.froguel@imperial.ac.uk) or J.S.B. (jacques.beckmann@chuv.ch).

R. G. Walters[1,2]*, S. Jacquemont[3]*, A. Valsesia[4–6], A. J. de Smith[1], D. Martinet[3], J. Andersson[1], M. Falchi[1], F. Chen[7], J. Andrieux[8], S. Lobbens[9], B. Delobel[10], F. Stutzmann[9], J. S. El-Sayed Moustafa[1], J.-C. Chèvre[9], C. Lecoeur[9], V. Vatin[9], S. Bouquillon[8], J. L. Buxton[1], O. Boute[11], M. Holder-Espinasse[11], J.-M. Cuisset[12], M.-P. Lemaitre[12], A.-E. Ambresin[13], A. Brioschi[14], M. Gaillard[3], V. Giusti[15], F. Fellmann[3], A. Ferrarini[3], N. Hadjikhani[7,16], D. Campion[17], A. Guilmatre[17], A. Goldenberg[18], N. Calmels[19], J.-L. Mandel[19], C. Le Caignec[20,21], A. David[20], B. Isidor[20], M.-P. Cordier[22], S. Dupuis-Girod[22], A. Labalme[22], D. Sanlaville[22,23], M. Béri-Dexheimer[24], P. Jonveaux[24], B. Leheup[24,25], K. Õunap[26], E. G. Bochukova[27], E. Henning[27], J. Keogh[27], R. J. Ellis[28], K. D. MacDermot[28], M. M. van Haelst[28]†, C. Vincent-Delorme[29], G. Plessis[30], R. Touraine[31], A. Philippe[32], V. Malan[32], M. Mathieu-Dramard[33], J. Chiesa[34], B. Blaumeiser[35], R. F. Kooy[35], R. Caiazzo[36,37], M. Pigeyre[37], B. Balkau[38], R. Sladek[39,40], S. Bergmann[4,6], V. Mooser[41], D. Waterworth[41], A. Reymond[42], P. Vollenweider[43], G. Waeber[43], A. Kurg[44], P. Palta[44], T. Esko[45,46], A. Metspalu[45,46], M. Nelis[45,46], P. Elliott[2], A.-L. Hartikainen[47], M. I. McCarthy[48,49], L. Peltonen[50–52], L. Carlsson[53], P. Jacobson[53], L. Sjöström[53], N. Huang[50], M. E. Hurles[50], S. O'Rahilly[27], I. S. Farooqi[27], K. Männik[44], M.-R. Jarvelin[2,54,55], F. Pattou[36,37], D. Meyre[9], A. J. Walley[1], L. J. M. Coin[2], A. I. F. Blakemore[1], P. Froguel[1,9] & J. S. Beckmann[3,4]

*These authors contributed equally to this work.

[1]Section of Genomic Medicine, Imperial College London, London W12 0NN, UK. [2]Department of Epidemiology and Public Health, Imperial College London, London W2 1PG, UK. [3]Service de Génétique Médicale, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. [4]Departement de Génétique Médicale, Université de Lausanne, CH-1015 Lausanne, Switzerland. [5]Ludwig Institute for Cancer Research, Université de Lausanne, CH-1015 Lausanne, Switzerland. [6]Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland. [7]Brain Mind Institute, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. [8]Laboratoire de Génétique Médicale, Centre Hospitalier Régional Universitaire, 59000 Lille, France. [9]CNRS 8090-Institute of Biology, Pasteur Institute, 59800 Lille, France. [10]Centre de Génétique Chromosomique, Hôpital Saint-Vincent de Paul, GHICL, 59020 Lille, France. [11]Service de Génétique Clinique, Hôpital Jeanne de Flandre, Centre Hospitalier Universitaire de Lille, 59000 Lille, France. [12]Service de Neuropédiatrie, Centre Hospitalier Régional Universitaire, 59000 Lille, France. [13]Unité Multidisciplinaire de Santé des Adolescents, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. [14]Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. [15]Service d'Endocrinologie, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. [16]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129, USA. [17]INSERM, U614, Faculté de Médecine, 76183 Rouen, France. [18]Service de Génétique, Centre Hospitalier Universitaire de Rouen, 76031 Rouen, France. [19]Laboratoire de Diagnostic Génétique, Nouvel Hôpital civil, 67091 Strasbourg, France. [20]Centre Hospitalier Universitaire Nantes, Service de Génétique Médicale, 44093 Nantes, France. [21]INSERM, UMR915, L'Institut du Thorax, 44007 Nantes, France. [22]Service de Génétique, Hospices Civils de Lyon, Hôpital de l'Hotel Dieu, 69288 Lyon, France. [23]EA 4171, Université Claude Bernard, 69622 Lyon, France. [24]Laboratoire de Génétique, Centre Hospitalier Universitaire, Nancy Université, 54511 Vandoeuvre les Nancy, France. [25]EA4368 Medical School Nancy, Université Henri Poincaré, 54003 Nancy, France. [26]Department of Genetics, United Laboratories, Tartu University Children's Hospital, 50406 Tartu, Estonia. [27]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's

Hospital, Cambridge CB2 0QQ, UK. [28]North West Thames Regional Genetics Service, Northwick Park & St Marks Hospital, Harrow HA1 3UJ, UK. [29]Centre Hospitalier D'Arras, Génétique Médicale, 62000 Arras, France. [30]Service de Génétique Médicale, Centre Hospitalier Universitaire Clemenceau, 14033 Caen, France. [31]Centre Hospitalier Universitaire–Hôpital Nord, Service de Génétique, 42055 Saint Étienne, France. [32]Département de Génétique et INSERM U781, Université Paris Descartes, Hôpital Necker-Enfants Malades, 75015 Paris, France. [33]Service de Génétique Clinique, Centre Hospitalier Universitaire, 80054 Amiens, France. [34]Laboratoire de Cytogénétique, Centre Hospitalier Universitaire Caremeau, 30029 Nîmes, France. [35]Department of Medical Genetics, University Hospital & University of Antwerp, 2650 Edegem, Belgium. [36]INSERM U859, Biotherapies for Diabetes, 59045 Lille, France. [37]Université Lille Nord de France, Centre Hospitalier Universitaire Lille, 59037 Lille, France. [38]INSERM U780-IFR69, 94807 Villejuif, France. [39]McGill University and Genome Quebec Innovation Centre, Montreal H3A 1A4, Canada. [40]Department of Medicine and Human Genetics, McGill University, Montreal H3A 1B1, Canada. [41]Division of Genetics, GlaxoSmithKline, Philadelphia, Pennsylvania 19101, USA. [42]The Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland. [43]Department of Medicine, Centre Hospitalier Universitaire Vaudois, CH-1011 Lausanne, Switzerland. [44]Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia. [45]Estonian Genome Project, University of Tartu, 50410 Tartu, Estonia. [46]Estonian Biocentre, 51010 Tartu, Estonia. [47]Department of Obstetrics and Gynaecology, University of Oulu, 90220 Oulu, Finland. [48]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK. [49]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [50]Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. [51]Institute of Molecular Medicine, Biomedicum, 00290 Helsinki, Finland. [52]Massachusetts Institute of Technology, The Broad Institute, Cambridge, Massachusetts 02142, USA. [53]Department of Molecular and Clinical Medicine and Center for Cardiovascular and Metabolic Research, The Sahlgrenska Academy, 413 45 Göteborg, Sweden. [54]Department of Child and Adolescent Health, National Public Health Institute, 90101 Oulu, Finland. [55]Institute of Health Sciences and Biocenter Oulu, University of Oulu, 90220 Oulu, Finland. †Present address: Department of Medical Genetics, University Medical Centre Utrecht, 3584 EA Utrecht, The Netherlands.

METHODS

Obesity phenotype. We used previously defined criteria to define overweight, obesity, and morbid (class III) obesity[6,7]. In adults, the thresholds were BMI $\geq$ 25, 30 and 40 kg m^{-2}, respectively. In children and adolescents, we used age-specific and sex-specific centiles of BMI, calculated from a French reference population[27,28], that approximately corresponded to these thresholds: overweight and obesity were defined by thresholds at the 90th and 97th centiles, respectively. Childhood morbid obesity was defined as BMI $\geq$ 4 standard deviations above the age-specific and sex-specific mean, which corresponds to a BMI of 40 kg m^{-2} between the ages of 20 and 30 years for both men and women; this threshold was used in the recruitment of the SCOOP severe early-onset obesity cohorts[7]. The age-specific and sex-specific thresholds used to define obesity and morbid obesity are shown in Fig. 1 and Supplementary Figs 1 and 2. No carriers of a 16p11.2 deletion were reported to be taking atypical antipsychotics (known to be associated with weight gain).

Patient and population cohorts. Patients referred for cognitive delay and obesity: a group of 33 patients was selected from those referred for genetic testing at the North West Thames Regional Genetics Service, based at Northwick Park Hospital in Harrow, UK, with approval from the Harrow Research Ethics Committee. Inclusion was based on three criteria: mental retardation, dysmorphology, and a weight greater than the 97th centile for age and gender. Abnormal karyotype, fragile X and Prader–Willi syndrome had previously been excluded.

A second group of 279 French children were selected from those referred to two centres (Laboratoire de Diagnostic Génétique, Nouvel Hôpital Civil, Strasbourg, France, and Centre de Génétique Chromosomique, Hôpital Saint-Vincent de Paul, GHICL, Lille, France). Inclusion was based on obesity plus at least one Prader–Willi-like syndromic feature (neonatal hypotonia and difficulty to thrive, mental retardation, developmental delay, behavioural problems, skin picking, facial dysmorphism, hypogenitalism or hypogonadism). Chromosomal abnormalities and Prader–Willi syndrome were excluded by karyotyping and DNA methylation analysis.

Patients referred for cognitive delay: patients with cognitive deficits are routinely referred to clinical genetics for aetiological work-ups including aCGH. We surveyed seven cytogenetic centres in France and Switzerland, identifying 3,870 patients ascertained for developmental delay and/or malformations. Also included in the study was a further 77 patients, ascertained on similar criteria, who were referred to the Department of Genetics, University of Tartu, Tartu, Estonia. These analyses were performed for clinical diagnostic purposes, all available phenotypic data (weight and height) being those provided anonymously by the clinician ordering the analysis. Consequently, research-based informed consent was not required by the institutional review board that approved the study.

CoLaus: this prospective population cohort was described previously[16]; 6,188 white individuals aged 35–75 years were randomly selected from the general population in Lausanne, Switzerland. These individuals underwent a detailed phenotypic assessment and were genotyped with the Affymetrix Mapping 500K array; 5,612 samples passed genotyping quality control. This study was approved by the institutional review boards of the University of Lausanne, and written consent was obtained from all participants. Because recruitment of this cohort required the ability to give informed consent, it is possible that the (statistically non-significant) lack of 16p11.2 deletions or duplications is due to an ascertainment bias. However, any such bias, if it exists, is very small and affects the identification of only one or two subjects carrying a deletion.

NFBC1966: the Northern Finland Birth Cohort 1966 is a prospective birth cohort of almost all individuals born in 1966 in the two northernmost provinces of Finland. Expectant mothers were enrolled, and clinical data collection took place prenatally, at birth, and at ages 6 months, 1 year, 14 years and 31 years. Biochemical and DNA samples were collected with informed consent at age 31 years. Genotyping with the Illumina Infinium 370cnvDuo array and phenotypic characteristics of the cohort were as described previously[17]. Phenotypic and genotyping data were available for 5,246 subjects after quality control.

EGPUT: the Estonian Genome Project is a biobank coordinated by the University of Tartu (EGPUT)[18]. The project is conducted in accordance with Estonian Gene Research Act, and all participants gave written informed consent. The cohort includes more than 39,000 individuals older than 18 years of age and reflects closely the age distribution in the Estonian population (33% male, 67% female; 83% Estonians, 14% Russians, 3% other). Subjects are recruited by general practitioners and hospital physicians and are then randomly selected. Computer Assisted Personal interview (CAPI) was filled during 1–2 h at the doctor's office. The data included personal data (such as place of birth, place(s) of living and nationality), family history (four generations), educational and occupational history, lifestyle and anthropometric data. A total of 1,090 randomly selected subjects were genotyped with the Illumina 370cnvDuo array, 998 passing the required criteria (nationality, genotyping call rate and phenotype availability).

Case-control familial obesity: the adult-obesity case-control groups and the child-obesity case control groups were as published previously[6], and were genotyped with the Illumina Human CNV370-duo array. In all, 643 children with familial obesity (BMI $\geq$ 97th centile corrected for gender and age, at least one obese first-degree relative, age less than 18 years), 581 non-obese children (BMI $\leq$ 90th centile), 705 morbidly obese adults with familial obesity (BMI $\geq$ 40 kg m^{-2}, at least one obese first-degree relative with BMI $\geq$ 35 kg m^{-2}, age $\geq$ 18 years) and 197 lean adults (BMI $\leq$ 25 kg m^{-2}) passed quality control; this cohort included a further 646 control subjects from the DESIR prospective cohort[19] (age at examination $\geq$ 45 years, normal fasting glucose in accordance with 1997 ADA criteria, BMI $<$ 27 kg m^{-2}) genotyped with the Illumina Hap300 array[20]. All participants or their legal guardians gave written informed consent, and all local ethics committees approved the study protocol.

Severe early-onset obesity cohort: the Genetics of Obesity Study (GOOS) cohort consists of more than 3,000 patients ascertained for severe obesity, defined as a BMI $\geq$ 4 standard deviations above the age-specific and sex-specific mean, and onset of obesity before 10 years of age. In this study we selected a discovery set of 1,000 UK Caucasian patients from this cohort in whom developmental delay had been excluded by routine clinical examination by experienced physicians (this cohort is referred to as SCOOP). Mutations in *LEPR*, *POMC* and *MC4R* were excluded by direct nucleotide sequencing and a karyotype was performed. DNA samples were analysed with Affymetrix GenomeWide Human SNP Array 6.0 by Aros, of which 931 passed quality control.

Bariatric surgery cohort: patients undergoing elective bariatric weight-loss surgery were recruited for the ABOS study at Lille Regional University Hospital. Genotyping was performed with the Illumina Human 1M-duo array, and data from 141 adults passed quality control. All participants gave written informed consent, and the study protocol was approved by the local ethics committee.

Swedish discordant sibling cohort: the SOS Sib Pair Study cohort was as published previously[21]. It includes 154 nuclear families, each with BMI discordant sibling pairs (BMI difference $>$ 10 kg m^{-2}), giving a total of 732 subjects. Genotyping data with the Illumina 610K-Quad array was available for 353 siblings from 149 families. Expression data from subcutaneous adipose tissue (sampled after overnight fasting) were available for 360 siblings from 151 families. Subjects received written and oral information before giving written informed consent. The Regional Ethics Committee in Gothenburg approved the studies.

Statistical methods. In view of the low frequency of the 16p11.2 deletions, all reported statistical tests were conducted with Fisher's exact test[29]. This was applied to comparisons of separately ascertained cohorts or categories and was performed on contingency tables constructed for the number of subjects carrying or lacking a 16p11.2 deletion (zero or one copies, because no homozygous deletions were observed) versus the obesity status or ascertainment of the individual. Because no homozygous deletions were observed, it was unnecessary to make a prior distinction between recessive, additive and dominant models of disease risk. For overall analysis of the obesity risk resulting from a deletion, cohorts were pooled in accordance with their obesity status determined according to the criteria described above, and the described tests were then applied to the pooled data. Odds ratios and 95% confidence limits were calculated as described[30].

CNV discovery and validation. Clinical identification of 16p11.2 deletions: all diagnostic procedures (aCGH, qPCR, QMPSF and FISH) were conducted in accordance with the relevant guidelines of good clinical laboratory practice for the respective countries. All rearrangements in probands were confirmed by a second technique, and karyotyping was performed in all cases to exclude a complex rearrangement.

cnvHap: CNVs were detected in the child/adult case-control, bariatric surgery, SOS sibpair and NFBC cohorts using the cnvHap algorithm (L.J.M.C., J. E. Asher, R.G.W., J.S.E.-S.M., A.J.d.S., R.S., D. J. Balding, P.F. and A.I.F.B., unpublished observations); this method is based on a hidden Markov model that models transitions between copy-number states at the haplotype level, improving sensitivity and accuracy by capturing linkage disequilibrium information between CNVs and SNPs. The compiled JAR and associated parameter files can be downloaded from http://www.imperial.ac.uk/medicine/people/l.coin/. Sample data from the algorithm applied to the NFBC cohort are illustrated in Supplementary Fig. 5a.

After clustering of genotyping data with the internal Illumina BeadStudio cluster files, values for log*R* ratio (LRR) and B-allele frequency (BAF) were exported from each project and normalized: effects of percentage GC content on LRR were removed by regressing on GC and GC^2, and wave effects[31] were removed by fitting a Loess function. Normalized data for probes within

2.5 megabases of the 16p11.2 deletion were analysed with cnvHap, and CNV calls intersecting the single-copy sequences within the deletion (chr16:29514353–30107356, build hg18) were extracted. 16p11.2 deletions were identified by a minimum 90% of probes within the deleted region being called as having a decreased copy number.

All called 16p11.2 deletions were validated by direct analysis of LRR. Data for each probe were normalized by first subtracting the median value across all samples (so that the distribution of LRR for each probes was centred on zero), and then dividing by the variance across all samples (to correct for variation in the sensitivity of different probes to copy-number variation). The normalized data were then smoothed by application of a nine-point moving average and visualized graphically (see Supplementary Fig. 6); putative deletions were checked by subsequent manual confirmation of loss of heterozygosity across the entire region. Equally, all deletions called by this method were confirmed by cnvHap.

Gaussian mixture model: for the CoLaus cohort, raw genotyping data were normalized using the aroma.affymetrix framework[32]. Normalization steps included allelic cross-talk calibration[33,34], intensity summarization using robust median average, and correction for any PCR amplification bias. Copy number (CN) ratios for a given sample, at a given SNP or CN probe, were computed as the $\log_2$ ratio of the normalized intensity of this probe divided by the median across all the samples. CN ratios were subsequently smoothed by fitting a Loess function[31]. CNV calling was performed with a new method based on a Gaussian mixture model (A.V., Z. Kutalik, T. Johnson, B. J. Stevenson, C. V. Jongeneel, D.W., V.M., P.V., G.W., J.S.B. and S.B., unpublished observations). This Gaussian mixture model fits four components (deletion, copy neutral, one additional copy and two additional copies) to CN ratios. The final copy number at each probe location is determined as the expected (dosage) copy number. The method has been validated by comparing test data sets with results from the CNAT[35] and CBS[36,37] algorithms and by replicating a subset of CoLaus subjects on Illumina arrays. All calls at the 16p11.2 locus made by the highly stringent CBS algorithm were replicated by the Gaussian mixture model. Principal components analysis detected no significant batch effects. Sample data from the algorithm applied to the CoLaus cohort are illustrated in Supplementary Fig. 5b.

PennCNV, QuantiSNP and Birdsuite: CNV discovery in the EGPUT cohort was performed with QuantiSNP[38], PennCNV[39] and BeadStudio GT module (Illumina). All analyses were conducted with the recommended settings, except changing EMiters to 25 and L to 1,000,000 in QuantiSNP. For PennCNV, the Estonian population-specific BAF file was used. Data from the SCOOP cohort were analysed with Affymetrix Power Tools and Birdsuite software[40].

Multiplex ligation-dependent probe amplification (MLPA): MLPA was performed with standard methods[41] using reagents obtained from MRC-Holland. The SALSA MLPA kit P343-B1 Autism-1 probe mix was used, which contained nine probes within the deleted region on 16p11.2, plus one probe upstream and one downstream of this locus (see Fig. 1a). MLPA products were separated with an AB3130 Genetic Analyser (Applied Biosystems) and outputs were analysed with GeneMarker software (Soft Genetics) and Microsoft Excel. Data normalization was performed by dividing the peak areas for each of the 11 test probes by the mean of 9 control probe peak areas. Normalized peak area data were then compared across the tested samples to determine which of them carried the 16p11.2 deletion.

31. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8,** R228, doi:10.1186/gb-2007-8-10-r228 (2007).
32. Bengtsson, H., Simpson, K., Bullard, J. & Hansen, K. aroma.affymetrix: a generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. (Department of Statistics, University of California, Berkeley, Technical Report 745, 2008).
33. Bengtsson, H., Irizarry, R., Carvalho, B. & Speed, T. P. Estimation and assessment of raw copy numbers at the single locus level. *Bioinformatics* **24,** 759–767 (2008).
34. Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics* **25,** 861–867 (2009).
35. Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1,** 287–299 (2004).
36. Olshen, A. B. & Venkatraman, E. S. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).
37. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (2007).
38. Collela, S. *et al.* QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **35,** 2013–2025 (2007).
39. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17,** 1665–1674 (2007).
40. Korn, J. N. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genet.* **40,** 1253–1260 (2008).
41. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57 (2002).