LETTER

doi:10.1038/nature10406

Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus

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

Both obesity and being underweight have been associated with increased mortality[1,2]. Underweight, defined as a body mass index (BMI) $\leq 18.5\,kg\,per\,m^2$ in adults and ≤ -2 standard deviations from the mean in children, is the main sign of a series of heterogeneous clinical conditions including failure to thrive[3–5], feeding and eating disorder and/or anorexia nervosa[6,7]. In contrast to obesity, few genetic variants underlying these clinical conditions have been reported[8,9]. We previously showed that hemizygosity of a ~600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities[10]. Here we show that the corresponding reciprocal duplication is associated with being underweight. We identified 138 duplication carriers (including 132 novel cases and 108 unrelated carriers) from individuals clinically referred for developmental or intellectual disabilities (DD/ID) or psychiatric disorders, or recruited from population-based cohorts. These carriers show significantly reduced postnatal weight and BMI. Half of the boys younger than five years are underweight with a probable diagnosis of failure to thrive, whereas adult duplication carriers have an 8.3-fold increased risk of being clinically underweight. We observe a trend towards increased severity in males, as well as a depletion of male carriers among non-medically ascertained cases. These features are associated with an unusually high frequency of selective and restrictive eating behaviours and a significant reduction in head circumference. Each of the observed phenotypes is the converse of one reported in carriers of deletions at this locus. The phenotypes correlate with changes in transcript levels for genes mapping within the duplication but not in flanking regions. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance.

Copy-number variants (CNVs) at the 16p11.2 locus have been associated with cognitive disorders including autism (deletions) and schizophrenia (duplications)[11–13], conditions that have been suggested to lie at opposite ends of a single spectrum of psychiatric phenotypes[14]. We and others have reported that a deletion of this region spanning 28 genes (Supplementary Table 1) increases the risk of morbid obesity 43-fold (Supplementary Fig. 1)[10,15]. We hypothesized that the reciprocal duplication, with its resulting increase in gene dosage, may influence BMI in a converse manner. The duplication was identified in 73 out of 31,424 patients with DD/ID, a frequency consistent with previous reports[13] (Table 1). Four additional cases were identified among 1,080 patients affected by bipolar disease or schizophrenia. Compared to its prevalence in seven European population-based genome-wide association study (GWAS) cohorts[16–18] (31 out of 58,635 individuals), the duplication was significantly more frequent in both the DD/ID cohorts ($P = 4.23 \times 10^{-13}$; odds ratio = 4.4, 95% confidence interval = 2.9–6.9) and the psychiatric cohorts ($P = 3.6 \times 10^{-3}$; odds ratio = 7.0, 95% confidence interval = 1.8–19.9) (Table 1), strengthening previous reports of similar associations[12,13]. Our data do not support a two-hit model[19] for the effects of 16p11.2 duplications or deletions (Supplementary Text and Supplementary Table 2).

We compared available data on weight, height and BMI for 106 independent duplication carriers (including published cases) to data for reference populations matched for gender, age and geographical location (Table 2, Methods and Supplementary Tables 3 and 4). The duplication was strongly associated with lower weight (mean *Z*-score −0.56; $P = 4.4 \times 10^{-4}$) and lower BMI (mean *Z*-score −0.47; $P = 2.0 \times 10^{-3}$) (Table 2 and Supplementary Table 5). Birth parameters ($n = 48$) were normal, indicating a postnatal effect. Adults carrying the duplication had a relative risk of being clinically underweight (BMI <18.5) of 8.3 (95% confidence interval = 4.4–15.9, $P = 1.53 \times 10^{-10}$) (see Methods). Concordantly, none of the 3,544 patients in our obesity cohorts[10,15] carried the duplication (Table 1).

To investigate these associations further, we carried out separate analyses of carrier patients (DD/ID and psychiatric) and non-medically ascertained carriers (population-based cohorts, plus 11 transmitting parents and three other affected first-degree relatives for whom data were available) (Table 2). Each category had significantly lower weight and BMI, with similar effect sizes. However, the proportion of underweight cases (BMI ≤ -2 s.d.) was higher in the first group than in the second group (17 out of 76 compared to 2 out of 40; $P = 0.017$). Note that the impact of the duplication on underweight status might be underestimated here owing to prescription of antipsychotic treatments that are often associated with weight gain[20] (Supplementary Table 6).

Having demonstrated an association of the duplication with being underweight, we investigated the implications of gender for the resulting phenotypes (Fig. 1, Supplementary Fig. 2 and Supplementary Table 7). In DD/ID patients, the impact of the duplication on being underweight is stronger in males; the effect in females is in the same direction, but is smaller and not statistically significant (Table 2). A similar and significant difference ($P = 0.0168$) was observed in adult carriers (all groups combined): the relative risk of being underweight for males is 23.2 (95% confidence interval = 9.1–59.3, $P = 4.6 \times 10^{-11}$); for females it is only 4.7 (95% confidence interval = 1.9–11.8, $P = 9.9 \times 10^{-4}$). A gender bias was also observed in the ascertainment of DD/ID duplication carriers, in which we have an excess of males (51 males:33 females, $P = 0.044$). By contrast, carriers from the general population showed a strong overrepresentation of females (10 males:21 females, $P = 0.035$) (Supplementary Text). A similar bias was observed among transmitting parents (7 males:23 females, $P = 5.53 \times 10^{-4}$). Thus, there is an overrepresentation of males in the medically ascertained group, and a depletion in the non-medically ascertained one. We suggest that males may be more likely than females to present severe phenotypes, and that this may account for the gender bias because severely affected males may be less likely to be recruited to adult population cohorts or to be reproductively successful.

As previously reported[21], the duplication was also associated with reduced head circumference (mean *Z*-score −0.89; $P = 7.8 \times 10^{-6}$) (Fig. 1), 26.7% presenting with microcephaly (head circumference ≤ -2 s.d.), whereas carriers of the reciprocal deletion had an increased head circumference (mean *Z*-score +0.57; $P = 1.79 \times 10^{-5}$) (Supplementary Fig. 3 and Supplementary Table 8): an additional instance of a mirror phenotype associated with reciprocal changes in copy number at this locus. Notably, head circumference *Z*-scores correlate

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | 16p11.2 rearrangements in cases and controls

Ascertainment	Cohorts	Duplication		Deletion		Total
		n	*P*†	*n*	*P*†	
Neuro-developmental disorders	Unspecified DD/ID* from 28 cytogenetic centres	72		113		30,323
	ADHD‡, deCODE	0		1		591
	Childhood autism‡, deCODE	0		2		159
	Childhood autism spectrum disorder‡, deCODE	1		3		351
	TOTAL	**73**	**4.23×10^{-13}**	**119**	**5.43×10^{-32}**	**31,424**
	Rearrangement frequency (95% CI)	**0.23%** (0.18–0.29)		**0.38%** (0.31–0.45)		
Family history	First-degree relatives of probands	30		35		43/62‖
Adult psychiatric symptoms	Schizophrenia, deCODE	0		1		657
	Bipolar disease, Rouen	1		0		156
	Schizophrenia, schizo-affective, Rouen	3		0		267
	TOTAL	**4**	**3.57×10^{-3}**	**1**	**3.78×10^{-1}**	**1,080**
	Rearrangement frequency (95% CI)	**0.37%** (0.01–0.73)		**0.09%** (0–0.27)		
Underweight	Eating disorder, Spain	1§		0		441
Obesity	Obesity, Spain	0		2		653
	Adult obesity, France	0		4		705
	Childhood obesity, France & UK	0		7		1,574
	Obesity bariatric surgery, France	0		2		141
	Obesity discordant siblings, Sweden	0		2		159
	Obesity and cognitive delay, France & UK	0		9		312
	TOTAL	**0**	**4.21×10^{-1}**	**26**	**2.52×10^{-19}**	**3,544**
	Rearrangement frequency (95% CI)	0		**0.73%** (0.45–1.01)		
Population-based cohorts	NFBC1966 Finnish	4		3		5,319
	CoLaus Swiss	5		0		5,612
	EGCUT Estonian	2		1		2,994
	deCODE Iceland	17		18		36,601
	SHIP Germany	1		2		4,070
	KORA F3+F4 Germany	2		1		3,458
	Paediatric family study	0		0		581
	TOTAL	**31**		**25**		**58,635**
	Rearrangement frequency (95% CI)	**0.05%** (0.03–0.07)		**0.04%** (0.03–0.06)		

CI, Confidence interval; ADHD, attention-deficit hyperactivity disorder. *Not a disease-specific cohort. Detailed distribution is provided in the online methods. †Fisher's exact test, compared to the combined frequency in general population groups. ‡There was no overlap between these 3 cohorts. §Atypical duplication (see Supplementary Fig. 5). ‖ Total number of parental pairs tested for duplication/deletion. 13 out of 43 duplications and 27 out of 62 deletion cases were *de novo*.

positively with those of BMI in carriers of both the duplication (rho = 0.37; $P = 2.65 \times 10^{-3}$) and the deletion (rho = 0.42; $P = 1.9 \times 10^{-5}$) (Supplementary Methods). This indicates that head circumference and BMI may be regulated by a common pathway, or that a causal relationship exists between these two traits in these patients. Alternatively, the two phenotypes may arise from distinct genes and pathways. A full list of malformations and secondary phenotypes reported in duplication carriers ascertained for DD/ID is available in Supplementary Table 9.

In view of the importance of modified eating behaviours in obesity and being underweight, the clinical reports of duplication carriers were screened for evidence of such modified behaviours. In 11 out of 77 clinically ascertained cases, clinicians had spontaneously reported low food intake and selective and restrictive eating behaviour, again mirroring one of the phenotypes—hyperphagia—seen in deletion carriers[10] (Supplementary Table 6) and indicating that the duplication may increase the risk of eating disorders. Consequently, we carried out multiplex ligation-dependent probe amplification (MLPA, Supplementary Table 10) to screen for 16p11.2 rearrangements in 441 patients diagnosed with eating disorders, including anorexia nervosa, bulimia and binge eating disorder (Table 1 and Supplementary Text). No duplications of the entire region were identified, but one out of 109 anorexia nervosa patients carried an atypical 136-kb duplication that encompasses the sialophorin (*SPN*) and quinolinate phosphoribosyltransferase (*QPRT*) genes (Supplementary Fig. 4). This single, smaller duplication does not allow us to draw any firm conclusions, but together with other atypical rearrangements, it may, in the future, be essential for establishing the roles of the 28 genes within the region.

Large genomic structural variants are known to affect the expression of genes not only within the affected region but also at a distance[22–25]. Therefore, it is possible that the phenotypes observed in 16p11.2 deletion and duplication individuals are due to effects on the expression of genes mapping outside the rearranged interval, rather than to gene dosage within the 600-kb deletion or duplication. We measured

Table 2 | Comparisons of the height, weight and BMI distributions in duplication carriers and controls.

		Combined†			DD/ID or psychiatric†			Non-medically ascertained‡		
	Strata	Mean *Z*-score	*P*-value	*n**	Mean *Z*-score	*P*-value	*n**	Mean *Z*-score	*P*-value	*n**
BMI	All	−0.47	**2.0×10^{-3}**	102	−0.56	**4.1×10^{-3}**	76	−0.45	**6.0×10^{-3}**	40
	Male	−0.54	**2.1×10^{-2}**	52	−0.71	**1.3×10^{-2}**	43	−0.31	2.0×10^{-1}	14
	Female	−0.4	**1.8×10^{-2}**	50	−0.37	8.3×10^{-2}	33	−0.52	**4.2×10^{-3}**	26
Weight	All	−0.56	**4.4×10^{-4}**	104	−0.65	**1.3×10^{-3}**	78	−0.61	**3.0×10^{-3}**	40
	Male	−0.64	**5.8×10^{-3}**	53	−0.79	**4.4×10^{-3}**	44	−0.57	8.8×10^{-2}	14
	Female	−0.47	**1.7×10^{-2}**	51	−0.47	6.5×10^{-2}	34	−0.63	**8.6×10^{-3}**	26
Height	All	−0.24	4.8×10^{-2}	103	−0.33	3.6×10^{-2}	77	−0.15	1.8×10^{-1}	40
	Male	−0.34	4.5×10^{-2}	52	−0.4	4.6×10^{-2}	43	−0.29	1.2×10^{-1}	14
	Female	−0.14	2.6×10^{-1}	51	−0.24	2.1×10^{-1}	34	−0.07	3.7×10^{-1}	26

The available BMI, weight and height data for duplication carriers were transformed to *Z*-scores using gender- and age-matched reference populations, and one-tailed *t*-tests were carried out to determine whether the mean *Z*-scores deviated from zero. Significant differences were identified by reference to cutoffs controlling the false discovery rate at 5% (see Methods): BMI, 0.022; weight, 0.032; height, 0.025. Significant results are indicated in bold. Data were not available for all subjects. *Relatives of probands were excluded as required, to avoid including more than one member of the same family in a single analysis. †Including 24 cases from the literature (Supplementary Table 3). ‡Population-based cases and first-degree relatives of probands.

©2011 Macmillan Publishers Limited. All rights reserved

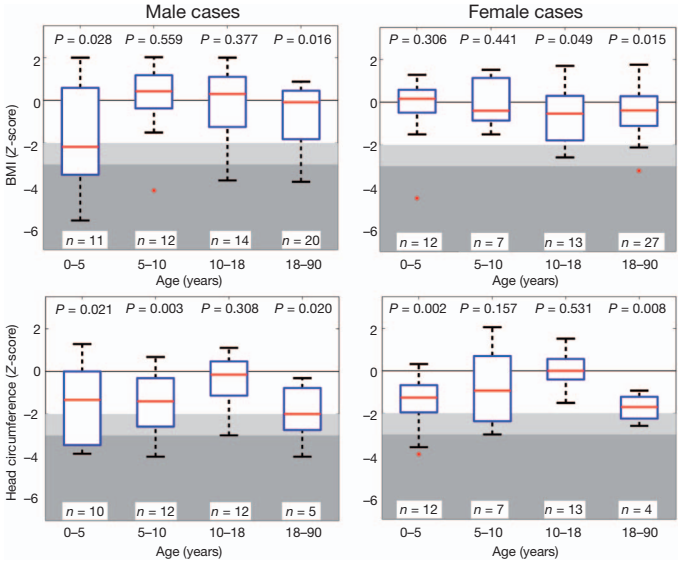

Figure 1 | Effect of the chromosome 16p11.2 duplication on BMI and head circumference. *Z*-score values of BMI and head circumference in carriers of the 16p11.2 duplication, stratified by gender and age group. The most severe effect is observed in children at 0–5 years of age. Boxplots represent the fifth, twenty-fifth, median, seventy-fifth and ninety-fifth percentile for each age group. Light grey and dark grey backgrounds represent ≤ -2 and ≤ -3 s.d., respectively, corresponding to the WHO definition of moderately and severely underweight[30]. BMI is decreased in adolescent and adult females.

relative transcript levels of 27 genes mapping within or near to the rearrangement, using lymphoblastoid cell lines (Supplementary Tables 1 and 11): six from deletion carriers, five from duplication carriers and ten from gender- and age-matched controls (Supplementary Table 12). Expression levels correlated positively with gene dosage for all genes in the CNV region (Fig. 2), consistent with published partial results from adipose tissue[10]. Mean relative transcript levels in deletion and duplication carriers were, respectively, 67% and 214% of the levels measured in controls (Supplementary Table 13). Although genes proximal (centromeric) to the rearrangement interval showed no significant variation in relative transcript levels between patients and controls (Fig. 2), distal (telomeric) genes showed a marked alteration in relative expression. However, their expression levels, including that of *SH2B1* (for which gene dosage and a nearby single nucleotide polymorphism (SNP) have been associated with obesity[15,26]), were similarly upregulated in cell lines of both deletion and duplication carriers, showing no apparent correlation between transcript level and either copy number or phenotype (Fig. 2). Although lymphoblastoid cells may not recapitulate obesity-relevant tissues, previous experiments have shown a high degree of correlation between expression levels in different tissues and cell lines[22], indicating that the same pathways may be similarly disrupted in different cell lineages. Thus, any involvement of these distal genes in the control of BMI in these subjects seems unlikely.

Our study demonstrates the power of very large screens (>95,000 samples: to our knowledge the largest of its kind so far) to characterize the clinical and molecular correlates of a rare functional genomic variant. We demonstrate unambiguously that carrying the 16p11.2 duplication confers a high risk of being clinically underweight, and show that reciprocal changes in gene dosage at this locus result in several mirror phenotypes. As in the schizophrenia/autism[14] and microcephaly/macrocephaly[21] dualisms, abnormal eating behaviours, such as hyperphagia and anorexia, could represent opposite pathological manifestations of a common energy-balance mechanism, although the precise relationships between these mirror phenotypes remain to be determined. We speculate that head circumference (which correlates with brain volume[27]), and thus neuronal circuitry, may affect cognitive function and energy balance in patients with 16p11.2 rearrangements (possibly through eating behaviour). Consistent with this are previous reports that a subgroup of children with microcephaly show a concomitant reduction in weight percentile[28]. Our findings also support the observation that severe overweight and underweight phenotypes correlate with lower cognitive functioning[4,29]. Thus, abnormal food intake may be a direct result of particular neurodevelopmental disorders. Although it is possible that the 16p11.2 region encodes distinct genes specific for each trait, a more parsimonious hypothesis is that these clinical manifestations of dysfunction of the central nervous system are all secondary to the disruption of a single neurodevelopmental step that is sensitive to gene dosage. Further resolution of this issue may

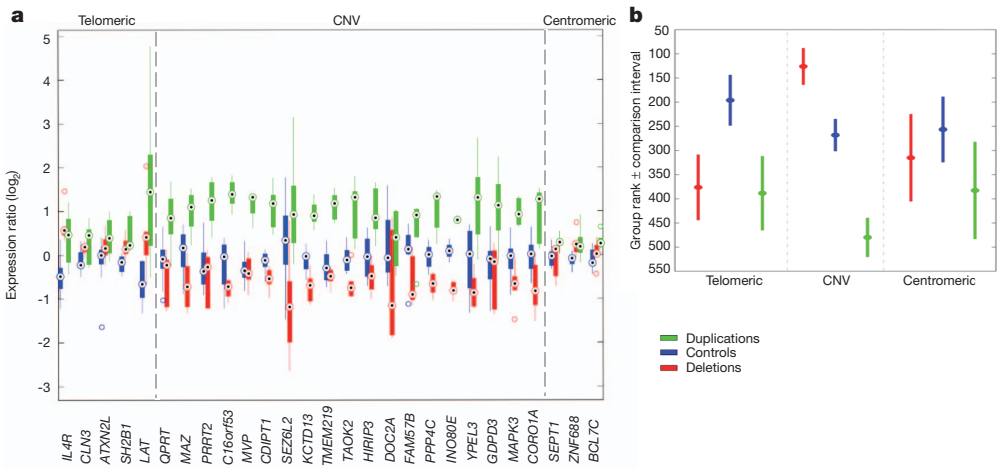

Figure 2 | Transcript levels for genes within and near to the 16p11.2 rearrangements. **a**, Relative expression levels of 27 genes mapping to 16p11.2 in deletion ($n = 6$) and duplication ($n = 5$) carriers (red and green, respectively), and in control cell lines ($n = 10$, blue). Grey lines denote the extent of the 16p11.2 CNV (29.5–30.1 megabases (Mb)). Complete lists of genes mapping within the rearranged interval, and of the quantitative PCR assays, are in Supplementary Tables 1 and 11, respectively. For the possible relevance of each of these genes to obesity/leanness and/or developmental delay/cognitive deficits, see ref. 10. **b**, Rank comparison (Kruskal–Wallis test) between the expression of 27 genes mapping to 16p11.2 in deletion and duplication carriers (red and green, respectively) and in control cell lines (blue). Genes are labelled as telomeric, centromeric or within the rearranged interval (CNV). Dots correspond to the mean group rank and bars indicate the comparison interval. Groups with non-overlapping intervals are significantly different (*P*-values were adjusted for multiple testing issues using a Bonferroni correction, where the number of tests is the number of pairwise comparisons; the resulting adjusted *P*-value was less than 0.05).

©2011 Macmillan Publishers Limited. All rights reserved

require the identification of additional patients with rare atypical rearrangements in this region.

METHODS SUMMARY

Underweight is defined in adults as BMI ≤ 18.5. In individuals younger than 18 years of age, it is defined as a Z-score ≤ -2.

Statistics. Two-tailed Fisher's exact test was used to compare frequencies of the rearrangement in patients and controls. Z-scores were computed for all data using gender-, age- and geographically-matched reference populations. One-tailed Student's t-test was performed to test BMI, height, weight and head circumference in duplication carriers for Z-scores of less than zero. We used Kruskal–Wallis tests for differences in gene expression patterns. P-values were adjusted using a Bonferroni correction, considering the number of pairwise comparisons; the resulting adjusted P-value was less than 0.05. The relative risk of being underweight was calculated as the ratio of the fraction of underweight individuals among duplication carriers versus our control group.

Discovery of CNVs. Carriers of 16p11.2 duplication and deletion were identified through various procedures: (1) comparative genomic hybridization with Agilent 44K, 60K, 105K, 180K, 244K arrays; (2) Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips; (3) Affymetrix 6.0, 500K genotyping arrays; (4) quantitative multiplex PCR of short fluorescent fragments (QMPSF); (5) fluorescent *in situ* hybridization (FISH); (6) MLPA. CNV analyses of GWAS data were carried out using cnvHap, a moving-window average-intensity procedure, a Gaussian mixture model, circular binary segmentation, QuantiSNP, PennCNV, BeadStudio GT module and Birdseed. At least two independent algorithms were used for each cohort.

Expression analyses. Lymphoblastoid cell lines were established from carriers and controls. SYBR Green quantitative PCR was performed to assess relative expression of genes.

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

Received 9 February; accepted 29 July 2011.
Published online 31 August 2011.

1. Berrington de Gonzalez, A. *et al.* Body-mass index and mortality among 1.46 million white adults. *N. Engl. J. Med.* **363,** 2211–2219 (2010).
2. Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. Cause-specific excess deaths associated with underweight, overweight, and obesity. *J. Am. Med. Assoc.* **298,** 2028–2037 (2007).
3. Olsen, E. M. *et al.* Failure to thrive: the prevalence and concurrence of anthropometric criteria in a general infant population. *Arch. Dis. Child.* **92,** 109–114 (2007).
4. Corbett, S. S. & Drewett, R. F. To what extent is failure to thrive in infancy associated with poorer cognitive development? A review and meta-analysis. *J. Child Psychol. Psychiatry* **45,** 641–654 (2004).
5. Rudolf, M. C. & Logan, S. What is the long term outcome for children who fail to thrive? A systematic review. *Arch. Dis. Child.* **90,** 925–931 (2005).
6. Bravender, T. *et al.* Classification of eating disturbance in children and adolescents: proposed changes for the DSM-V. *Eur. Eat. Disord. Rev.* **18,** 79–89 (2010).
7. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR* 39–134 and 583–596 (Am. Psychiatric Assoc., 2000).
8. Scherag, S., Hebebrand, J. & Hinney, A. Eating disorders: the current status of molecular genetic research. *Eur. Child Adolesc. Psychiatry* **19,** 211–226 (2010).
9. Wang, K. *et al.* A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol. Psychiatry* doi:10.1038/mp.2010.107 (2010).
10. Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463,** 671–675 (2010).
11. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
12. McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41,** 1223–1227 (2009).
13. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
14. Crespi, B., Stead, P. & Elliot, M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc. Natl Acad. Sci. USA* **107** (Suppl 1), 1736–1741 (2010).
15. Bochukova, E. G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463,** 666–670 (2010).
16. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **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 (2009).
19. Girirajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum. Mol. Genet.* **19,** R176–R187 (2010).
20. Pramyothin, P. & Khaodhiar, L. Metabolic syndrome with the atypical antipsychotics. *Curr. Opin. Endocrinol. Diabetes Obes.* **17,** 460–466 (2010).
21. Shinawi, M. *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **5,** 332–341 (2009).
22. Merla, G. *et al.* Submicroscopic deletion in patients with Williams–Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am. J. Hum. Genet.* **79,** 332–341 (2006).
23. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848–853 (2007).
24. Henrichsen, C. N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nature Genet.* **41,** 424–429 (2009).
25. Ricard, G. *et al.* Phenotypic consequences of copy number variation: insights from Smith–Magenis and Potocki–Lupski syndrome mouse models. *PLoS Biol.* **8,** e1000543 (2010).
26. 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).
27. Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. *J. Am. Med. Assoc.* **290,** 337–344 (2003).
28. Baxter, P. S., Rigby, A. S., Rotsaert, M. H. & Wright, I. Acquired microcephaly: causes, patterns, motor and IQ effects, and associated growth changes. *Pediatrics* **124,** 590–595 (2009).
29. Li, Y., Dai, Q., Jackson, J. C. & Zhang, J. Overweight is associated with decreased cognitive functioning among school-age children and adolescents. *Obesity (Silver Spring)* **16,** 1809–1815 (2008).
30. de Onis, M., Blossner, M., Borghi, E., Frongillo, E. A. & Morris, R. Estimates of global prevalence of childhood underweight in 1990 and 2015. *J. Am. Med. Assoc.* **291,** 2600–2606 (2004).

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

Acknowledgements We thank the Vital-IT high-performance computing centre of the Swiss Institute of Bioinformatics. S.J. is recipient of a bourse de relève académique de la Faculté de Biologie et Médecine de l'Université de Lausanne. This work was supported by the Leenaards Foundation Prize (S.J., D.M. and A.Reymond), the Jérôme Lejeune Foundation (A.Reymond), the Telethon Action Suisse Foundation (A.Reymond), the Swiss National Science Foundation (A.Reymond, J.S.B., S.B. and S.E.A.), a Swiss National Science Foundation Sinergia grant (S.J., D.M., S.B., J.S.B. and A.Reymond), the European Commission anEUploidy Integrated Project grant 037627 (A.Reymond, S.B., X.E., H.G.B. and S.E.A.), the Ludwig Institute for Cancer Research (A.V.), the Swiss Institute of Bioinformatics (S.B. and Z.K.), an Imperial College Department of Medicine PhD studentship (J.S.E.-S.M), the Comprehensive Biomedical Research Centre, Imperial College Healthcare NHS Trust, and the National Institute for Health Research (P.E.), the Wellcome Trust and the Medical Research Council (A.I.F.B. and P.F.), the Instituto de Salud Carlos III (ISCIII)-FIS, the German Mental Retardation Network funded through a grant of the German Federal Ministry of Education and Research (NGFNplus 01GS08160) (A.Reis), European Union-FEDER (PI081714, PS09/01778) (F.F.A., M.G. and X.E.), SAF2008-02278 (C.R.), the Belgian National Fund for Scientific Research, Flanders (N.V.A. and R.F.K.), the Dutch Organisation for Health Research and Development (ZON-MW grant 917-86-319) and Hersenstichting Nederland (B.B.A.d.V.), grant 81000346 from the Chinese National Natural Science Foundation (Y.G.Y.), the Simons Foundation Autism Research Initiative, Autism Speaks and NIH grant GM061354 (J.F.G.), and Oesterreichische Nationalbank (OENB) grant 13059 (A.K.-B.). Y.S. holds a Young Investigator Award from the Children's Tumor Foundation and a Catalyst Award from Harvard Medical School. B.L.W. holds a Fudan Scholar Research Award from Fudan University, a grant from Chinese National '973' project on Population and Health (2010CB529601) and a grant from the Science and Technology Council of Shanghai (09JC1402400). E.R.S. and S.L., recipients of the Michael Smith Foundation for Health Research Scholar award, acknowledge the CIHR MOP 74502 operational grant. The Estonian Genome Center of the University of Tartu (EGCUT) received support from the EU Centre of Excellence in Genomics and FP7 grants 201413 and 245536, and from Estonian Government SF0180142s08, SF0180026s09 and SF0180027s10 (A.M., K.M. and A.K.). D.S. thanks the Direction Générale de l'Organisation des Soins from the French Ministry of Health for their support in the development of several array-CGH platforms, and the Centres Labellisés Anomalies du Development in France. The Helmholtz Zentrum Munich and the State of Bavaria financed the KORA study, also supported by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823), the German Federal Ministry of Education and Research (BMBF), and the Munich Center of Health Sciences (MC Health, LMUinnovativ). CIBEROBN and CIBERESP are initiatives of ISCIII (Spain). S.W.S. holds the GlaxoSmithKline-Canadian Institutes of Health Chair in Genetics and Genomics at the University of Toronto and the Hospital for Sick Children, and is supported by Genome Canada and the McLaughlin Centre. Funding for deCODE came in part from NIH grant MH071425 (K.S.), EU grant HEALTH-2007-2.2.1-10-223423 (Project PsychCNV) and EU grant IMI-JU-NewMeds. NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), 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), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank M. Hass, Z. Jaros, M. Jussila, M. Koiranen,

©2011 Macmillan Publishers Limited. All rights reserved

P. Rantakallio, M. C. Rudolf, V. Soo, O. Tornwall, S. Vaara, T. Ylitalo and the French DHOS national CGH network for their help, as well as all participating patients and clinicians. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions S.J., A.Reymond, P.F. and J.S.B. wrote the manuscript with contributions from F.Z., L.H., R.G.W., N.D.B., Z.K., A.I.F.B. and A.V. L.H., A.V. and A.Reymond produced and analyzed the expression data. Z.K., A.V., R.G.W. and N.D.B. conducted the statistical analyses, guided by S.J., A.Reymond, P.F. and J.S.B. S.J., A.Reymond, F.Z., L.H., D.M., Y.S., G.T., M.B., S.B., D.C., N.d.L., B.B.A.d.V., B.A.F., F.F.A., M.G., A.G., J.H., A.K., C.L.C., K.M., O.S.P. D.S., M.M.V.H., S.V.G., A.T.V.-v.S., F.W., B.-L.W., Y.Y., J.A., X.E., J.F.G., A.M., S.W.S., K.S., U.T., A.I.F.B., J.S.B., P.F. and all other authors phenotyped and/or genotyped patients and/or individuals of the general population. S.J., A.Reymond and J.S.B. designed the study. All authors commented on and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. J.S.B. and P.F. are equally contributing senior authors and F.Z., L.H., R.G.W. and Z.K. are equally contributing second authors. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.B. (Jacques.Beckmann@chuv.ch).

Sébastien Jacquemont[1]*, Alexandre Reymond[2]*, Flore Zufferey[1], Louise Harewood[2], Robin G. Walters[3], Zoltán Kutalik[4,5], Danielle Martinet[1], Yiping Shen[6,7], Armand Valsesia[4,5,8], Noam D. Beckmann[1], Gudmar Thorleifsson[9], Marco Belfiore[1], Sonia Bouquillon[10], Dominique Campion[11], Nicole de Leeuw[12], Bert B. A. de Vries[12], Tõnu Esko[13,14], Bridget A. Fernandez[15], Fernando Fernández-Aranda[16], José Manuel Fernández-Real[17], Mònica Gratacòs[18], Audrey Guilmatre[11], Juliane Hoyer[19], Marjo-Riitta Jarvelin[20,21,22], R. Frank Kooy[23], Ants Kurg[13], Cédric Le Caignec[24], Katrin Männik[13], Orah S. Platt[6], Damien Sanlaville[25], Mieke M. Van Haelst[3,26], Sergi Villatoro Gomez[18], Faida Walha[2], Bai-lin Wu[6,27], Yongguo Yu[6,28], Azzedine Aboura[29], Marie-Claude Addor[1], Yves Alembik[30], Stylianos E. Antonarakis[31], Benoît Arveiler[32,33], Magalie Barth[34], Nathalie Bednarek[35], Frédérique Béna[31], Sven Bergmann[4,5], Mylène Beri[36], Laura Bernardini[37], Bettina Blaumeiser[23], Dominique Bonneau[34], Armand Bottani[31], Odile Boute[38], Han G. Brunner[12], Dorothée Cailley[33], Patrick Callier[39], Jean Chiesa[40], Jacqueline Chrast[2], Lachlan Coin[3], Charles Coutton[41], Jean-Marie Cuisset[42], Jean-Christophe Cuvellier[42], Albert David[24], Bénédicte de Freminville[43], Bruno Delobel[44], Marie-Ange Delrue[32,33], Bénédicte Demeer[45], Dominique Descamps[46], Gérard Didelot[2], Klaus Dieterich[47], Vittoria Disciglio[48,110], Martine Doco-Fenzy[49], Séverine Drunat[29], Bénédicte Duban-Bedu[44], Christèle Dubourg[50,51], Julia S. El-Sayed Moustafa[3], Paul Elliott[52,53], Brigitte H. W. Faas[12], Laurence Faivre[54], Anne Faudet[55], Florence Fellmann[1], Alessandra Ferrarini[1], Richard Fisher[56], Elisabeth Flori[30], Lukas Forer[57], Dominique Gaillard[49], Marion Gerard[29], Christian Gieger[58], Stefania Gimelli[31], Giorgio Gimelli[59], Hans J. Grabe[60], Agnès Guichet[34], Olivier Guillin[11], Anna-Liisa Hartikainen[61], Délphine Heron[62,63], Loyse Hippolyte[1], Muriel Holder[38], Georg Homuth[64], Bertrand Isidor[24], Sylvie Jaillard[50,65], Zdenek Jaros[66], Susana Jiménez-Murcia[16], Géraldine Joly Helas[67], Philippe Jonveaux[36], Satu Kaksonen[68], Boris Keren[55], Anita Kloss-Brandstätter[57], Nine V. A. M. Knoers[26,69], David A. Koolen[12], Peter M. Kroisel[70], Florian Kronenberg[57], Audrey Labalme[25], Emilie Landais[49], Elisabetta Lapi[71], Valérie Layet[72], Solenn Legallic[11], Bruno Leheup[73], Barbara Leube[74], Suzanne Lewis[75], Josette Lucas[65], Kay D. MacDermot[76], Pall Magnusson[77], Christian Marshall[78], Michèle Mathieu-Dramard[45], Mark I. McCarthy[79, 80,81], Thomas Meitinger[82,111], Maria Antonietta Mencarelli[48], Giuseppe Merla[83], Alexandre Moerman[38], Vincent Mooser[84], Fanny Morice-Picard[32,33], Mafalda Mucciolo[48], Matthias Nauck[85], Ndeye Coumba Ndiaye[86], Ann Nordgren[87], Laurent Pasquier[88], Florence Petit[38], Rolph Pfundt[12], Ghislaine Plessis[89], Evica Rajcan-Separovic[90], Gian Paolo Ramelli[91], Anita Rauch[92], Roberto Ravazzolo[93], Andre Reis[19], Alessandra Renieri[48], Cristobal Richart[94], Janina S. Ried[58], Claudine Rieubland[95], Wendy Roberts[96], Katharina M. Roetzer[70], Caroline Rooryck[32,33], Massimiliano Rossi[25], Evald Saemundsen[97], Véronique Satre[41], Claudia Schurmann[64], Engilbert Sigurdsson[98], Dimitri J. Stavropoulos[99], Hreinn Stefansson[9], Carola Tengström[100], Unnur Thorsteinsdóttir[9,101], Francisco J. Tinahones[102], Renaud Touraine[43], Louis Vallée[42], Ellen van Binsbergen[26], Nathalie Van der Aa[23], Catherine Vincent-Delorme[103], Sophie Visvikis-Siest[86], Peter Vollenweider[104], Henry Völzke[105], Anneke T. Vulto-van Silfhout[12], Gérard Waeber[104], Carina Wallgren-Pettersson[106], Robert M. Witwicki[2], Simon Zwolinksi[56], Joris Andrieux[10], Xavier Estivill[18], James F. Gusella[7], Omar Gustafsson[9,107], Andres Metspalu[13,14], Stephen W. Scherer[108], Kari Stefansson[9], Alexandra I. F. Blakemore[3], Jacques S. Beckmann[1,4] & Philippe Froguel[3,109]

[1]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. [2]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. [3]Department of Genomics of Common Disease, Imperial College London, London W12 0NN, UK. [4]Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland. [5]Swiss Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. [6]Laboratory Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [7]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [8]Ludwig Institute for Cancer Research, University of Lausanne, 1015 Lausanne, Switzerland. [9]deCODE Genetics, Sturlugata 8, IS-101 Reykjavík, Iceland. [10]Laboratoire de Génétique Médicale, Hopital Jeanne de Flandre, CHRU de Lille, 59000 Lille, France. [11]INSERM U614, University of Rouen and Centre Hospitalier du Rouvray, 76000 Rouen, France. [12]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [13]Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia. [14]Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia. [15]Discipline of Genetics and Medicine, Memorial University of Newfoundland, St John's A1B 3V6, Newfoundland, Canada. [16]Department of Psychiatry, University Hospital of Bellvitge-IDIBELL, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), 08907 Barcelona, Spain. [17]Department of Diabetes, Endocrinology, and Nutrition, Hospital Universitari de Girona Dr. Josep Trueta, Institut d'Investigació Biomèdica de Girona, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 17007 Girona, Spain. [18]Genes and Disease Program, Center for Genomic Regulation (CRG-UPF), CIBER en Epidemiología y Salud Pública (CIBERESP), 08003 Barcelona, Catalonia, Spain. [19]Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany. [20]Department of Epidemiology and Biostatistics, Imperial College London, School of Public Health, London W2 1PG, UK. [21]Department of Child and Adolescent Health, National Institute for Health and Welfare, Box 310, 90101 Oulu, Finland. [22]Institute of Health Sciences, University of Oulu, and Biocenter Oulu, University of Oulu, Box 5000, 90014 University of Oulu, Finland. [23]Department of Medical Genetics, University and University Hospital Antwerp, B-2650 Antwerp, Belgium. [24]Service de Génétique Médicale, CHU Nantes, 44093 Nantes, France. [25]Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CHU de Lyon and Neuroscience Research Center, TIGER team, UCBL1, Lyon, F-69000, France. [26]Department of Medical Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. [27]Children's Hospital and Institutes of Biomedical Science, Fudan University, 200032 Shanghai, China. [28]Shanghai Children's Medical Center, 200127 Shanghai, China. [29]Department of Genetics, APHP-Robert DEBRE University Hospital, 75019 Paris, France. [30]Service de Génétique Médicale, CHU Strasbourg, Hôpital de Hautepierre, 67000 Strasbourg, France. [31]Service of Genetic Medicine, University Hospitals of Geneva, 1205 Geneva, Switzerland. [32]Laboratoire Maladies Rares-Génétique et Métabolisme, Université Bordeaux 2, 33076 Bordeaux, France. [33]Service de Génétique Médicale, CHU de Bordeaux, 33076 Bordeaux, France. [34]Service de Génétique, CHU Angers, 49933 Angers, France. [35]Service Pédiatrie, CHU Hôpital Alix de Champagne, 51100 Reims, France. [36]Laboratoire de génétique, CHU Nancy, 54511 Vandoeuvre les Nancy, France. [37]Mendel Laboratory, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni Rotondo, Italy. [38]Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, 59037 Lille, France. [39]Laboratoire de Cytogénétique, CHU le Bocage, 21070 Dijon, France. [40]Laboratoire de Cytogénétique, CHU Caremeau, 30029 Nimes, France. [41]Laboratoire de Génétique Chromosomique, CHU de Grenoble, BP 217, 38043 Grenoble, France. [42]Service de Neurologie pédiatrique, Hôpital Roger Salengro, CHRU de Lille, 59037 Lille, France. [43]CHU-Hôpital Nord, Service de Génétique, CHU Saint Etienne, 42055 St Etienne, France. [44]Centre de Génétique Chromosomique, Hôpital Saint-Vincent de Paul, GHICL, 59160 Lille, France. [45]Service de Génétique Médicale, CHRU Amiens, 80000 Amiens, France. [46]Centre hospitalier de Béthune, 62408 Bethune, France. [47]Service de Génétique Clinique, CHU Grenoble, 38043 Grenoble, France. [48]Medical Genetics, Department of Biotechnology, University of Siena, 53100 Siena, Italy. [49]Service de Génétique, HMB, CHU REIMS, IFR 53, 51092 Reims, France. [50]UMR 6061 CNRS, IFR 140 GFAS, Université de Rennes 1, 35043 Rennes, France. [51]Service de Génétique Moléculaire, CHU, 35033 Rennes, France. [52]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [53]MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [54]Centre de Génétique, CHU Dijon, 21079 Dijon, France. [55]Département de Génétique et Cytogénétique, APHP-GH Pitié-Salpêtrière, 75013 Paris, France. [56]Institute of Human Genetics, International Centre for Life, Newcastle Upon Tyne NE1 4EP, UK. [57]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria. [58]Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany. [59]Laboratorio di Citogenetica, G. Gaslini Institute, 16147 Genova, Italy. [60]Department of Psychiatry and Psychotherapy, Ernst-Moritz-Arndt University Greifswald, 17475 Greifswald and D-18437 Stralsund, Germany. [61]Institute of Clinical Medicine, Department of Obstetrics and Gynecology, University of Oulu, 90570 Oulu, Finland. [62]Embryologie AP-HP, Université Pierre et Marie Curie, 75005 Paris, France. [63]Département de Génétique Cytogénétique, Hôpital Pitié-Salpêtrière, 75013 Paris, France. [64]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt University Greifswald, D-17487 Greifswald, Germany. [65]Laboratoire de Cytogénétique et Biologie Cellulaire, CHU Rennes, 35033 Rennes, France. [66]Abteilung für Kinder und Jugendheilkunde, Landesklinikum Waldviertel Zwettl, 3910 Zwettl, Austria. [67]Department of Genetics, CHU Rouen, 76183 Rouen, France. [68]The Habilitation Unit of Folkhalsan, Folkhalsan, SF 00250 Helsinki, Finland. [69]Department of Human Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. [70]Institute of Human Genetics, Medical University of Graz, 8010 Graz, Austria. [71]Medical Genetics Unit, Children's Hospital Anna Meyer, 50139 Firenze, Italy. [72]Unité de Génétique, Groupe Hospitalier du Havre, 76600 le Havre, France. [73]Service de Médecine Infantile III et Génétique Clinique, CHU-Nancy et PRES de l'Université de Lorraine UHP Nancy, 54511 Vandoeuvre les Nancy, France. [74]Institute of Human Genetics and Anthropology, Heinrich-Heine University Hospital Duesseldorf, D-40001 Duesseldorf, Germany. [75]Department of Medical Genetics, The University of British Columbia and Child and Family Research Institute, Vancouver V6H 3N1, Canada. [76]North West Thames Regional Genetics Service, Northwick Park & St Marks Hospital, Harrow HA1 3UJ, UK. [77]Child and Adolescent Psychiatry, Landspitali University Hospital, IS-105 Reykjavík, Iceland. [78]The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [79]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. [80]Wellcome Trust Centre for Human Genetics, University of Oxford, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [81]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. [82]Institute of Human Genetics,Hemholtz Center Munich, German Research Center for Environmental Health and Institute of Human Genetics, Technical University Munich, 85764 Neuherberg, Germany. [83]Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni

©2011 Macmillan Publishers Limited. All rights reserved

Rotondo, Italy. [84]Genetics, GlaxoSmithKline R&D, 720 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. [85]Institute of Clinical Chemistry and Laboratory Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. [86]Cardiovascular Genetics Research Unit, EA4373, Université Henri Poincaré, 54000 Nancy, France. [87]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Stockholm, Sweden. [88]Service de Génétique-CLAD Ouest, 35203 Rennes cedex 2, France. [89]Service de Génétique, CHU Clémenceau, 14033 Caen, France. [90]Department of Pathology, University of British Columbia and Child and Family Research Institute, Vancouver, British Columbia V5Z 4H4, Canada. [91]Division of Pediatrics, Ospedale San Giovanni, 6500 Bellinzona, Switzerland. [92]Institute of Medical Genetics, University of Zurich, 8603 Schwerzenbach, Switzerland. [93]Department of Pediatrics and CEBR, G. Gaslini Institute, University of Genova, 16126 Genova, Italy. [94]Department of Internal Medicine, University Hospital Juan XXIII, Universitat Rovira y Virgili, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 43005 Tarragona, Spain. [95]Division of Human Genetics, Department of Paediatrics, Inselspital, University of Bern, 3010 Bern, Switzerland. [96]The Autism Research Unit, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [97]State Diagnostic and Counseling Center, 200 Kopavogur, Iceland. [98]University of Iceland and Landspitali University Hospital, 101 Reykjavik, Iceland. [99]Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [100]Genetic Services, Rinnekoti Research Foundation, Kumputie 1, SF-02980 Espoo, Finland. [101]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. [102]Department of Endocrinology and Nutrition, Clinic Hospital of Virgen de la Victoria, Ciber Fisiopatologia Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 29010 Malaga, Spain. [103]Centre de Maladies Rares, Anomalies du Developpement Nord de France, CH Arras-CHRU Lille, 59000 Lille, France. [104]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. [105]Institute for Community Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. [106]Department of Medical Genetics, Haartman Institute, University of Helsinki and Folkhälsan Institute of Genetics, 00251 Helsinki, Finland. [107]Department of Psychiatry, Oslo University Hospital, N-0407 Oslo, Norway. [108]The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G IL7, Canada. [109]CNRS 8090-Institute of Biology, Pasteur Institute, 59800 Lille, France. [110]UOC Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy. [111]Institute of Human Genetics, Technische Universität München, Klinikum rechts der Isar, 81675 München, Germany.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Study cohorts. For the description of these cohorts, refer to Supplementary Information.

CNV detection. Cases ascertained for intellectual disabilities and developmental delay were identified through standard medical diagnostic procedures. CNV analyses of GWAS data were variously carried out using cnvHap[31]; a moving-window average-intensity procedure; a Gaussian mixture model (Valsesia *et al.*, submitted); circular binary segmentation[32,33]; QuantiSNP[34]; PennCNV[35]; BeadStudio GT module (Illumina Inc.); and Birdseed[36] (see below). At least two independent algorithms were used for each cohort.

Patients referred for intellectual disabilities and developmental delay. All diagnostic procedures (CGH, quantitative PCR and/or quantitative multiplex PCR of short fluorescent fragments) were carried out according to the relevant guidelines of good clinical laboratory practice for the respective countries. All rearrangements in probands were confirmed by a second independent method and karyotyping was performed in all cases to exclude a complex rearrangement.

Northern Finland 1966 birth cohort (NFBC). CNV calling has been previously described[10]. In brief, data were normalized using Illumina BeadStudio, then GC effects on ratios were removed by regressing on GC and GC2, and wave effects were removed by fitting a Loess function[37]. CNV analysis was done using cnvHap[31]. All called 16p11.2 duplications were validated by direct analysis of $\log_2$ ratios. Data for each probe were normalized by first subtracting the median value across all samples (so that the distribution of ratios for each probe 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). All CNV calls were confirmed by MLPA.

deCODE genetics. Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips were used for CNV analysis. BeadStudio (version 2.0) was used to call genotypes, normalize the signal intensity data and establish the log R ratio (LRR) and B allele frequency (BAF) at every SNP according to standard Illumina protocols. All samples passed a standard SNP-based quality control procedure with a SNP call rate greater than 0.97. PennCNV[35], a free, open-source tool, was used for detection of CNVs. The input data for PennCNV are LRR, a normalized measure of the total signal intensity for the two alleles of the SNP, and BAF, a normalized measure of the allelic intensity ratio of the two alleles. These values are derived with the help of control genotype clusters (HapMap samples), using the Illumina BeadStudio software. PennCNV employs a hidden Markov model to analyse the LRR and BAF values across the genome. CNV calls are made on the basis of the probability of a given copy state at the current marker, as well as on the probability of observing a copy-state change from the previous marker to the current one. PennCNV uses a built-in correction model for GC content[38].

Cohorte Lausannoise (CoLaus). Data normalization and CNV calling have been previously described[10]. Data normalization included allelic cross-talk calibration[39,40], intensity summarization using robust median average, and correction for any PCR amplification bias. Wave effects were corrected by fitting a Loess function[37]. CNV calling was done using a Gaussian mixture model (Valsesia *et al.*, submitted) that fits four components (deletion, copy-neutral, one additional copy and two additional copies) to copy-number 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[41] and CBS[32,33] algorithms, and by replicating a subset of CoLaus subjects on Illumina arrays. Only duplications found by both Gaussian mixture model and CBS were considered.

Estonian genome center of the University of Tartu (EGCUT). Genotypes were called by BeadStudio software GT module v3.1 or GenomeStudio GT v1.6 (Illumina Inc.). Values for LRR and BAF produced by BeadStudio were formatted for further CNV analysis and break-point mapping with Hidden-Markov-Model-based softwares QuantiSNP (ver.1.1)[34] and PennCNV[42] or CNVPartition 2.4.4 (Illumina Inc.). All analyses were carried out using the recommended settings, except changing EMiters to 25 and L to 1,000,000 in QuantiSNP. For PennCNV, the Estonian-population-specific SNP allele frequency data was used. All detected duplications were confirmed by quantitative PCR.

Study of health in Pomerania (SHIP). Raw intensities were normalized using Affymetrix power tools (Affymetrix); CNV analysis was done using Birdseye from the Birdsuite software package[36] and PennCNV[35]. PennCNV predictions with confidence scores less than 10 were removed. Birdsuite predictions were filtered as in ref. 15: CNVs were kept if their linkage disequilibrium (LOD) score was >10, length >1 kb, number of probes ≥ 5 and size per number of probes $<10{,}000$.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) F3 and F4. Genotyping for KORA F3 was performed using the Affymetrix 500K array set, consisting of two chips (StyI and NspI). The KORA F4 samples were genotyped with the Affymetrix human SNP array 6.0. For both studies, genomic DNA from blood samples was used for analysis. Hybridization of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotyping was done in the Genome Analysis Centre of the Helmholtz Centre Munich. Genotypes were determined using BRLMM clustering algorithm (Affymetrix 500K array set) and Birdseed2 clustering algorithm (Affymetrix array 6.0). For quality control purposes, we applied a positive control and a negative control DNA every 48 samples (KORA F3) or 96 samples (KORA F4). On the chip level, only subjects with overall genotyping efficiencies of at least 93% were included. In addition, the called gender had to agree with the gender in the KORA study database. After exclusions, 1,644 individuals remained in KORA F3 and 1,814 in KORA F4 for further analysis.

MLPA analysis. We used MLPA to determine changes in the copy number of a region of about 2 Mb on chromosome 16p11.2. Briefly, we designed, using hg18, nine probes within the targeted region, one control probe outside the rearranged region and seven control probes targeting unique position in the genome (Supplementary Table 10). Assays were performed with MRC-Holland reagents according to the manufacturer's protocol[43]. The analysis of the amplification products was performed by capillary electrophoresis in the DNA Analyser 3730XL and using the GeneMapper software v3.7 (Applied Biosystems). The calculations were performed independently for each experiment: we first normalized the MLPA data to minimize the amount of experimental variation, summing all signal values of each control probe for each sample, and then dividing each signal value of each sample by the sum. The normalized signal values were compared to signal values from all other samples in the same experiment, dividing the normalized signal values by the average calculated from all the samples in the same experiment. The product of this calculation is termed dosage quotient (DQ). A DQ value of less than 0.65 or more than 1.25 was considered as copy-number loss or gain, respectively, as previously described[44–46].

Custom array-CGH for the short arm of chromosome 16. DNA samples were labelled with Cy3 and cohybridized to custom-made Nimblegen arrays with Cy5-labelled DNA from the CEPH cell line GM12042. These arrays contained 71,000 probes spread across the short arm of chromosome 16 from 22.0 Mb to 32.7 Mb (at a median space of 45 bp between 27.5 Mb and 31.0 Mb), and 1,000 control probes situated in invariable regions of the X chromosome. DNA labelling, hybridization and washing were performed according to Nimblegen protocols. Scanning was performed using an Agilent G2565BA microarray scanner. Image processing, quality control and data extraction were performed using the Nimblescan software v.2.5.

Defining underweight. Underweight was defined throughout the study as BMI ≤ 18.5 kg per m^2 in adults and ≤ -2 s.d. in children[30,47,48].

Weight, height, BMI and head circumference *Z*-scores as a function of age. For paediatric cases, weight, height, BMI and head circumference *Z*-scores were determined for paediatric cases (0–18 years of age) using clinical growth charts specific to the country of origin. Children were ascertained from nine different countries. If charts were only available in percentiles, those measures were transformed into *Z*-scores using gender-, age- and geographically-matched reference populations (see Statistics).

For the USA and Canada, data from the Center for Disease Control and National Center for Health Statistics (CDC/NCHS) were used to calculate *Z*-scores[49].

For the French paediatric population, we used French national growth charts[50,51]. For the Swiss paediatric population, we used Swiss national growth charts[52]. For Dutch participants, Dutch national growth charts were used[53]. For Italian, German, Finnish and Austrian cases ($n = 6$), height, weight and BMI *Z*-scores were estimated using WHO growth charts[54].

To check for discrepancies generated by the use of different growth charts, height, weight and BMI *Z*-scores were recalculated using WHO growth charts for all cases under five years of age, regardless of origin (http://www.who.int/childgrowth/standards/en/54). *Z*-scores obtained using the WHO data were not significantly different. These growth standards, developed by the World Health Organization multicentre growth reference study, describe normal child growth from birth to 5 years under optimal environmental conditions. These standards can be applied to all children everywhere, regardless of ethnicity, socioeconomic status and type of feeding[55,56].

If necessary, percentile values were transformed to *Z*-scores by the inverse-normal density function. When growth charts were unavailable, we used reported LMS parameters (median (M), generalized coefficient of variation (S) and skewness (L)) to obtain *Z*-scores via the formula:

$$Z\text{-score} = \begin{cases} \frac{(X/M)^L - 1}{L*S}, & L > 0 \\ \ln(X/M)/S, & L = 0 \end{cases},$$

in which X is the observed value.

In adults (>18 years of age), we estimated LMS parameters when these were unavailable from the available sex-, age- and origin-matched Swiss (CoLaus),

©2011 Macmillan Publishers Limited. All rights reserved

Estonian or French control populations. For cases identified from population-based cohorts, *Z*-scores were directly inferred from the cohort.

Gene expression. We established lymphoblastoid cell lines from deletion and duplication carriers, as well as from controls (Supplementary Table 12), by transforming peripheral blood mononuclear cells with Epstein–Barr virus. Patients and controls were enrolled after obtaining appropriate informed consent via the physicians in charge, and approval by the ethics committee of the University of Lausanne. More control cell lines were obtained from Coriell Institute for Medical Research (http://www.coriell.org/) (Supplementary Table 12). SYBR Green real-time quantitative PCR (RT–PCR) was performed as previously described[22,57]. Briefly, 1 μg of total RNA from lymphoblastoid cell lines was converted to complementary DNA using Superscript VILO (Invitrogen) primed with a mixture of oligo(dT) and random hexamers. Oligos were designed using the PrimerExpress program (Applied Biosystems) with default parameters (Supplementary Table 11). Non-intron-spanning assays were tested for genomic contamination in standard ± reverse transcriptase reactions. The amplification efficiency of each primer pair was tested in a cDNA dilution series, as previously described[58]. A full list of genes mapping in the rearranged interval, and exclusion criteria, are presented in Supplementary Table 1. All RT–PCR reactions were performed in a 10-μl final volume and triplicates per sample. The setup in a 384-well plate format was performed using a Freedom EVO robot (TECAN) and assays were run in an ABI 7900 sequence detection system (Applied Biosystems) with the following amplification conditions: 50 °C for 2 min, 95 °C for 10 min, and 45 cycles of 95 °C 15 s, then 60 °C for 1 min. A final incubation of 95 °C for 15 s followed by 60 °C for 15 s was carried out to establish a dissociation curve. Each plate included the appropriate normalization genes to control for any variability between plate runs. Raw threshold cycles (Ct) values were obtained using SDS2.4 (Applied Biosystems). To calculate the normalized relative expression ratio of individuals carrying the CNV and of controls, we used Biogazelle qBase Plus software[59] including geNorm[60]. This program identified appropriate normalization genes (*EEF1A1*, *RPL13*, *GUSB* and *TBP*) having a gene-stability measure of M = 0.25. We note that one gene, *LAT*, showed a very high expression profile in one of the duplication samples (DASYL, Supplementary Table 13), reaching a relative expression value of 27.3 (s.e.m. = 1.37), compared to an average expression for other duplications of 1.89 (s.e.m. = 0.51). We cannot exclude that this finding is genuine (and confirmed it in a second experiment), but it was removed from further analyses as an outlier to give a more accurate overview of expression profiles for these genes.

In silico analysis was performed to check for brain, and specifically hypothalamus, expression of genes in the rearranged 16p11.2 interval (Supplementary Table 1). This was done using Allen Brain Atlas Resources, available from http://www.brain-map.org.

Cases with major neurological signs. Major neurological signs were defined by moderate to severe hypotonia, hypertonia, ataxia, spasticity, hypereflexia, hyporeflexia and/or extra-pyramidal signs, and by the presence of epilepsy.

Statistics. Student's *t*-test: one-tailed *t*-tests were performed to test whether duplication carriers have *Z*-score values lower than zero for BMI, height and weight. We found this analysis more suitable than linear regression analysis, correcting for confounding factors such as sex and age, because these anthropometric traits have a highly nonlinear dependence on these factors, as can be observed in control populations.

Kruskal–Wallis test: this was used to test differences in the gene expression pattern between deletion and duplication carriers and control individuals. Because expression values are not necessarily normally distributed, this test is more adequate than a classical one-way analysis of variance. To test pairwise differences, we computed the difference in mean group rank with its 95% confidence interval (as provided by the multcompare function in Matlab). Correction for multiple testing was done using a Bonferroni adjustment.

Multiple testing: we determined false-discovery-rate-based thresholds for association *P*-values for each phenotype, to correct for multiple testing. For each phenotype, we replaced the observed *Z*-scores with numbers randomly drawn from a standard normal distribution and performed the same *t*-tests for the same strata. The procedure was repeated 1,000 times. For various *P*-value thresholds, we asked how many tests would be declared significant for the null set on average (over the 1,000 random draws). The false discovery rate was estimated as the ratio of this number and the actual number obtained for the observed *Z*-scores. Thus, we controlled the dependence between nested tests.

Relative risk: among adults, we defined underweight as a BMI <18.5 (WHO criteria). The estimated relative risk is the ratio of the fraction of underweight individuals among duplication carriers versus our control group. The standard error of log(relative risk) and its significance were calculated as previously described[61]. In our control group (population-based cohorts), the frequency of being underweight is 1.9% (38 males and 148 females out of 9,470). Owing to the fact that being underweight decreases with age in the general population, we resampled our control group to ensure precise age-matching.

31. Coin, L. J. *et al.* cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nature Methods* **7,** 541–546 (2010).
32. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).
33. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23,** 657–663 (2007).
34. Colella, 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).
35. 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).
36. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genet.* **40,** 1253–1260 (2008).
37. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8,** R228 (2007).
38. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36,** e126 (2008).
39. 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).
40. 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).
41. Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1,** 287–299 (2004).
42. Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316,** 1194–1198 (2007).
43. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30,** e57 (2002).
44. Bunyan, D. J. *et al.* Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. *Br. J. Cancer* **91,** 1155–1159 (2004).
45. Fernández, L. *et al.* Comparative study of three diagnostic approaches (FISH, STRs and MLPA) in 30 patients with 22q11.2 deletion syndrome. *Clin. Genet.* **68,** 373–378 (2005).
46. Slater, H. R. *et al.* Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J. Med. Genet.* **40,** 907–912 (2003).
47. Mei, Z. *et al.* Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am. J. Clin. Nutr.* **75,** 978–985 (2002).
48. Physical. status: the use and interpretation of anthropometry. Report of a WHO expert committee. *World Health Organ. Tech. Rep. Ser.* **854,** 1–452 (1995).
49. Kuczmarski, R. J. *et al.* CDC growth charts: United States. *Adv. Data* **314,** 1–27 (2000).
50. Sempé, M., Pedron, G. & Roy-Pernot, M. P. *Auxologie, Méthode et Séquences.* (Théraplix, 1979).
51. Rolland-Cachera, M. F. *et al.* Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45,** 13–21 (1991).
52. Prader, A., Largo, R. H., Molinari, L. & Issler, C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv. Paediatr. Acta.*, Suppl. 52, 1–125 (1989).
53. Fredriks, M. *Growth Diagrams, 1997. Fourth Dutch Nation-wide Survey.* 233–242 (Bohn Stafleu Van Loghum, 1997).
54. de Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85,** 660–667 (2007).
55. de Onis, M., Garza, C., Onyango, A. W. & Borghi, E. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *J. Nutr.* **137,** 144–148 (2007).
56. Mei, Z., Ogden, C. L., Flegal, K. M. & Grummer-Strawn, L. M. Comparison of the prevalence of shortness, underweight, and overweight among US children aged 0 to 59 months by using the CDC 2000 and the WHO 2006 growth charts. *J. Pediatr.* **153,** 622–628 (2008).
57. Molina, J. *et al.* Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki–Lupski syndrome. *Hum. Mol. Genet.* **17,** 2486–2495 (2008).
58. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25,** 402–408 (2001).
59. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8,** R19 (2007).
60. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** 1–11 (2002).
61. Morris, J. A. & Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br. Med. J. (Clin. Res. Ed.)* **296,** 1313–1316 (1988).

©2011 **Macmillan Publishers Limited. All rights reserved**

Copyright of Nature is the property of Nature Publishing Group and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.