Large Copy-Number Variations Are Enriched in Cases With Moderate to Extreme Obesity

Kai Wang,1 Wei-Dong Li,2 Joseph T. Glessner,1 Struan F.A. Grant,1,3 Hakon Hakonarson,1,3 and R. Arlen Price2

OBJECTIVE—Obesity is an increasingly common disorder that predisposes to several medical conditions, including type 2 diabetes. We investigated whether large and rare copy-number variations (CNVs) differentiate moderate to extreme obesity from never-overweight control subjects.

RESEARCH DESIGN AND METHODS—Using single nucleotide polymorphism (SNP) arrays, we performed a genome-wide CNV survey on 430 obese case subjects (BMI >35 kg/m²) and 379 never-overweight control subjects (BMI <25 kg/m²). All subjects were of European ancestry and were genotyped on the Illumina HumanHap550 arrays with ~550,000 SNP markers. The CNV calls were generated by PennCNV software.

RESULTS—CNVs >1 Mb were found to be overrepresented in case versus control subjects (odds ratio [OR] = 1.5 [95% CI 0.5–5]), and CNVs >2 Mb were present in 1.3% of the case subjects but were absent in control subjects (OR = infinity [95% CI 1.2–infinity]). When focusing on rare deletions that disrupt genes, even more pronounced effect sizes are observed (OR = 2.7 [95% CI 0.5–27.1] for CNVs >1 Mb). Interestingly, obese case subjects who carry these large CNVs have moderately high BMI and do not appear to be extreme cases. Several CNVs disrupt known candidate genes for obesity, such as a 3.3-Mb deletion disrupting NAP1L5 and a 2.1-Mb deletion disrupting UCP1 and IL15.

CONCLUSIONS—Our results suggest that large CNVs, especially rare deletions, confer risk of obesity in patients with moderate obesity and that genes impacted by large CNVs represent intriguing candidates for obesity that warrant further study.

Diabetes 59:2690–2694, 2010

Obesity has become the most common health disorder worldwide. Obesity predisposes to multiple diseases, particularly diabetes, and it has been estimated that life expectancy may diminish in the next generation as a result (1). Numerous studies have shown that body weight and obesity are strongly influenced by genetic factors, with heritability estimates in the range of 65–80% (2). However, single gene mutations are quite rare, and common variation (e.g., in FTO [3] and MC4R [4]) account for a small percentage of familial risk. Recent large-scale meta-analysis of genome-wide association studies (GWASs) identified six additional genes that associate with BMI, but all eight genes collectively explain merely 0.84% of the BMI variation in human populations (5). Therefore, it is unlikely that expansion of sample sizes in GWASs will identify common variants with major effect sizes.

The examination of copy-number variations (CNVs) offers novel insights into the genetic architecture of common and complex human diseases. CNVs are defined as a chromosomal segment whose copy number varies across individuals in the population (6). Recurrent CNVs such as 16p11.2 deletions were reported to account for 0.7% of morbid obesity cases (7). In addition, several reports demonstrated that large and rare CNVs collectively associate with schizophrenia (8–10), extreme early-onset obesity (11), and variation in BMI (12).

In the current study, we investigated the potential role of rare variants in obesity, by performing comparative CNV analysis on obese case and control subjects who were genotyped by Illumina single nucleotide polymorphism (SNP) arrays. Case subjects had moderate to extreme obesity, and control subjects had never been overweight. Although our sample size precludes the definitive identification of specific CNVs/gene that associate with obesity, we demonstrate that large, yet rare, CNVs, as a group, are collectively associated with obesity. Furthermore, we identified previously implicated obesity candidate genes in some of these large and rare CNVs, making them especially attractive for additional follow-up studies and functional assays.

RESEARCH DESIGN AND METHODS

Obese case and control subjects. The case subjects were obese (BMI ≥35 kg/m²) with a lifetime BMI >40 kg/m². Independent control subjects were selected who had a current and lifetime BMI ≤25 kg/m². All the case and control subjects who participated in the current study were part of a previous candidate gene study (13). Sample characteristics are summarized in Table 1 for 430 case and 370 control subjects passing quality control. The median age at obesity onset was 12 years, and 90% had an onset prior to age 26 years. All subjects gave informed consent, and the protocol was approved by the committee on studies involving human beings at the University of Pennsylvania.

SNP genotyping. DNA was extracted from whole blood or lymphoblastoid cell lines using a high-salt method and genotyped on the Illumina HumanHap550 SNP arrays (Illumina, San Diego, CA). Standard Illumina data normalization procedures and canonical genotype clustering files were used to process the genotyping signals. All case and control subjects passed call-rate (>95%) measures and were genetically inferred to be of European ancestry, based on multidimensional scaling analysis (supplementary Fig. 1 in the online appendix, available at http://diabetes.diabetesjournals.org/cgi/content/full/db10-0192/DC1).

CNV calling. Using log R ratio and B allele frequency measures for all markers, the CNV calls were generated by PennCNV software (Version 2009Aug27) (14). The quality-control procedure was described in detail in supplementary Fig. 2. We removed samples with low quality of signal intensity

From the 1Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; the 2Center for Neurobiology and Behavior, Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania; and the 3Department of Pediatrics, University of Pennsylvania, Philadelphia, Pennsylvania. Corresponding author: R. Arlen Price, arlen@mail.med.upenn.edu. Received 8 February 2010 and accepted 27 June 2010. Published ahead of print on http://diabetes.diabetesjournals.org on 9 July 2010. DOI: 10.2337/db10-0192. © 2010 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See http://creativecommons.org/licenses/by-nc-nd/3.0/ for details.
values, so that the remaining samples have log R ratio < -0.3, B allele frequency_ drift < 0.01, wave factor < 0.05, and that the number of calls is < 50. We removed CNV calls with <10 SNPs or with a confidence score < 10, sparse calls (average intermarker distance > 50 kb), calls in the immunoglobulin regions, and calls in centromeric regions and telomeric regions (100 kb within the start or end of the chromosomes). The overlapping genes or exons for CNV calls were annotated using the scan_region.pl program, based on RefSeq gene annotation (15). We compiled a set of common CNV regions (cCNVRs), which occur at >1% frequency, and then classified the CNV call as common or rare by the scan_region.pl program: if >50% of a CNV call overlaps with a cCNVR, it is referred to as a common CNV. The comparison of number of CNV calls in case versus control subjects was performed by t test, while the comparison of fraction samples with large CNVs was performed by the Fisher exact test.

RESULTS

CNV calling and quality control. To examine whether CNVs represent genetic risk factors for obesity, we analyzed CNV calls on 430 obese case and 379 control subjects who were genotyped by Illumina SNP arrays and passed quality-control measures for CNV analysis. The sample characteristics were described in Table 1. We first compared the general characteristics of CNV calls between case and control subjects. The number of CNVs per subject did not differ between case and control subjects (5.8 ± 3.9 vs. 6.0 ± 3.1, P = 0.35). The number of gene-disrupting CNVs per subject is similar in case versus control subjects (3.8 ± 3.1 vs. 4.2 ± 2.7, P = 0.07). Similarly, the number of exonic CNVs per subject is similar in case versus control subjects (3.2 ± 2.9 vs. 3.6 ± 2.6, P = 0.06). We compiled a list of common CNV regions and found that 38.2% of CNV calls can be classified as rare CNVs. The number of rare CNVs per subject did not differ between obese case and control subjects (2.3 ± 2.4 vs. 2.2 ± 1.8, P = 0.41).

Large CNVs are overrepresented in obese case subjects. We next performed comparative analysis on CNV calls stratified by their sizes, common/rare status, and deletion/duplication status. Interestingly, with the increasing size thresholds, we observe a stronger trend of association (odds ratio [OR]) between CNVs and obesity (Table 2). Similar to previous reports in schizophrenia cases (8), we found that 5/427 (1.2%) of the case subjects but none of the control subjects carry CNVs > 2 Mb (OR = infinity [95% CI 1.16 to infinity]), P = 0.04. The frequency of obese case subjects carrying CNVs > 2 Mb in our study are similar to the Kirov et al. study (16) (6 of 471, 1.3%) and the Need et al. study (8) (14 of 1,013, 1.4%) on schizophrenia cases. Among five CNVs > 2 Mb observed in our study, three are deletions and two are duplications. We listed all 16 CNVs > 1 Mb in case and control subjects in Table 2, and the signal intensity patterns are provided in supplementary Fig. 3 as a visual means of validation. We also assessed whether large and rare gene-disrupting deletion CNVs tend to be enriched in case versus control subjects. Not surprisingly, the ORs for conferring risk of obesity are even higher for this group of CNVs (2.7 [0.47–27.1] for > 1 Mb CNVs, infinity for > 2 Mb CNVs) (Table 2), though this does not reach statistical significance due to the rare nature of the events.

Multiple large and rare CNVs disrupt obesity candidate genes. We next examined CNVs > 1 Mb and found several genes that are a priori candidates for obesity. Two of the strongest candidates are UCP1 and IIL15, which are located within the same 2.1-Mb deletion on chromosome 4q31 (Fig. 1). The case carrying this CNV has moderate obesity (BMI 46.2 kg/m²). Numerous studies relate UCP1 to obesity in animal models (17), and associations have been reported in humans (18). We validated this CNV by a CNV-typing platform, the Affymetrix Cytogenetic arrays (supplementary Fig. 4). Since parental DNA is also available, we assessed both parents and found that the CNV is inherited from the father. Another large CNV on chromosome 4q22.1 contains two potential candidate genes (NAPIL5 and SNCA), and it is present in a subject with moderate obesity (BMI 49.0 kg/m²) (Fig. 2). NAPIL5 is an imprinted gene, which is of interest because of associations of body weight and obesity with genomic imprinting (19). Differences in paternal and maternal copies of this gene have been related to body weight at birth and in adulthood in mice (20). We validated this CNV by the Affymetrix Cytogenetic platform (supplementary Fig. 4) and also found that the CNV is inherited from the father. SNCA is another gene within this CNV that has been reported to have interactive effects on response to a high-fat diet in dietary obesity (21), yet SNCA duplication is a well-known risk factor for Parkinson’s disease. Several other candidate genes, such as CTSC, NOX4, DLG2, ME3, and MIPEP, are also found within the collection of rare CNVs in case subjects (Table 3). We acknowledge that this list is relatively small and that none of them occur twice in case subjects; as a result, we detected the collective

<table>
<thead>
<tr>
<th>Size (kb)</th>
<th>n (case subjects with CNVs)</th>
<th>n (controls with CNVs)</th>
<th>OR (95% CI)</th>
<th>n (case subjects with gene-disrupting deletion)</th>
<th>n (control subjects with gene-disrupting deletion)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;100</td>
<td>352</td>
<td>313</td>
<td>0.99 (0.65–1.39)</td>
<td>88</td>
<td>66</td>
<td>1.22 (0.84–1.77)</td>
</tr>
<tr>
<td>&gt;500</td>
<td>34</td>
<td>30</td>
<td>1.01 (0.46–1.37)</td>
<td>9</td>
<td>6</td>
<td>1.33 (0.42–4.58)</td>
</tr>
<tr>
<td>&gt;1 Mb</td>
<td>10</td>
<td>6</td>
<td>1.49 (0.48–5.00)</td>
<td>6</td>
<td>2</td>
<td>2.67 (0.47–27.1)</td>
</tr>
<tr>
<td>&gt;2 Mb</td>
<td>5</td>
<td>0</td>
<td>Infinity (1.16 to infinity)</td>
<td>3</td>
<td>0</td>
<td>Infinity (0.69 to infinity)</td>
</tr>
<tr>
<td>&gt;5 Mb</td>
<td>1</td>
<td>0</td>
<td>Infinity (0 to infinity)</td>
<td>0</td>
<td>0</td>
<td>Infinity</td>
</tr>
</tbody>
</table>

K. WANG AND ASSOCIATES
association with obesity but cannot identify specific CNVs/gens that are more penetrant than others. Finally, we also did an exploratory examination to determine whether some CNVs are unique to the extremely obese case subjects. We chose a BMI threshold of 70 kg/m², which doubles the minimum entry criteria for case subjects. However, compared with case subjects with moderate obesity, the extremely obese case subjects do not appear to have larger CNVs or more well-characterized candidate genes.

**Examination of previously reported obesity-associated CNVs.** An association between BMI and a chromosome 10q11 CNV was recently reported in a Chinese cohort (12). We observed three case subjects carrying this CNV (BMI 56, 41, and 43 kg/m², respectively), but it is not present in control subjects. Two genes in this region are **GPRIN2** and **PPYR1**, which are worthy of follow-up studies in larger sample sets. Additionally, a highly penetrant deletion on 16p11.2 was recently reported to be associated with obesity (7,11). In our data, one obese subject (BMI 44.9 kg/m²) carries this deletion and one control subject (BMI 19.1 kg/m²) carries the reciprocal duplication. Therefore, our data are consistent with the possibility that the 16p11.2 deletion is associated with obesity.

**DISCUSSION**

In the current study, we assayed a sample collection of obese case subjects and never-overweight control subjects and found strong support that large and rare CNVs contribute to obesity. Collectively, the OR for large CNVs observed in our study is higher than common SNPs (for example, the OR for **FTO** is 1.3 [3] and for **MC4R** in severe childhood obesity is 1.3 [4]), suggesting that rare CNVs may represent more penetrant risk factors for obesity.

One interesting implication of our study relates to the hypothesized genetic architecture of obesity. Although it is well known that obesity results from multiple genetic risk factors as well as environmental factors, it is not clear what and how many genetic risk factors are involved. Recent GWASs identified a few obesity genes, but they collectively only explain a minor fraction of interindividual differences in obesity (5). Therefore, even though more common susceptibility variants may be identified by increasing sample size, they will be very unlikely to account for a significant proportion of genetic risk. On the other hand, our study suggests that rare variants with much higher ORs may also contribute to risk of obesity. Given the rare nature of the CNVs, we could not discern which of these large CNVs are truly causal for obesity, so some less penetrant or noncausal large CNVs will dilute the effect sizes. Therefore, the observed effect sizes for large CNVs may represent underestimation of the true effect size of causal CNVs for obesity.

Another interesting implication is how quantitative genetics relates to disease phenotypes. How distinct alleles, including modest-effect alleles and major-effect alleles, may interact to shape disease presentation is not well studied. For obesity, although **FTO** represents consistently the strongest gene in many association studies, it has never been implicated from studies of monogenic forms of obesity. Similarly, although **MC4R** has been implicated in monogenic forms of obesity, analysis of common variants have been highly inconsistent until large-scale GWASs are conducted (4). Therefore, it is likely that rare alleles work together with common alleles to shape the onset of obesity in human populations and that some genes with rare causal alleles may never show up from studies on common variants.

In conclusion, we have identified large, yet rare, CNVs representing major risk factors for obesity. Some of these large CNVs encompass known obesity genes or potential
TABLE 3
Description of CNVs >1 Mb in case and control subjects

<table>
<thead>
<tr>
<th>Region (NCBI 36)</th>
<th>Number of SNPs</th>
<th>Length</th>
<th>Type</th>
<th>Phenotype</th>
<th>BMI</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr2:106245033–108062350</td>
<td>279</td>
<td>1,562,513</td>
<td>Del</td>
<td>Case</td>
<td>52.09</td>
<td>LOC729121, PLG1A, RGP3D, ST6GAL2</td>
</tr>
<tr>
<td>chr10:141598764–143656669</td>
<td>403</td>
<td>2,057,906</td>
<td>Del</td>
<td>Case</td>
<td>46.16</td>
<td>BMS1, LOC414666, LOC84269, RET, ZNF353, ZNF37B</td>
</tr>
<tr>
<td>chr16:80739605–82222770</td>
<td>822</td>
<td>1,483,166</td>
<td>Del</td>
<td>Case</td>
<td>45.37</td>
<td>CDH13, MPHOSPH6</td>
</tr>
<tr>
<td>chr3:22153141–24201255</td>
<td>649</td>
<td>2,048,115</td>
<td>Del</td>
<td>Case</td>
<td>36.94</td>
<td>C11orf73, CCDC31, CCDC83, CCDC89, CREB2P, DLOD, E2D, E3, PICALM, SYTL2, TMEM126A, TMEM126B, CHORDC1, CTCF, FOLHL1B, GRM3, LOC72938, NAALAD2, NOX3, RAB38, TMEM135, TMIM49, TMIM53, TMIM64, TRIM64B, TRIM77, TUR, UBRFL1, ATP12A, C1QTNF9, C1QTNF9B, LOC37449, MIPER, MIR2276, PARP4, PCOGH, SACS, SGCG, SPATA13, TFR5S1F9, ABC2C1, ABCB6, C16orf65, C16orf63, KIAA0430, MIR38, MPV17L, MYH11, NDE1, NTAN1, PDXDC1, RNR3</td>
</tr>
<tr>
<td>chr18:4663080–6830148</td>
<td>525</td>
<td>2,167,069</td>
<td>Del</td>
<td>Control</td>
<td>51.12</td>
<td>PABPCP2</td>
</tr>
<tr>
<td>chr5:103816450–104873156</td>
<td>172</td>
<td>1,056,707</td>
<td>Del</td>
<td>Control</td>
<td>21.16</td>
<td>RAB9P1</td>
</tr>
<tr>
<td>chr16:15032942–16197033</td>
<td>201</td>
<td>1,164,092</td>
<td>Del</td>
<td>Control</td>
<td>20.13</td>
<td>(not gene disrupting)</td>
</tr>
<tr>
<td>chr17:14063278–15411904</td>
<td>461</td>
<td>1,348,627</td>
<td>Del</td>
<td>Control</td>
<td>20.12</td>
<td>CDRT15, CDRT9, FAM18B2, HGSTST21B, MGC12916, PMP22, TEKT3</td>
</tr>
</tbody>
</table>

Underlining indicates DNA from cell line. All others are from blood.

We thank all the case and control subjects who donated blood samples for genetic research purposes.

REFERENCES

LARGE COPY-NUMBER VARIATIONS IN OBESITY


Price RA, Li WD, Zhao H. FTO gene SNPs associated with extreme obesity in cases, controls and exceptionally discordant sister pairs. BMC Med Genet 2009;9:4

