

RESEARCH ARTICLE

Open Access



A vast genomic deletion in the C56BL/6 genome affects different genes within the *Ifi200* cluster on chromosome 1 and mediates obesity and insulin resistance

Heike Vogel^{1,2}, Markus Jähnert^{1,2}, Mandy Stadion^{1,2}, Daniela Matzke^{1,2}, Stephan Scherneck³ and Annette Schürmann^{1,2*}

Abstract

Background: Obesity, the excessive accumulation of body fat, is a highly heritable and genetically heterogeneous disorder. The complex, polygenic basis for the disease consisting of a network of different gene variants is still not completely known.

Results: In the current study we generated a BAC library of the obese-prone NZO strain to clarify the genomic alteration within the gene cluster *Ifi200* on chr.1 including *Ifi202b*, an obesity gene that is in contrast to NZO not expressed in the lean B6 mouse. With the PacBio sequencing data of NZO BAC clones we identified a deletion spanning approximately 261.8 kb in the B6 reference genome. The deletion affects different members of the *Ifi200* gene family which also includes the original first exon and 5'-regulatory parts of the *Ifi202b* gene and suggests to be the relevant cause of its expression deficiency in B6. In addition, the generation and characterization of congenic mice carrying the critical fragment on the B6 background demonstrate its crucial role for obesity and insulin resistance.

Conclusions: Our data reveal the reconstruction of a complex genomic region on mouse chr.1 resulting from deletions and duplications of *Ifi200* genes and suggest to be relevant for the development of obesity. The results further demonstrate the complexity of the disease and highlight the importance for studying rare genetic variants as they can be causal for large effects.

Keywords: BAC library, Gene cluster, Deletion, Obesity

Background

Obesity is the consequence of an imbalance between food intake and energy expenditure resulting in an excess accumulation of body fat. Progress and course of obesity and its associated diseases are dependent on nutritional conditions and on other lifestyle parameters (e.g. physical activity). However, its main basis is the complex, polygenic predisposition consisting of a

network of variant genes which are still not completely known and difficult to identify in humans. Genome-wide association studies (GWAs) are commonly used for the identification of disease genes. Nevertheless, the identified loci accounting only for a small proportion of the heritability of a complex disease like obesity [1]. Genomic structural variants (GSVs) may explain rare variants with large effects, which are not readily identifiable via SNP-based methods [2–4].

Genomic linkage studies in rodents are a suitable approach to identify and to study such chromosomal alterations. In a previous study we reported the identification of a major obesity QTL (*Nob3*) on distal mouse chr.1 in an outcross population of the New Zealand obese

* Correspondence: schuermann@dife.de

¹Department of Experimental Diabetology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert Allee 114-116, D-14558 Nuthetal, Germany

²German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, 85764 München-Neuherberg, Germany

Full list of author information is available at the end of the article



(NZO) strain, a polygenic mouse model for obesity, and the lean C57BL/6J (B6) mouse. By generating recombinant congenic lines and expression studies we finally identified the *Ifi202b* (*Interferon inducible gene 202b*) gene as the causal variant of the *Nob3* locus. The transcriptional regulator *Ifi202b* is a member of the *Ifi200* gene family, which has also been annotated as the PYHIN family, acknowledging the defining features of an N-terminal pyrin domain and C-terminal HIN domain [5]. The proteins are involved in the defense against infection through recognition of foreign DNA, whereas *Ifi202b* was also shown to be involved in the development of obesity [6]. The gene family is arranged as a cluster on mouse chromosome 1 (1q band H3) between the *Cell adhesion molecule 3* (*Cadm3*) gene and a cluster of olfactory receptors. The *Ifi202b* gene is expressed in various tissues of the NZO strain but not transcribed in B6 mice and we hypothesized that this is due to a deletion of the first exon and the 5'-regulatory region [6]. The lack of *Ifi202b* is specific for C57BL mice (e.g. C57BL/10J, C57BL/6J, C57BLKS/J, and C57BR/sdJ), whereas most other strains (e.g. SJL/Bm, DBA2/J, BALB/cJ, C3H/HeJ, and FVB/NJ) express this gene [7].

In the current study we clarified the exact genomic structural variation causing the *Ifi202b* deficiency and demonstrated that a rare genomic alteration on mouse chr.1 is responsible for the development of obesity. We generated a NZO BAC library and performed a *de novo* assembly of the complex *Ifi200* region on mouse chr. 1 by using PacBio long reads, a third generation sequencing (TGS) approach and characterized mice with the affected region in respect to different metabolic traits.

Methods

Bacterial artificial chromosome (BAC) library construction and screening

NZO (NZO/HIBomDife) BAC library was constructed from high molecular weight (HMW) genomic DNA processed at Amplicon Express Inc. (Pullman, WA, USA) from liver tissue. All animal experiments were approved by the ethics committee of the State Office of Environment, Health and Consumer Protection (V3-2347-21-2012, Federal State of Brandenburg, Germany). With the restriction enzyme HindIII the HMW DNA was partially digested (average size 135 kb) and ligated into the pCC1BAC vector. Ligations were transformed into DH10B *E.coli* cells and plated on LB agar. Clones were picked and arranged onto 384-well plates, replicated and frozen at -80°C . Screening of the BAC library was also processed by Amplicon Express Inc. by using nylon filters with arrayed library clones (18,432 clones) and digoxigenin (DIG)-labeled probes representing position 11,239–11,453 in the genomic sequence of *Ifi202b*

(NC_000067). The DIG-labeled probe was generated from gDNA by PCR using the primers *Ifi202b_for*: TCTTCAGAGTGATGGTGTTTCG and *Ifi202b_rev*: TGTTTGCAAGTGAAGATCACAA. The *Ifi202b* probe was found to hybridize to 14 BAC clones with a size of 90–196 kb. Two positive clones with a size of 147 kb and 196 kb were selected for sequencing. Isolation of the high molecular weight plasmid from the *E.coli* cultures was performed with the PhasePrepTM BAC DNA Kit (Micro Scale Preparation, Sigma-Aldrich, Steinheim, Germany) and the BACMAXTM DNA Purification Kit (Biozym, Hessisch Oldendorf, Germany) according to the manufacturer's instructions. The PhasePrep BAC DNA Kit was used for cell harvesting, lysis, neutralization, and nucleic acid precipitation, whereas digestion of the residual RNA, removal of residual impurities and final precipitation was done with the BACMAX Kit.

BAC sequencing and sequence assembly

Sequencing of the two BAC clones (mixture, ratio 1:1) and assembling was processed by GATC Biotech AG (Konstanz, Germany) using the SMRT[®] Technology PacBio RS II. *De novo* assembly of BAC inserts was performed with the standard SMRT Portal Software including quality filtering of the reads, improvement of long reads through alignment of short reads, assembly of long reads, and assembly correction. The assembly of the reads was based on the hierarchical genome-assembly process (HGAP).

Comparative genomic hybridization assay

Genomic DNA was prepared from the tail of C57BL/6J and NZO/HIBomDife mice. Unamplified genomic DNA was labeled with Cy3 (NZO) or Cy5 (reference strain, C57BL/6J) and hybridization was performed by imagoGenes (Berlin, Germany) using the NimbleGen platform.

Animals

Breeding and genotyping

All animal experiments were approved by the ethics committee of the State Office of Environment, Health and Consumer Protection (Federal State of Brandenburg, Germany). NZO mice from our own colony (NZO/HIBomDife) and C57BL/6J (Charles River, Sulzfeld, Germany) were used throughout the study. Mice were kept at a temperature of $20 \pm 2^{\circ}\text{C}$ with a 12:12 h light-dark cycle and had *ad libitum* access to drinking water and to a high-fat diet (HFD) containing 45 kcal% from fat, 35 kcal% from carbohydrates, and 20 kcal% from protein (D12451, Research Diets, Inc., New Brunswick, USA). Congenic mice were generated on a B6 background and the offspring was selected in each generation for carrying the fragment 163.5–177.7 Mbp from NZO

on chr.1 (*Nob3.14*). Phenotypical characterization of female congenic mice were performed in the F10N8 generation. For genotyping, DNA was prepared from mouse tails with a DNA isolation kit based on a salt precipitation method (InViTek, Berlin, Germany) and used for tests with polymorphic microsatellite markers. Microsatellites (D1Mit143 and D1Mit115) were genotyped by PCR with oligonucleotide primers obtained from MWG (Ebersberg, Germany), and the microsatellite length was determined by non-denaturing polyacrylamide gel electrophoresis.

Body composition and blood glucose

Fat mass of *Nob3.14* mice were determined by nuclear magnetic resonance (EchoMRI™-100H, EchoMRI LCC, Houston, USA) and blood glucose levels were measured in the morning (7–10 a.m.) using a CONTOUR® XT glucometer (Bayer, Leverkusen, Germany).

Histological analysis of adipose tissue

Paraffin sections (2 µm) of gonadal white adipose tissue (gonWAT) of 30-week-old *Nob3.14* mice were stained with hematoxylin and eosin. Microscopic images were captured with the Keyence BZ-9000 fluorescent microscope and the corresponding BZ-II Analyzer software (Keyence International, Mechelen, Belgium).

Metabolic phenotyping

Oral glucose tolerance tests (OGTT) were performed in 22-week-old mice. Mice were fasted overnight and received 2 g/kg body weight of glucose (Glucosteril® 20%, Fresenius Kabi, Bad Homburg, Germany). Blood glucose and insulin concentrations were detected up to 120 min.

Plasma analysis

Plasma insulin levels were analyzed using the Mouse Ultrasensitive Insulin ELISA (ALPCO Diagnostics, Salem, USA) following the manufacturer's instructions.

Protein extraction and western blotting

Adipose tissue of *Nob3.14* mice were homogenized in TES buffer (20 mM TrisHCl, 1 mM EDTA, 8.7% sucrose, pH 7.4, supplemented with protease inhibitor cocktail). Proteins were separated by SDS-PAGE, transferred to a PVDF membrane (Immobilon-P Membrane, Merck Milipore, Darmstadt, Germany) and targeted proteins were detected by ECL Prime Detection Reagent (GE Healthcare Europe GmbH, Freiburg, Germany) using the FUSION-SL4 advanced chemiluminescence system (Peqlab Biotechnologie GmbH, Erlangen, Germany). Primary antibodies against PPARγ (ab41928, Abcam, Cambridge, UK), pHSL (#4139S, Cell Signaling, Beverly, MA, USA), tHSL (#4107S, Cell Signaling), β-ACTIN (A3854, Sigma-Aldrich, St. Louis, USA), and

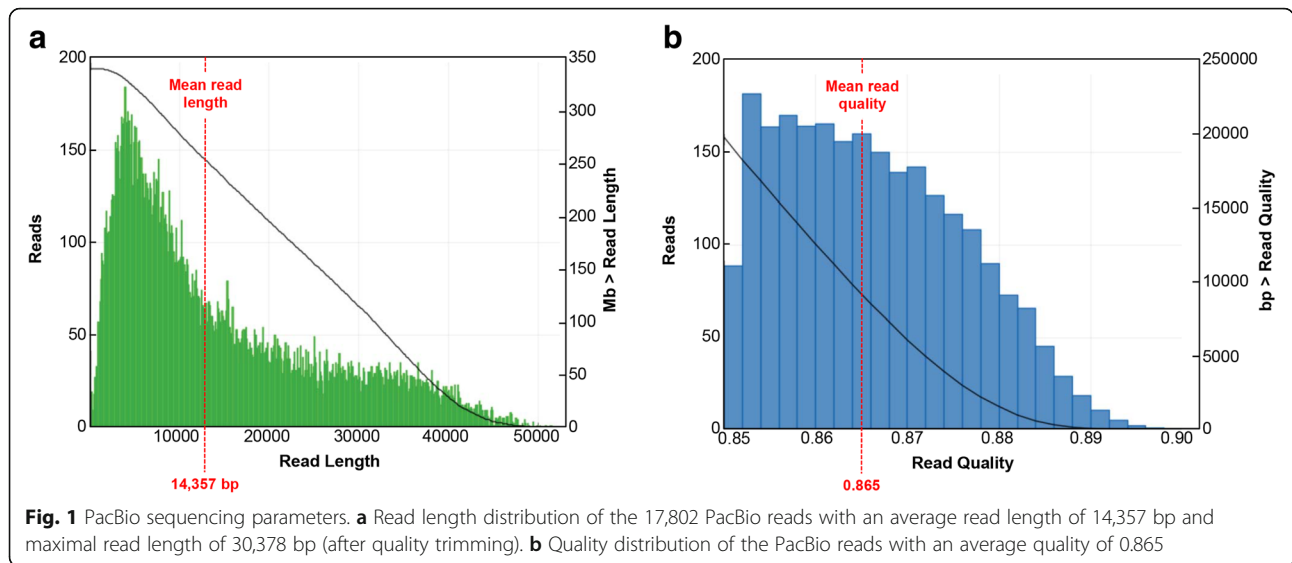
appropriate horseradish peroxidase-labeled secondary antibodies (Dianova, Hamburg, Germany) were applied.

Results and discussion

The *Ifi200* gene cluster developed as a consequence of gene duplications and rearrangements resulting in a divergence in the number of genes between various inbred strains of mice and in repetitive sequences even in coding regions between the different gene members. In order to clarify the genomic alteration responsible for the *Ifi202b* deficiency in the B6 mouse we used the PacBio system, single-molecule real-time (SMRT) sequencing approach, for *de novo* assembling of the critical region in the NZO strain.

For the screening of the NZO BAC clones containing the relevant *Ifi202b* upstream sequence a probe matching a unique *Ifi202b* sequence was used. Additionally a probe specific for the *Olf432* gene was chosen to define the distal border of the region of interest; in contrast to the genomic *Ifi200* region the *Olf432* gene represents a unique sequence within the mouse genome. In total, sequencing of the NZO BAC clones mapped 17,802 PacBio RS reads with a mean read length of 14,357 kb (maximal read length 30,378 kb) and a mean read quality of 0.865. *De novo* assembly of the reads resulted in 4 contigs. However, two of them were not considered for further analysis (unitig2: 35 kb, mean coverage 24 and unitig3: 38 kb, mean coverage 26) due to poor sequence quality. With the two remaining contigs (unitig1: 36.5 kb, mean coverage 365 and unitig0: 300 kb, mean coverage 603; Fig. 1a and b) it was possible to assemble a region covering 6 genes that belongs to the *Ifi200* gene family and the olfactory receptor *Olf433* as the distal boundary (Fig. 2b, upper panel). As described earlier the NZO strain carries two copies of the *Ifi202b* gene which differ in only 8 bp within the coding region, respectively 7 amino acids [6]. In addition, sequence analysis of the BAC identified two copies of other family members; *Ifi205* and *Ifi203*. Interestingly, by comparing the assembled NZO sequence with the B6 reference genome we identified a 261,797 bp deletion affecting the *Ifi200* locus in respect to gene duplications.

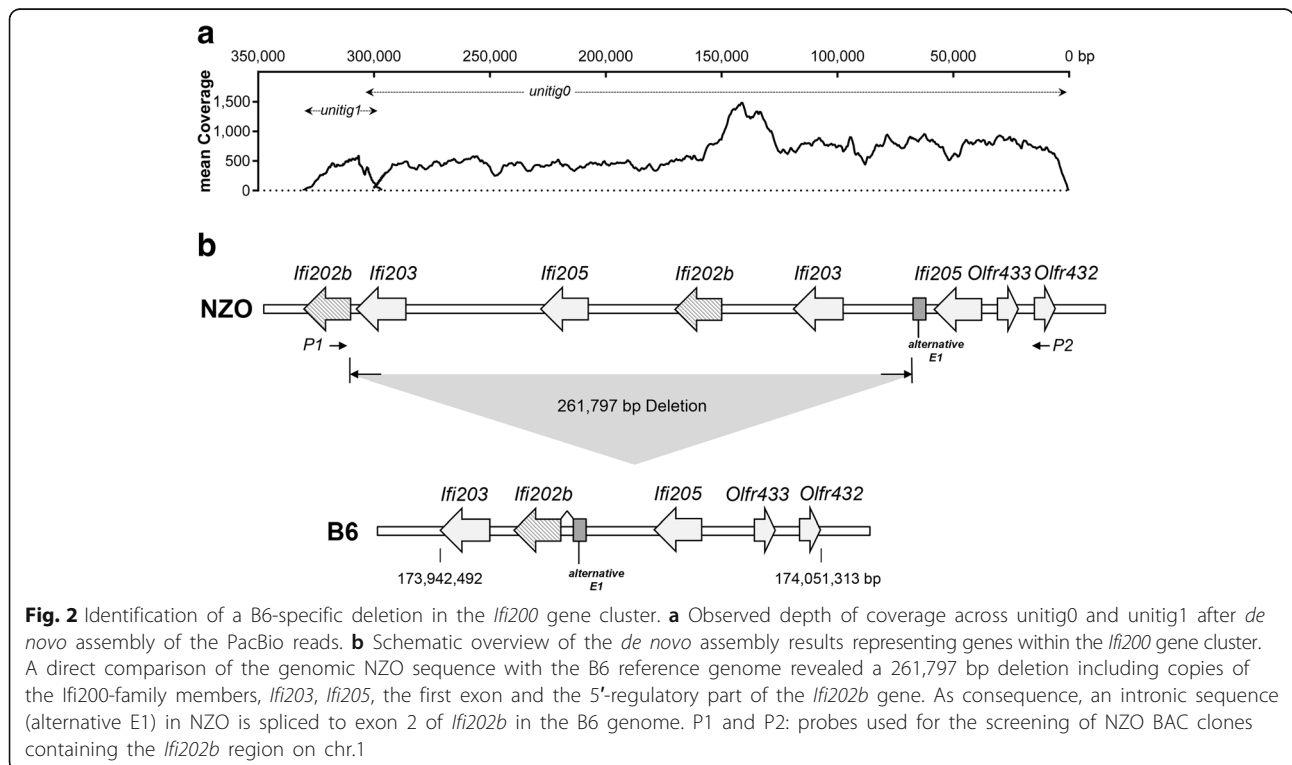
With a second-generation sequencing (SGS) approach it would have been impossible to solve the organization of the *Ifi200* cluster in NZO as sequences are mapped to the B6 reference genome and gaps within the reference genome will result in an incorrect alignment [8]. While the SGS approach is efficient for accurately identifying SNPs in the genome, it does not enable a thorough characterization of structural variations such as insertions and deletions [9–11]. The short sequence read data has complicated the assembly of repetitive structures leading to the translation into gaps, missing data and more incomplete assembly [12–14]. In contrast, the



main advantage of TGS is the long read nature, which was reported to be as long as 3,000 bp on average, and some reads are supposed to be 20,000 bp or even longer. The long read length provides an important benefit for *de novo* assemblies, it allows the discovery of large structural variants, and it provides accurate microsatellite lengths, detection of sensitive SNPs, and haplotype blocks [8, 15–18]. TGS has successfully been used for *de novo* assembling of hundreds of microbial genomes and reconstruction of plant and animal genomes [18–23]. It

has also been applied to resequencing analysis, to create detailed maps of structural variations and phasing variants across large regions of human chromosomes [23–25].

The evolutionary analysis revealed a remarkable plasticity in the mammalian *Ifi200* genes, suggesting the existence of strong evolutionary pressures that have shaped the *Ifi200* sequences and functions throughout the mammalian lineage [26]. Here, we report the identification of structural variations within the *Ifi200* (*PYHIN*) gene



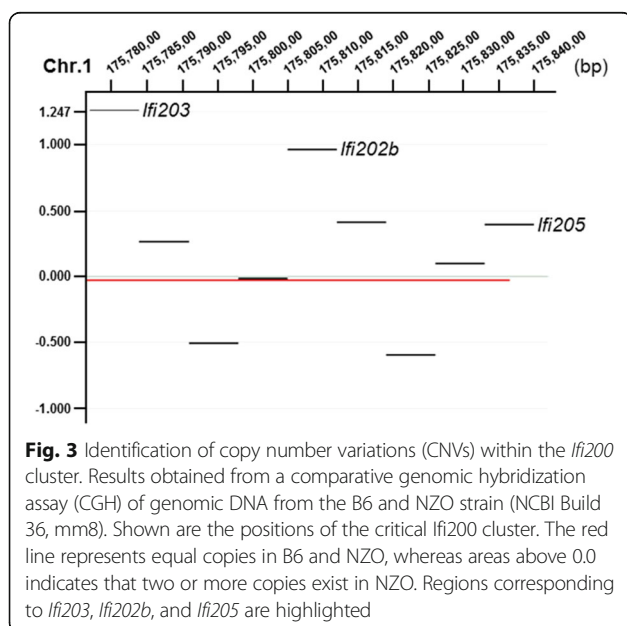
cluster in the obese NZO strain. Cridland and colleagues presented a map comparing the human, C57BL/6 mouse, and rat *Ifi200* gene loci. The mouse contains at least 14 mouse *Ifi200* genes, whereas the human and rat genome expresses only 4, respectively 5 [5]. It was already published that the *Ifi200* gene locus is divergent between various mouse strains as the number of genes present at the locus and the sequence is different [5, 6]. The number of predicted mouse genes has increased with each new update of the mouse genome database and in the current study with *de novo* assembling of the PacBio sequencing reads we can strengthen and expand this assumption to the obese NZO strain [5]. The NZO strain carries two copies of *Ifi202b* (*Ifi202a* and *b*) which was also found in the 129X1/SvJ mouse genome in addition to a pseudogene (*Ifi202c*), whereas only one truncated copy is present in C57BL/6 that is not expressed in metabolically relevant tissues [6, 27, 28]. Another family member, *Ifi203*, showed two extra copies in NZO in comparison to B6. Also the *Ifi205* gene was duplicated as two regions, spanning the coding sequence of the gene, could be mapped in the NZO BAC clones (Fig. 2b). To further verify the sequencing results we performed a comparative genomic hybridization assay (CGH) of genomic DNA obtained from the B6 and the NZO strain to detect copy number variations (CNVs) within the cluster. This analysis further supports that the NZO strain carries at least two copies of the genes *Ifi202b*, *Ifi203*, and *Ifi205* (Fig. 3). Other studies also show the presence of gene duplications. She and colleagues (2008) assessed CNVs between the B6 strain and

15 mouse strains (including NZO) which were used for genetic association studies, sequencing, and the Mouse Phenome Project [29]. The analysis also showed a duplication of the *Ifi203* gene. Similar results were detected for *Ifi205* in the study by Cahan et al., 2009 where CNVs in 17 mouse strains were analyzed [30]. In conclusion, *de novo* assembling of the NZO BAC clone reads and the analysis of CNVs revealed structural variations between different inbred strains of mice within a complex region on chr. 1 caused by duplications and genomic alterations.

It is also documented that the corresponding region in humans is affected by genomic alterations. According to the 1000 Genomes project several deletions, CNV, and duplications can be mapped within this locus [31]. Cagliani and colleagues performed an evolutionary analysis of the human family members (*MNDA*, *PYHIN1*, *IFI16*, and *AIM2*) by analyzing inter- and intraspecies diversity and revealed that the genes have been repeatedly targeted by natural selection. Especially the *IFI16* gene region shows a high nucleotide diversity in human populations and indicates that the region has been a target of long-standing balancing selection [32].

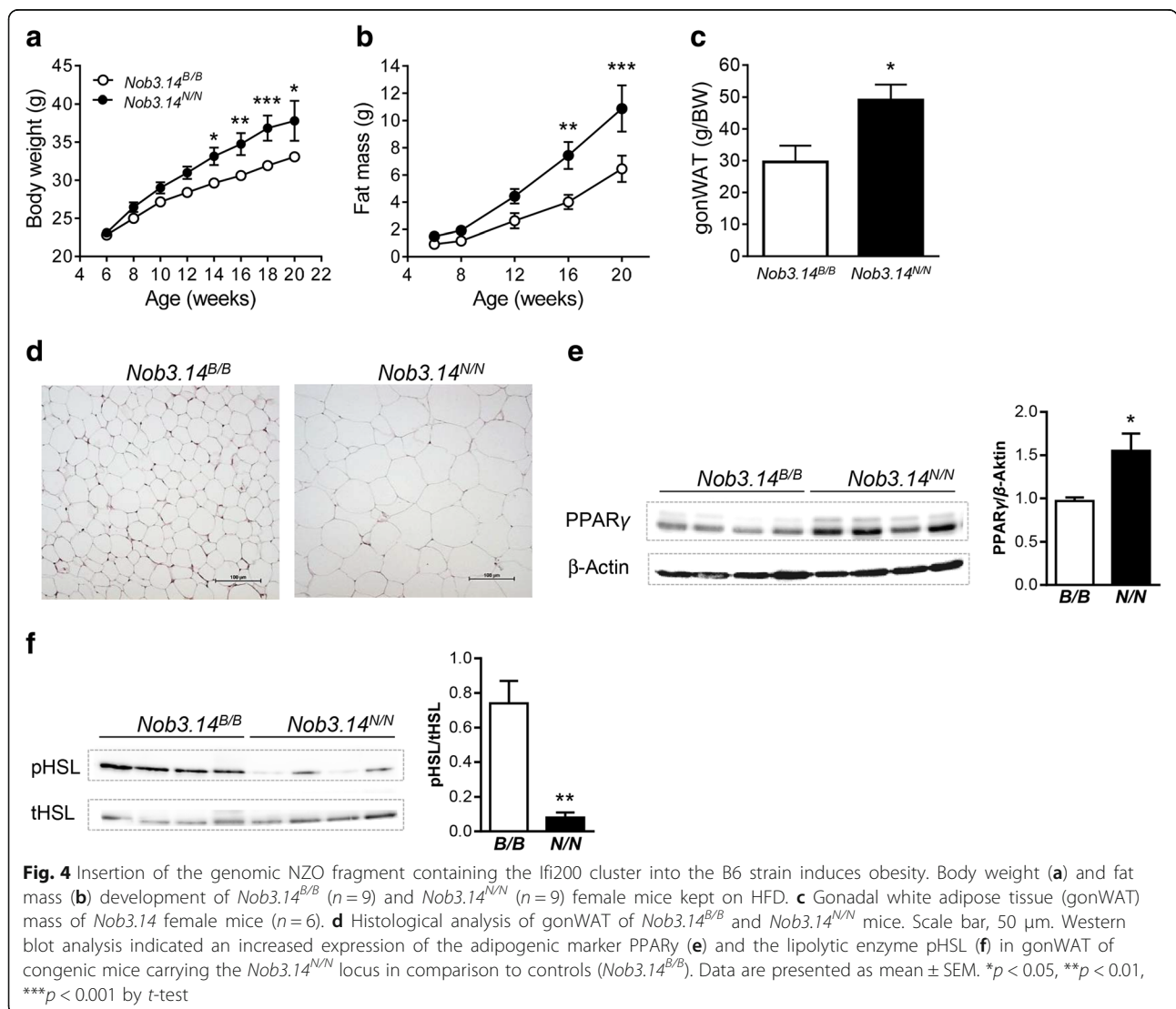
The main goal of the current study was to analyze the chromosomal alterations leading to the *Ifi202b* deficiency in the B6 strain. With the BAC sequencing we identified a deletion spanning approximately 261.8 kb within the B6 genome, a sequence present in NZO. The deletion includes different copies of *Ifi200*-family members, *Ifi203*, *Ifi205*, and exon 1 of *Ifi202b* (Fig. 2b). In our previous study we identified an alternative first exon in the B6 reference genome (Vogel et al., 2012). With the current study we are finally able to define the exact chromosomal region deleted in B6 and we can explain how this alternative exon 1 - which is an intronic sequence in NZO - is spliced to exon 2 of *Ifi202b* in the B6 genome (Fig. 2b, lower panel). The fact that B6 do not express *Ifi202b* in the same tissues (e.g. adipose tissue, liver, and skeletal muscle) as NZO indicates that in addition to the first exon also the promoter or at least part of it was deleted as well.

It is also reasonable to assume that the deleted region in B6 contains enhancer motifs/long-range control elements that drive and regulate the expression of other genes. In a previous study we reported that the genes *Lefty1*, *Pcp4l1*, and *Apoa2*, located in the same diabetes susceptibility locus as *Ifi202b* (*Nob3*), are exclusively present in islets of the diabetes-resistant B6 strain in contrast to the diabetes-prone NZO mouse. The identified genes are furthermore involved in the adaptive islet hyperplasia and prevention from severe diabetes in B6-*ob/ob* mice [33]. With the hereby reported data we hypothesize that the genomic alterations within the cluster may also include enhancer elements that carry the



potential to regulate the expression of *Lefty1*, *Pcp4l1*, and *Apoa2*. By using the *Nsite* program, a computer tool to search for regulatory elements (REs), we found 5 predictive enhancer motifs that are located within the deleted sequence in the B6 genome which can potentially be responsible for the described expression differences. A number of long-range regulatory disruptions affecting the expression of genes have already been described [34, 35]. One of the oldest examples of a human gene in which long-range regulations has been implicated and studied is *SOX9*, a gene responsible for autosomal sex reversal and Campomelic Dysplasia (CD). All rearrangements including deletions are found from 50 kb to 950 kb upstream of *SOX9* suggesting that a similar mechanism could also account for the expression differences between the diabetes-prone NZO and diabetes-resistant B6 strain of genes located within the *Nob3* locus [34, 35].

Finally, to elucidate whether the genomic alteration on chr. 1 is also associated with metabolic alterations we generated and characterized congenic mice carrying 14.2 Mbp (163.5-177.7 Mbp) of the NZO genome (*Nob3.14^{N/N}*), including the *Iffi200* gene cluster, on B6 background. On HFD, homozygous NZO allele carriers developed a higher body weight and fat mass (Fig. 4a and b), in particular gonadal white adipose tissue (gonWAT, Fig. 4c), than the corresponding controls (*Nob3.14^{B/B}*). Histological analysis of the gonWAT demonstrated that the adipocytes were larger in the *Nob3.14^{N/N}* group than those of *Nob3.14^{B/B}* mice (Fig. 4d). As these data points towards a role of the cluster in adipose tissue biology we tested the expression of proteins involved in adipocyte differentiation and lipolysis. Western blot analysis indicated an increased expression of the adipogenic marker PPAR γ (Peroxisome proliferator-activated receptor gamma)



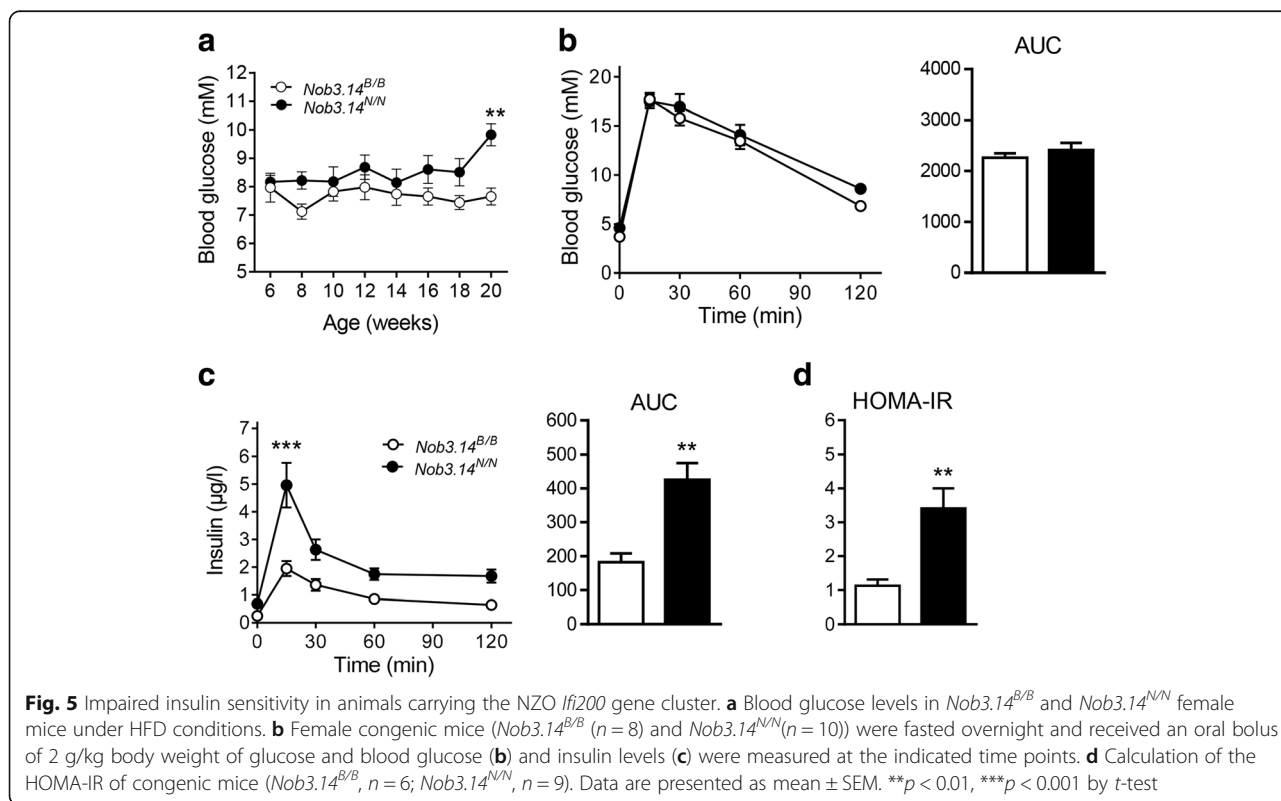
and a decreased activation of the lipolytic enzyme HSL (Hormone sensitive lipase) in gonWAT of NZO allele carriers in comparison to controls (Fig. 4e and f). As obesity and hypertrophy of adipose tissue are also known to impair insulin sensitivity and glucose tolerance, we measured the glucose levels of the congenic lines. Blood glucose levels were measured randomly and started to differ at the age of 20 weeks between the two groups with higher concentrations in NZO allele carriers (Fig. 5a). Glucose clearance during oral glucose tolerance tests was not different between the two genotypes (Fig. 5b). However, the *Nob3.14^{N/N}* mice required higher levels of insulin than *Nob3.14^{B/B}* mice to clear blood glucose, pointing towards an insulin resistance (Fig. 5c) which is also indicated by calculating the HOMA-IR (Fig. 5d). In conclusion, introducing the genomic region of the *Ifi200* gene cluster of the NZO genome into the B6 genome results in the development of obesity and is associated with insulin resistance which demonstrates the functional consequences of the alteration on chr.1.

In different reports it was already published that rare GSVs are associated with obesity [36]. A rare (0.7%), 593 kb deletion on chromosome 16p11.2 (at 29.5–30.1 Mbp) was shown to be significantly ($p = 6.4 \times 10^{-8}$) enriched in obese patients compared to controls, whereas a

duplication of the same locus has the opposite effect, being associated with underweight [1, 37, 38]. Another study by Wang et al. [39] also showed large and rare CNVs that are associated with a higher risk to develop obesity. They reported several CNVs that affect known candidate genes for obesity, such as a 3.3-Mbp deletion disrupting *NAP1L5* and a 2.1-Mbp deletion disrupting *UCP1* and *IL15*. One prominent example for chromosomal syndromes with obesity is the Prader-Willi syndrome (PWS) in which a 5–7 Mb deletion of the paternally inherited chromosomal 15q11.2-q13 region is responsible for a neurobehavioral disorder manifested by infantile hypotonia and feeding difficulties in infancy, followed by morbid obesity secondary to hyperphagia [40].

Conclusions

In summary, by using TGS it was possible to assemble a complex genomic region on mouse chr. 1 containing different genes of the *Ifi200* cluster. This approach further leads to the identification of a vast chromosomal deletion including the regulatory part of the obesity-associated gene *Ifi202b*, as well as one copy of *Ifi203* and one of *Ifi205* in the B6 strain which finally leads to an altered expression and consequently affecting the susceptibility to develop obesity.



Abbreviations

AIM2: Absent in melanoma 2; *Apoa2*: Apolipoprotein A-II; *B/B*: B6/B6; B6: C57BL/6; *BAC*: Bacterial artificial chromosome; *Cadm3*: Cell adhesion molecule 3; *CGH*: Comparative genomic hybridization; *chr.*: Chromosome; *CNVs*: Copy number variations; *Cy*: Cyanine; *DIG*: Digoxigenin; *E.coli*: *Escherichia coli*; *E1*: Exon 1; *EDTA*: Ethylenediaminetetraacetic acid; *ELISA*: Enzyme linked immunosorbent assay; *for*: Forward; *gDNA*: Genomic DNA; *gonWAT*: Gonadal white adipose tissue; *GSVs*: Genomic structural variants; *GWAs*: Genome-wide association studies; *HFD*: High-fat diet; *HGAP*: Hierarchical genome-assembly process; *HMW*: High molecular weight; *HOMA-IR*: Homeostasis model assessment of insulin resistance; *HSL*: Hormone sensitive lipase; *Irf16*: Interferon gamma inducible protein 16; *Irf200*: Interferon inducible gene 200 family; *Irf202b*: Interferon inducible gene 202b; *Irf203*: Interferon inducible gene 203; *Irf205*: Interferon inducible gene 205; *IL15*: Interleukin 15; *LB*: Lysogeny broth; *Lefty1*: Left right determination factor 1; *MNDA*: Myeloid cell nuclear differentiation antigen; *N/N*: NZO/NZO; *NAP1L5*: Nucleosome assembly protein 1 like 5; *Nob3*: NZO obesity 3; *NZO*: New Zealand obese; *OGTT*: Oral glucose tolerance test; *Olf432*: Olfactory receptor 432; *Olf433*: Olfactory receptor 433; *PacBio*: Pacific biosciences; *Pcp4l1*: Purkinje cell protein 4-like 1; *PPARγ*: Peroxisome proliferator-activated receptor gamma; *PVDF*: Polyvinylidene difluoride; *PWS*: Prader-Willi syndrome; *PYHIN*: Pyrin and HIN domain-containing protein; *P1/2*: Primer 1/2; *QTL*: Quantitative trait locus; *rev*: Reverse; *SDS-PAGE*: Sodium dodecyl sulfate polyacrylamide gel electrophoresis; *SGS*: Second-generation sequencing; *SMRT*: Single-molecule real-time sequencing; *SNP*: Single nucleotide polymorphism; *SOX9*: *SRY* (sex determining region Y)-box 9; *TGS*: Third-generation sequencing; *UCP1*: Uncoupling protein 1

Acknowledgements

We gratefully thank Malte Neubauer for technical assistance. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

Funding

The study was supported by grants from the German Ministry of Education and Research (DZD, 01GI0922 and 01GI0925; NEUROTARGET: 01GI0847).

Availability of data and materials

Sequence data that support the findings of this study have been deposited in GenBank with the accession numbers KX668626 and KX668627.

Authors' contributions

HV and AS conceived and designed the experiments. HV, MJ, DM, MS, and SS carried out all experiments. HV, MJ, DM, MS, SS, and AS analyzed data. HV and AS wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval

All animal experiments were approved by the ethics committee of the State Office of Environment, Health and Consumer Protection (Federal State of Brandenburg, Germany).

Author details

¹Department of Experimental Diabetology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert Allee 114-116, D-14558 Nuthetal, Germany. ²German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, 85764 München-Neuherberg, Germany. ³Institute of Pharmacology and Toxicology, University of Braunschweig, Mendelssohnstr. 1, 38106 Braunschweig, Germany.

Received: 22 December 2016 Accepted: 3 February 2017

Published online: 15 February 2017

References

- Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczak K, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 2010;463:666–70.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–53.
- McClellan J, King M-C. Genetic heterogeneity in human disease. *Cell*. 2010;141:210–7.
- Gorlov IP, Gorlova OY, Frazier ML, Spitz MR, Amos CI. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet*. 2011;79:199–206.
- Cridland JA, Curley EZ, Wykes MN, Schroder K, Sweet MJ, Roberts TL, et al. The mammalian PYHIN gene family: phylogeny, evolution and expression. *BMC Evol Biol*. 2012;12:140.
- Vogel H, Scherneck S, Kanzleiter T, Benz V, Kluge R, Stadion M, et al. Loss of function of *Irf202b* by a microdeletion on chromosome 1 of C57BL/6J mice suppresses 11 β -hydroxysteroid dehydrogenase type 1 expression and development of obesity. *Hum Mol Genet*. 2012;21:3845–57.
- Vogel H, Montag D, Kanzleiter T, Jonas W, Matzke D, Scherneck S, et al. An Interval of the Obesity QTL *Nob3.38* within a QTL Hotspot on Chromosome 1 Modulates Behavioral Phenotypes. *PLoS One*. 2013;8:e53025.
- Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11:31–46.
- Bentley DR. Whole-genome re-sequencing. *Curr Opin Genet Dev*. 2006;16:545–52.
- Wang JJ, Wang W, Li R, Li Y, Tian G, Fan W, et al. The diploid genome sequence of an Asian individual. *Nature*. 2008;456:60–5.
- Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet*. 2010;19:R227–40.
- Alkan C, Cardone MF, Catacchio CR, Antonacci F, O'Brien SJ, Ryder OA, et al. Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Res*. 2011;21:137–45.
- Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, et al. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Res*. 2012;22:557–67.
- Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, et al. Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res*. 2014;24:688–96.
- Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol*. 2013;14:405.
- Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*. 2012;13:375.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Chen X, Li Q, Li Y, Qian J, Han J. Chloroplast genome of *Aconitum barbatum* var. *puberulum* (Ranunculaceae) derived from CCS reads using the PacBio RS platform. *Front Plant Sci*. 2015;6:42.
- Koren S, Harhay GP, Smith TPL, Bono JL, Harhay DM, Mcvey SD, et al. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol*. 2013;14:R101.
- Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods*. 2015;12:733–5.
- Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol*. 2015;33:623–30.
- Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, et al. Long-read sequence assembly of the gorilla genome. *Science*. 2016;352:aae0344.
- Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, et al. Third-generation sequencing and the future of genomics. *bioRxiv*. 2016;Table 1, p. 048603.
- Chaisson MJP, Wilson RK, Eichler EE. Genetic variation and the de novo assembly of human genomes. *Nat Rev Genet*. 2015;16:627–40.
- Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, et al. Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol*. 2014;32:261–6.

26. Brunette RL, Young JM, Whitley DG, Brodsky IE, Malik HS, Stetson DB. Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. *J Exp Med*. 2012;209:1969–83.
27. Deschamps S, Meyer J, Chatterjee G, Wang H, Lengyel P, Roe BA. The mouse Ifi200 gene cluster: Genomic sequence, analysis, and comparison with the human HIN-200 gene cluster. *Genomics*. 2003;82:34–46.
28. Wang H, Chatterjee G, Meyer JJ, Liu CJ, Manjunath NA, Bray-Ward P, et al. Characteristics of three homologous 202 genes (Ifi202a, Ifi202b, and Ifi202c) from the murine interferon-activatable gene 200 cluster. *Genomics*. 1999;60:281–94.
29. She X, Cheng Z, Zöllner S, Church DM, Eichler EE. Mouse segmental duplication and copy number variation. *Nat Genet*. 2008;40:909–14.
30. Cahan P, Li Y, Izumi M, Graubert TA. The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nat Genet*. 2009;41:430–7.
31. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
32. Cagliani R, Forni D, Biasin M, Comabella M, Guerini FR, Riva S, et al. Ancient and recent selective pressures shaped genetic diversity at AIM2-like nucleic acid sensors. *Genome Biol Evol*. 2014;6:830–45.
33. Kluth O, Matzke D, Kamitz A, Jähnert M, Vogel H, Scherneck S, et al. Identification of Four Mouse Diabetes Candidate Genes Altering β -Cell Proliferation. *PLoS Genet*. 2015;11:e1005506.
34. Pfeifer D, Kist R, Dewar K, Devon K, Lander ES, Birren B, et al. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *Am J Hum Genet*. 1999;65:111–24.
35. Kleinjan DA, Van Heyningen V. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am J Hum Genet*. 2005; 76:8–32.
36. Walters RG, Coin LJM, Ruukonen A, de Smith AJ, El-Sayed Moustafa JS, Jacquemont S, et al. Rare Genomic Structural Variants in Complex Disease: Lessons from the Replication of Associations with Obesity. *PLoS One*. 2013;8: e58048.
37. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010;463:671–5.
38. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*. 2011;478:97–102.
39. Wang K, Li WD, Glessner JT, Grant SFA, Hakonarson H, Price RA. Large copy-number variations are enriched in cases with moderate to extreme obesity. *Diabetes*. 2010;59:2690–4.
40. Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet*. 2010;18:1196–201.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

