

Common nonsynonymous variants in *PCSK1* confer risk of obesity

Michael Benzinou^{1,2}, John W M Creemers³, Helene Choquet², Stephane Lobbens², Christian Dina², Emmanuelle Durand², Audrey Guerardel², Philippe Boutin², Beatrice Jouret⁴, Barbara Heude⁵, Beverley Balkau⁵, Jean Tichet⁶, Michel Marre⁷⁻⁹, Natascha Potoczna⁶, Fritz Horber¹⁰, Catherine Le Stunff¹¹, Sebastien Czernichow¹², Anelli Sandbaek¹³, Torsten Lauritzen¹³, Knut Borch-Johnsen^{14,15}, Gitte Andersen¹⁴, Wieland Kiess¹⁶, Antje Körner¹⁶, Peter Kovacs¹⁷, Peter Jacobson¹⁸, Lena M S Carlsson¹⁸, Andrew J Walley¹, Torben Jørgensen¹⁹, Torben Hansen¹⁴, Oluf Pedersen^{14,15}, David Meyre² & Philippe Froguel^{1,2}

Mutations in *PCSK1* cause monogenic obesity. To assess the contribution of *PCSK1* to polygenic obesity risk, we genotyped tag SNPs in a total of 13,659 individuals of European ancestry from eight independent case-control or family-based cohorts. The nonsynonymous variants rs6232, encoding N221D, and rs6234-rs6235, encoding the Q665E-S690T pair, were consistently associated with obesity in adults and children ($P = 7.27 \times 10^{-8}$ and $P = 2.31 \times 10^{-12}$, respectively). Functional analysis showed a significant impairment of the N221D-mutant PC1/3 protein catalytic activity.

Four independent genome-wide linkage studies, two of which were done in independent sets of individuals of French European ancestry, delineate a common 5.6-Mb interval on chromosome 5q linked with obesity-associated traits¹⁻⁴. This interval contains the *PCSK1* (prohormone convertase 1/3) gene (Supplementary Table 1 online), which encodes an enzyme expressed in neuroendocrine cells that converts prohormones into functional key hormones that regulate central and/or peripheral energy metabolism. Mutations in *PCSK1* have been found to lead to human congenital PC1/3 deficiency, a syndrome characterized by obesity and small intestinal dysfunction⁵⁻⁷. With the hypothesis that frequent (minor allele frequency (MAF) >

0.05) *PCSK1* SNPs could be involved in the pathogenesis of common obesity, we used a staged approach involving the testing of 13,659 participants of European ancestry. Initially, the *PCSK1* coding regions, plus 3.5 kb extending both 5' and 3', were sequenced in 48 unrelated obese adults that contributed to the chromosome 5q obesity linkage^{1,3}. Of 19 frequent SNPs identified (MAF > 0.05), 8 tagging SNPs (tSNPs; r^2 threshold = 0.8) were selected, together with the nonsynonymous SNP rs6234, encoding Q665E, which is highly correlated with the nonsynonymous tSNP rs6235, encoding S690T ($r^2 > 0.96$; Fig. 1a and Supplementary Tables 2 and 3 online). Together, these SNPs tagged 92% of the common variation with a mean r^2 of 0.98 in the 103.7 kb of haplotype blocks that incorporate *PCSK1* (Supplementary Fig. 1 online).

We genotyped these nine SNPs in 1,045 obese adults (BMI > 30 kg/m²) and 1,265 non-obese controls (Supplementary Methods and Supplementary Table 4 online), all of French European ancestry. In a logistic regression test that adjusted for age and gender, six SNPs showed significant association with obesity (P values ranged from 0.045 to 0.00004; Supplementary Table 5 online). Because they had the highest genetic evidence for association with obesity and a putative functional detrimental role, we chose to further examine rs6232 in *PCSK1* exon 6, encoding N221D, and rs6234-rs6235 in exon 14, encoding Q665E-S690T, in a further two independent adult case-control studies. rs6235 was selected over rs6234 for genotyping purposes because it had both a higher phastCons conservation score and a higher genotyping success rate (Supplementary Table 5). The comparison of 3,074 obese Danish adults with 2,790 non-obese Danish controls (Supplementary Methods and Supplementary Table 4) confirmed the association with obesity for rs6232 (OR = 1.17 (1.01–1.36), $P = 0.019$) and rs6235 (OR = 1.12 (1.03–1.21), $P = 0.005$) (Table 1). Similarly, the comparison of 551 class III obese Swiss individuals (BMI > 40 kg/m²) to 542 randomly selected blood donors (Supplementary Methods and Supplementary Table 4) further demonstrated association of the rs6235[C] at-risk allele and of the rs6232[G] at-risk allele with obesity (OR = 1.38 (1.15–1.67), $P = 0.00035$; OR = 1.38 (0.95–2.01), $P = 0.045$; respectively) (Table 1).

We then assessed the contribution of these *PCSK1* nonsynonymous variants to childhood obesity risk in a further three case-control studies. Allele frequencies of the selected SNPs were compared

¹Genomic Medicine, Imperial College London, Hammersmith Hospital, London W120NN, UK. ²Centre National de la Recherche Scientifique (CNRS) 8090-Institute of Biology, Pasteur Institute, Lille 59000, France. ³Department of Human Genetics, University of Leuven, Leuven B-3000, Belgium. ⁴Institut National de la Santé et de la Recherche Médicale (INSERM) U563, Children's Hospital, Toulouse 31000, France. ⁵INSERM, U780-IFR69, Villejuif, France and Université Paris Sud, Faculty of Medicine, Le Kremlin-Bicêtre 94270, France. ⁶The Regional Institut for Health, Tours 37000, France. ⁷INSERM U695, Paris 75018, France. ⁸Université Paris Diderot - Paris 7, Paris 75018, France. ⁹Department of Endocrinology-Diabetology and Nutrition, Bichat Claude Bernard Hospital, Paris 75870, France. ¹⁰Klinik Lindberg, Winterthur 8400, Switzerland. ¹¹Department of Pediatric Endocrinology and INSERM U561, Saint Vincent de Paul Hospital, René Descartes University, Paris 75014, France. ¹²INSERM U557, Institut Scientifique de Recherche Agronomique (INRA) U1125, CNAM EA3200, Université Paris 13, CRNH IdF and Hôpital Avicenne (AP-HP), Bobigny F-93017, France. ¹³Department of General Practice, University of Aarhus, Aarhus DK-8000, Denmark. ¹⁴Steno Diabetes Center, DK-2820 Gentofte, Copenhagen, Denmark. ¹⁵Faculty of Health Science, University of Aarhus, Aarhus DK-8000, Denmark. ¹⁶University Hospital for Children and Adolescents, University of Leipzig, Leipzig D-04109, Germany. ¹⁷Department of Internal Medicine III, Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig D-04109, Germany. ¹⁸Department of Molecular and Clinical Medicine Institute of Medicine, The Sahlgrenska Academy, Göteborg University, Göteborg SE-431, Sweden. ¹⁹Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup DK-2600, Denmark. Correspondence should be addressed to P.F. (p.froguel@imperial.ac.uk).

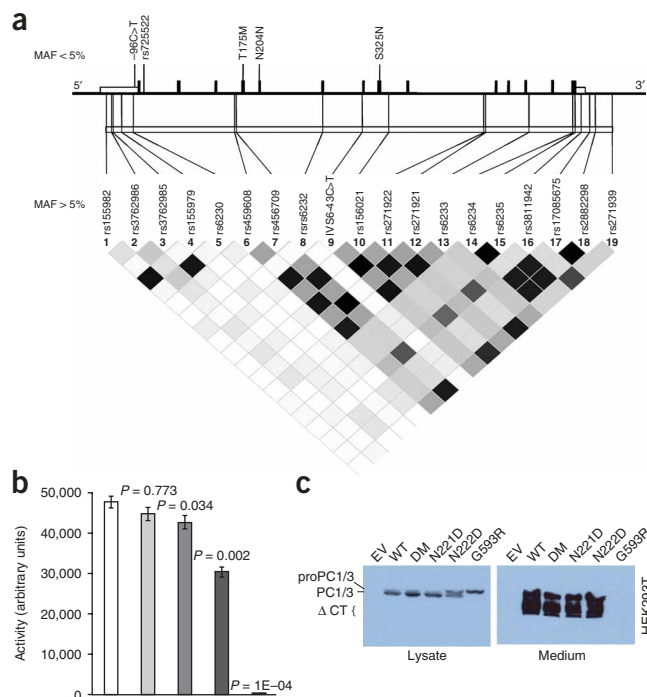


Figure 1 Linkage disequilibrium map of the *PCSK1* genomic region and functional characterization of the N221D and Q665E-S690T substitutions. **(a)** Linkage disequilibrium map (r^2 measure) of the 19 frequent polymorphisms of *PCSK1* identified in 48 obese subjects produced using Haploview (shading is used to indicate LD between pairs of SNPs: darker shades indicate stronger LD). **(b)** Functional characterization of the N221D and Q665E-S690T substitutions. HEK293T cells were transfected with empty vector (EV), wild-type PC1/3, Q665E-S690T (DM), N221D, N222D or G593R. Activity of recombinant PC1/3 proteins immunopurified from transfected HEK293T cells using the fluorogenic substrate p-Glu-Arg-Thr-Lys-Arg-amino methylcoumarin and normalized for expression levels. The diagrams represent means of quadruplicate measurements. Data shown are for one representative of three independent experiments. We confirmed the lack of catalytic activity for the recombinant G593R protein (98.7% decrease; $P = 10^{-4}$) and a 36.3% decrease in activity of the N222D protein. The cluster Q665E-S690T induced a nonsignificant reduction of the activity enzymatic of PC1/3 (6.1% decrease; $P = 0.773$). The N221D amino acid substitution induced a 10.4% significant reduction of the activity ($P = 0.034$) in comparison with the wild-type PC1/3. **(c)** Protein blot of cell lysates and medium of transfected HEK293T cells using Flag M2 for detection of recombinant PC1/3 proteins. No marked difference in maturation and secretion was observed for N221D and for Q665E-S690T when compared to WT, neither in lysate nor in medium. ΔCT, PC1/3 truncated at the C terminus.

between 1,010 non-obese French individuals and 580 unrelated French obese children (**Supplementary Methods** and **Supplementary Table 4**) and replicated the association for both rs6232 and rs6235 (OR = 1.67 (1.21–2.32), $P = 0.0009$ and OR = 1.23 (1.05–1.45), $P = 0.005$, respectively) (**Table 1**). The comparison of 532 non-obese young French adults from the east of France to 505 French obese children (**Supplementary Methods** and **Supplementary Table 4**) showed similar trends of association (rs6232, OR = 1.57 (1.07–2.31), $P = 0.009$ and rs6235, OR = 1.50 (1.23–1.83), $P = 0.00003$, respectively) (**Table 1**). Further testing of these SNPs in 715 non-obese children and 283 overweight or obese children from eastern Germany (**Supplementary Methods** and **Supplementary Table 4**) demonstrated that rs6232[G] was associated with obesity (OR = 1.56 (1.05–2.32), $P = 0.013$) whereas no significant association was found for rs6235 (**Table 1**).

We obtained further confirmation of the association of rs6232 and rs6235 with obesity using a cohort of 154 families of Swedish descent, discordant for severe obesity (with at least one class III obese sib and one lean sib; **Supplementary Methods**). We observed significant overtransmission of the obesity at-risk alleles to obese offspring: 60.6% transmitted for rs6232[G] ($P = 0.037$) and 59% transmitted for rs6235[C] ($P = 0.027$) (**Table 1**).

Additionally, we tested the 5,195 subjects of the 9-year followed-up French general middle-aged D.E.S.I.R. cohort⁸. The limited number of obese incident cases (N) carrying at-risk alleles in this cohort ($N_{rs6232} = 15$; $N_{rs6235} = 77$) only allowed the study of the predictive effect of the most frequent variant, rs6235, on obesity. Survival analysis identified an increased incidence of class II obesity associated with the rs6235 at-risk allele under an additive model, during the 9-year follow-up (Cox model adjusted for gender, hazard ratio = 1.47 (1.10–1.97), $P = 0.009$).

To provide an estimate of the global odds ratio despite the differences in ascertainment strategies of our studies, we used the Mantel-Haenszel method on the seven independent OR estimates

described above, excluding the D.E.S.I.R. cohort because of its correlation with the first French control set. We found that rs6232 and rs6235 exceeded genome-wide levels of significance ($< 5 \times 10^{-7}$) with combined OR of 1.34 (1.20–1.49) ($P = 7.27 \times 10^{-8}$) and 1.22 (1.15–1.29) ($P = 2.31 \times 10^{-12}$), respectively (**Table 1**). No significant heterogeneity among studies was identified ($P = 0.32$ and $P = 0.053$ for rs6232 and rs6235, respectively). The population attributable risk (PAR) to develop obesity due to the rs6232[G] at-risk allele and rs6235[C] at-risk allele and calculated from the case-control studies was estimated at 1.9% (1.1–2.7) and 5.7% (3.9–7.3), respectively. The most likely inheritance model was found to be additive with no deviation from linearity ($P = 0.16$). To test whether rs6232 and rs6234-rs6235 had independent effects on the risk of obesity, we carried out a two-SNP analysis in our obese and control cohorts. When we excluded the pair effect of rs6235-rs6234 from the model, we found an independent significant effect of rs6232 ($P = 2.7 \times 10^{-6}$). Similarly, when we excluded the rs6232 variant effect, we found a significant effect of the rs6235-rs6234 pair on the risk of obesity ($P = 0.002$; **Supplementary Table 6** online). To test for the contribution of *PCSK1* to the chromosome 5q linkage, we ran a GIST procedure in the 109 pedigrees that showed linkage for obesity in this region³. Neither rs6232 ($P = 0.54$) nor the rs6234-rs6235 pair ($P = 0.51$) were found to significantly contribute to the 5q15 linkage.

The S690T and Q665E substitutions are both located in the C-terminal region of the protein. The C terminus of the protein has been shown to be important for correct targeting and specificity of PC1/3 and the sorting in secretory granules⁹. The N221D substitution is located in the catalytic domain near two other polymorphic sites (Glu250stop and Ala213del) that, respectively, truncate the PC1/3 protein and delete a highly conserved alanine residue near the catalytically critical His208 (ref. 6). Of note, N221D has the closest contact with the His208 residue (<http://genetics.bwh.harvard.edu/pph/>). Furthermore, N221D is immediately adjacent to the N222D, a variation of the murine PC1/3 catalytic domain that leads to maturity-onset obesity and increased body fat content in homozygous mutant mice¹⁰. N222D-heterozygous mice also had increased body fat content compared to wild-type mice¹⁰. *In vitro* studies allowed testing for the effect of N221D and Q665E-S690T substitutions on PC1/3

Table 1 Association between rs6232 and rs6235 and risk of obesity: original case-control, replication studies and pooled analysis

	rs6232 (encoding N221D)										rs6235 (encoding S690T)									
	Genotypes (n)					MAF	Allelic P value	Odds ratio estimate (CI)	Genotypes (n)					MAF	Allelic P value	Odds ratio estimate (CI)				
AA	AG	GG	CC	GC	GG				GC	CC	GC	CC	GC				CC	GC	CC	
Independent adult case-control obesity studies																				
French adult obesity	1,159	83	4	3.7	0.0042	1.51 (1.14–2.01)	707	483	68	24.6	0.00004	1.31 (1.15–1.50)								
Obese adults	930	100	6	5.4	0.019*	1.17 (1.01–1.36)	503	439	90	30	0.005*	1.12 (1.03–1.21)								
Controls	2,464	316	7	5.9	0.019*	1.17 (1.01–1.36)	1,476	1,059	193	26.6	0.005*	1.12 (1.03–1.21)								
Danish adult obesity	2,599	381	15	6.9	0.045*	1.38 (1.05–2.01)	1,497	1,174	253	28.7	0.00035*	1.38 (1.15–1.67)								
Obese adults	491	4	1	4.7	0.045*	1.38 (1.05–2.01)	301	202	34	25.1	0.00035*	1.38 (1.15–1.67)								
Controls	461	59	4	6.4			245	238	51	31.8										
Independent case-control studies on childhood obesity																				
French child, obesity 1	869	70	3	4	0.0009*	1.67 (1.21–2.32)	518	382	62	26.3	0.005*	1.23 (1.05–1.46)								
Obese children	498	69	3	6.6	0.009*	1.57 (1.07–2.31)	317	226	59	30.6	0.00003*	1.50 (1.23–1.83)								
Controls	481	45	1	4.5	0.009*	1.57 (1.07–2.31)	277	179	29	22.6	0.12*	1.12 (0.92–1.43)								
French child, obesity 2	421	56	5	6.8	0.013*	1.56 (1.05–2.32)	237	194	49	30.4										
Obese children	646	66	2	4.9			415	252	47	24.2										
Controls	241	40	1	7.4			150	110	20	26.8										
Independent family-based studies																				
Swedish adult obesity	33	20	13	0.037*	1.54 (0.69–3.47)	105	62	43	0.027*	1.44 (0.98–2.11)										
Overall significance of the independent studies (Mantel-Haenszel test on seven independent OR estimates)																				
Combined case-control studies and Swedish family study																				
Informative meioses					Informative meioses					Informative meioses										
T _r					T _r					T _r										
Non T _r					Non T _r					Non T _r										
TDT P value					TDT P value					TDT P value										
Odds ratio estimate (CI), P value					Odds ratio estimate (CI), P value					Odds ratio estimate (CI), P value										
P value = 0.32					P value = 0.052					P value = 0.0052										
1.34 (1.20–1.49), 7.27 × 10 ⁻⁸					1.54 (0.69–3.47)					1.22 (1.15–1.29), 2.31 × 10 ⁻¹²										

Odds ratios and P values are from logistic regression, adjusted for age and gender, in the adult obesity case-control studies. For the childhood obesity case-controls, the tests were only adjusted for gender. *P values are one-sided, as we tested the specific hypothesis of increased frequency of rs6232(G) and rs6235(C) in obese children and adult replication cohorts.

processing and catalytic activity. Similarly to a previous report⁷, the Q665E-S690T cluster was not found to significantly alter PC1/3 enzymatic activity (Fig. 1b). Conversely, N221D induced a 10.4% significant reduction of activity ($P = 0.03$) when compared to the wild-type PC1/3 (Fig. 1b). N221D and Q665E-S690T amino acid changes were not found to affect PC1/3 maturation and secretion in lysate and medium (Fig. 1c). G593R, a substitution resulting in the human PC1/3 deficiency⁵ and N222D were tested as experimental controls. As previously reported, we detected decreased enzymatic activity of the N222D protein (36.3% decrease; $P = 0.002$)¹⁰, a reduction of catalytic activity to background levels (98.7% decrease; $P = 10^{-4}$) and an altered cleavage and maturation of the recombinant G593R protein⁵. Altogether, our functional data suggest a modest deleterious role of the N221D, but no significant impact of the Q665E-S690T amino acid substitutions was identified. In view of the previously described independent genetic effect of the Q665E-S690T cluster, the exploration of other, noncatalytic functions of PC1/3 may be of interest. Moreover, the variants found to be associated with obesity in the initial French case-control and the five known noncoding SNPs (one 3' UTR and four intronic variations) in high LD ($r^2 \geq 0.8$) with the rs6234-rs6235 pair should not be excluded as having individual or cumulative consequences on PC1/3 expression and activity levels.

In conclusion, the multiple replications and the compelling overall significance of associations with obesity found for frequent nonsynonymous *PCSK1* variants in a total of 13,659 participants of European ancestry of all ages and the significant catalytic effect found for the frequent N221D substitution support the contribution of *PCSK1* to polygenic obesity. Although additional functional analysis and replication in other cohorts will be needed, these findings firmly place *PCSK1* on the short list of genes reproducibly associated with common obesity and emphasize the need to investigate PC1/3 and its related substrates for identification of specific therapeutic targets for treatment of common obesity.

Note: Supplementary information is available on the Nature Genetics website.

AUTHOR CONTRIBUTIONS

M.B. was responsible for the study design, project follow-up and statistical analysis, was involved in the *in vitro* study and wrote the manuscript. J.W.M.C. was responsible for the *in vitro* study. H.C. was involved in the *in vitro* study. S.L., E.D. were involved in the genotyping. C.D. was involved in the statistical analysis. A.G. was involved in the sequencing. P.B. was involved in the study design. B.J., B.H., B.B., J.T., M.M., N.P. provided the DNA and quantitative data for the D.E.S.I.R. study. F.H. provided the DNA and quantitative data for the Swiss study. C.L.S. and S.C. provided the DNA and quantitative data for the second French children obesity study. W.K., A.K. and P.K. provided the DNA and quantitative data for the German study. P.J. and L.M.S.C. provided the DNA and quantitative data for the Swedish study. A.J.W. helped draft the manuscript. A.S., T.L., K.B.-J., G.A., T.J., T.H. and O.P. provided the DNA and quantitative data for the Danish case-control study. D.M. was involved in the study design and helped draft the manuscript. P.F. was the principal investigator on the project was involved in the study design, data analyses and in the manuscript writing.

Published online at <http://www.nature.com/naturegenetics/>
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Hager, J. *et al. Nat. Genet.* **20**, 304–308 (1998).
- Chagnon, Y.C. *et al. J. Appl. Physiol.* **90**, 1777–1787 (2001).
- Bell, C.G. *et al. Diabetes* **53**, 1857–1865 (2004).
- Chen, G. *et al. Int. J. Obes. (Lond.)* **29**, 255–259 (2005).
- Jackson, R.S. *et al. Nat. Genet.* **16**, 303–306 (1997).
- Jackson, R.S. *et al. J. Clin. Invest.* **112**, 1550–1560 (2003).
- Farooqi, I.S. *et al. J. Clin. Endocrinol. Metab.* **92**, 3369–3373 (2007).
- Balkau, B. *Rev. Epidemiol. Sante Publique* **44**, 373–375 (1996).
- Zhou, Y. & Lindberg, I. *J. Biol. Chem.* **269**, 18408–18413 (1994).
- Lloyd, D.J., Bohan, S. & Gekakis, N. *Hum. Mol. Genet.* **15**, 1884–1893 (2006).