

RESEARCH ARTICLE

Open Access

A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes

Luca Fontanesi^{1,2*}, Giuseppina Schiavo^{1,2†}, Giuliano Galimberti^{3†}, Daniela Giovanna Calò^{3†}, Emilio Scotti¹, Pier Luigi Martelli^{2,4}, Luca Buttazoni⁵, Rita Casadio^{2,4} and Vincenzo Russo¹

Abstract

Background: Carcass fatness is an important trait in most pig breeding programs. Following market requests, breeding plans for fresh pork consumption are usually designed to reduce carcass fat content and increase lean meat deposition. However, the Italian pig industry is mainly devoted to the production of Protected Designation of Origin dry cured hams: pigs are slaughtered at around 160 kg of live weight and the breeding goal aims at maintaining fat coverage, measured as backfat thickness to avoid excessive desiccation of the hams. This objective has shaped the genetic pool of Italian heavy pig breeds for a few decades. In this study we applied a selective genotyping approach within a population of ~ 12,000 performance tested Italian Large White pigs. Within this population, we selectively genotyped 304 pigs with extreme and divergent backfat thickness estimated breeding value by the Illumina PorcineSNP60 BeadChip and performed a genome wide association study to identify loci associated to this trait.

Results: We identified 4 single nucleotide polymorphisms with $P \leq 5.0E-07$ and additional 119 ones with $5.0E-07 < P \leq 5.0E-05$. These markers were located throughout all chromosomes. The largest numbers were found on porcine chromosomes 6 and 9 ($n=15$), 4 ($n=13$), and 7 ($n=12$) while the most significant marker was located on chromosome 18. Twenty-two single nucleotide polymorphisms were in intronic regions of genes already recognized by the Pre-Ensembl Sscrofa10.2 assembly. Gene Ontology analysis indicated an enrichment of Gene Ontology terms associated with nervous system development and regulation in concordance with results of large genome wide association studies for human obesity.

Conclusions: Further investigations are needed to evaluate the effects of the identified single nucleotide polymorphisms associated with backfat thickness on other traits as a pre-requisite for practical applications in breeding programs. Reported results could improve our understanding of the biology of fat metabolism and deposition that could also be relevant for other mammalian species including humans, confirming the role of neuronal genes on obesity.

Keywords: GWA, Backfat, Fatness, Obesity, Heavy pig

* Correspondence: luca.fontanesi@unibo.it

†Equal contributors

¹Department of Agricultural and Food Science and Technology, University of Bologna, Viale Fanin 46, Bologna 40127, Italy

²Centre for Genome Biology, University of Bologna, Bologna 40126, Italy

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

Background

Fat deposition is a key biological process that has important similarities between humans and pigs, potentially useful to elucidate mechanisms determining human obesity. This trait has practical and economical implications in pig breeding as it indirectly affects feeding efficiency and determines carcass value and consumers' acceptance of pork.

Following consumer demands, breeding goals for fresh pork generally aim at reducing carcass fatness and increasing lean meat content which has adversely affected pork quality (e.g. [1]).

The Italian pig breeding industry is mainly devoted to the production of high quality Protected Designation of Origin (PDO) dry cured hams for which pigs are raised until they reach about 160 kg live weight and appropriate fat coverage of the hams is required [2,3]. Therefore, breeding objectives aim at maintaining fat coverage measured as backfat thickness (BFT). This objective has shaped the genetic pool of Italian heavy pig breeds for a few decades.

To investigate molecular genetic aspects of fat deposition in these pigs, we have recently applied a systematic candidate gene approach and have identified tens of single nucleotide polymorphisms (SNPs) associated with BFT and/or intermuscular fat content in Italian Large White and Italian Duroc pigs [4-10]. For example, a list of more than 30 SNPs has been associated with BFT in Italian Large White, including SNPs already found by other authors in *IGF2* [11], *MCAR* [12], *TBC1D1* [8], *PPARG* [13] genes or newly identified in the *PCSK1* [14], *ACPD2*, *CALR*, *JAK3*, and *NTSE*, among several other genes [10]. Moreover, many other SNPs in additional candidate genes have been shown to explain a proportion of genetic variability of fat deposition traits in pigs [4,15,16]. In addition, a large number of QTLs for a variety of fat deposition and related traits have been already reported and listed in the Pig QTL database [17,18].

Recently, with the development of a commercial high throughput SNP genotyping tool in pig (PorcineSNP60 BeadChip [19]), a number of genome wide association (GWA) studies have been carried out in this species focusing on reproduction [20,21], boar taint [22,23], disease resistance [24], structural and body composition, including BFT [25].

We have herein applied a selective genotyping approach in the Italian Large White pig breed and genotyped extreme and divergent pigs for BFT estimated breeding value (EBV) by the Illumina PorcineSNP60 BeadChip (www.illumina.com) tool to identify chromosome regions and markers associated with BFT.

Results and discussion

SNP data

A total of 304 performance tested Italian Large White pigs were genotyped with the Illumina PorcineSNP60

BeadChip, interrogating 62,163 loci. One pig was excluded from further analysis due to a call rate below 90%. A call rate ≥ 0.90 was obtained for 58,680 SNPs (for 2,293 SNP, call rate was 0.0; 1,190 SNP had $0.0 < \text{call rate} < 0.90$). About 15.8% (9,287 SNPs) of these potentially useful SNPs had a minor allele frequency < 0.05 and were discarded. The remaining 49,393 SNPs were re-mapped on the Sscrofa10.2 genome assembly.

Genome wide association (GWA) results

Only individuals with extreme phenotypes were genotyped for association study. Several authors have shown that this approach allows to attain the same power with less genotyped individuals (e.g. [26-28]). A recent GWA study for human obesity showed that this design can obtain very similar results to previous studies on general body mass index performed on unselected cohorts of tens of thousands of subjects [29].

In our study, genotyped pigs had extreme and divergent EBV for BFT: 151 had the lowest (thinnest BFT) and 152 the highest (thickest BFT) EBV. These animals were two generation unrelated gilts taken from the performance test of the National selection program of the Italian Large White breed carried out by the National Pig Breeders Association (ANAS).

Figure 1 reports a Manhattan plot showing significant ($P \leq 5.0E-07$) and suggestively significant ($5.0E-07 < P \leq 5.0E-05$) SNPs ($P_{\text{nominal value}}$ thresholds for significant results were those indicated by the Wellcome Trust Case Control Consortium, WTCCC [30]). Using these values, 4 SNPs were significantly associated (Table 1) whereas 119 SNPs were suggestively associated with BFT (Additional file 1: Table S1).

The WTCCC criteria to reduce the number of false positive are rather conservative: the suggestive threshold for significance ($P = 5.0E-05$) corresponds to a FDR of 0.02. Had we assumed a false discovery rate (FDR) of 0.05 ($P_{\text{nominal value}} = 0.000412$), a total of 410 SNPs would have been below this threshold and therefore considered at least "suggestively associated" (data not shown).

Single nucleotide polymorphisms with $P \leq 5.0E-05$ were located in all porcine autosomal chromosomes (SCC), and on SSCX, and 5 SNPs were in unassembled scaffolds of the Sscrofa10.2 genome version. Among the mapped SNPs, the largest number was on SSC6 and SSC9 ($n = 15$), SSC4 ($n = 13$), SSC7 ($n = 12$) and SSC1 ($n = 11$) (Additional file 1: Table S1). Twenty-two SNPs were in intronic regions of recognized genes in the Pre-Ensembl Sscrofa10.2 assembly. The closest gene for the remaining mapped SNPs ($n = 96$) was located in a range from 481 bp to 4.69 Mb (mean = 287.6 kb \pm 580 kb, median = 88 kb).

The most significant SNP (ALGA0098168; $P = 3.07E-11$) was on SSC18 (Table 1). This SNP was localized in intron 3 of the phosphodiesterase 1C, calmodulin-dependent

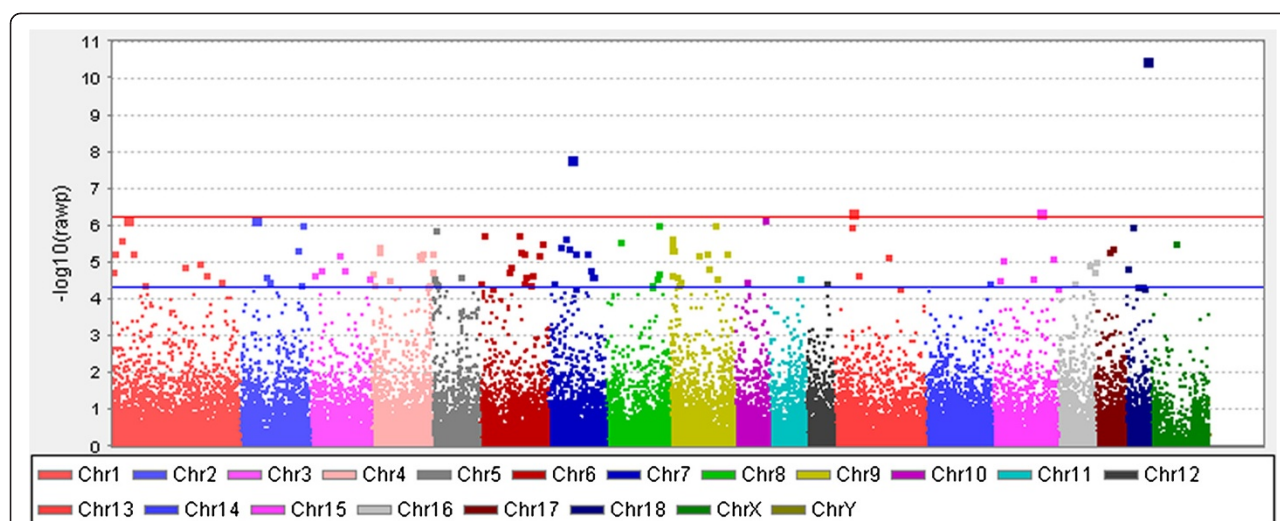


Figure 1 Manhattan plot of genome wide association results for backfat thickness in Italian Large White pigs. Red line: significance threshold ($P=5.0E-07$). Blue line ($P=5.0E-05$): threshold for suggestively significant results ($5.0E-07 < P < 5.0E-05$).

70kDa (*PDE1C*) gene. *PDE1C*, highly expressed in brain and heart, is involved in the regulation of the cellular level of adenosine 3',5'-cyclic monophosphate (cAMP) and guanosine 3',5'-cyclic monophosphate (cGMP) that play critical roles in signal transduction [31]. The second most significant SNP (M1GA0010276; $P=1.45E-08$) was localized on SSC7 at about 3.6 kb from the cysteine-rich secretory protein 1 (*CRISP1*) gene whose known function in reproduction processes is not directly linked to any fat or energy related biological function. The third top SNP (ALGA0109557; $P=3.81E-07$) was mapped on SSC15 at about 460 kb from the signal transducer and activator of transcription 4 (*STAT4*) gene. *STAT4* is a member of the STAT family of transcription factors that transduces interleukin and type 1 interferon cytokine signals in T cells and monocytes, leading to important immunological functions. Reduction of *STAT4* activation has been proposed to control obesity-induced inflammation [32]. The fourth most significant marker (ALGA0069549; $P=3.87E-07$) was located

on SSC13 at about 20 kb from the stabilin 1 (*STAB1*) gene. Another close marker (ALGA0109216; at position 38330168 of SSC13; Additional file 1: Table S1) was suggestively significant ($P=1.01E-06$). The SSC13 region bracketed by these two SNPs includes the *STAB1*-nischarin (*NISCH*) gene interval that in human has been shown to be associated with waist-hip ratio (a measure of body fat distribution) [33].

Several other genes close or within the additional suggestively significant SNPs (Additional file 1: Table S1) have been already involved in obesity related biological mechanisms. Among this list it is worth mentioning: ATP-binding cassette, sub-family B (MDR/TAP), member 1 (*ABCB1*) gene (SSC9; ALGA0109564, $P=9.01E-07$) whose altered function contributes to steatosis and obesity in mice [34] and a polymorphism in this gene has been associated with obesity risk in Japanese subjects [35]; galanin receptor 3 (*GALR3*) gene (SSC5; M1GA0007458, $P=1.25E-06$) that is upregulated in adipose tissues of mice fed a high fat diet [36], and whose function is to bind

Table 1 Significant SNPs ($P < 5.0E-07$), their chromosome positions and their closest genes in Sscrofa10.2 (Pre-Ensembl)

SNP	CHR:position ¹ (Sscrofa10.2)	CHR:position ¹ (Sscrofa10)	CHR:position ¹ (Sscrofa9.2)	P ²	FDR ³	SNP position/ distance ⁴	Closest gene coordinates	Closest gene protein in Pre- Ensembl	Gene symbol
ALGA0098168	18:45408799	18:44540120	18:25516667	3.07E-11	1.51E-06	Intron 3	18:45404849-45567252	ENSSSCP00000017656	<i>PDE1C</i>
M1GA0010276	7:50272760	7:50024255	7:50440974	1.45E-8	3.58E-04	3657	7:50276417-50297466	ENSSSCP00000001892	<i>CRISP1</i>
ALGA0109557	15:107079255	15:102388547	-	3.81E-7	4.17E-3	4609542	15:102429318-102469713	ENSP000000351255	<i>STAT4</i>
ALGA0069549	13:37851945	13:37353675	13:23719928	3.87E-7	4.17E-3	20143	13:37872088-37876725	ENSSSCP00000012196	<i>STAB1</i>

¹Chromosome and nucleotide position in the different genome versions.

²P-rav value.

³False Discovery Rate.

⁴Localization of the SNP in the corresponding gene or distance from the closest gene (in bp).

galanin, a neuropeptide that regulates food intake, neurogenesis, memory, and gut secretion; olfactory receptor genes (two genes on SSC9, *OR52N2* and *OR56A3*) have been associated with eating behaviour and adiposity in humans [37]; Parkinson protein 2 (*PARK2*) gene on SSC1 (ALGA0108518, $P=5.48E-06$) that is regulated in a lipid-dependent manner and modulates systemic fat uptake via ubiquitin ligase-dependent effects [38]; phosphodiesterase 4B, cAMP-specific (*PDE4B*) on SSC6 (ALGA0109354, $P=5.95E-06$) that has been already shown to be associated with BFT in pigs as well as with obesity in humans [39]; vacuolar protein sorting 13 homolog B (yeast) (*VPS13B*) on SSC4 (ALGA0024658, $P=3.00E-05$) that causes Cohen syndrome, characterized by truncal obesity [40]; iroquois homeobox 3 (*IRX3*) gene on SSC6 (M1GA0008432, $P=4.66E-05$), that is involved in the stress response after fat loss [41] and could be linked to obesity and type 2 diabetes through its pancreatic function [42]. Interestingly the second closest gene to this latter SSC6 SNP was *FTO*, that is well known to affect human obesity (i.e. [43]).

Even though the annotation of the pig genome available at present in Pre-Ensembl should be considered preliminary, we further evaluated the potential functional role of regions around associated or suggestively associated SNPs with BFT in our pig population (Additional file 1: Table S1). For this evaluation we used Gene Ontology (GO) information of their corresponding closest genes. Table 2 reports GO terms enriched in this dataset. Interestingly, most of the statistically significant GO terms were related to nervous system development and regulation. This indication might support and extend the role of the nervous system in the genetic predisposition of fat accumulation in mammals, as in part reported in large GWA studies in humans [44] and, subsequently, in pigs [39]. Among the genes listed in these neuronal GO categories (Table 2), few have been already reported to be indirectly associated or involved in obesity related traits. Apart from those already described above (*IRX2* and *PARK2*), it is interesting to mention the delta-like 1 (*Drosophila*) gene (*DLL1*) as this gene is located in a quite large region (~1 Mb) associated to type 1 diabetes on human chromosome 6 [45]. For several other genes involved in neuronal processes, at present, there is no direct reported link with obesity or fat metabolism. It would be important to further explore their role in affecting the investigated phenotype as a possible strategy to identify new pathways and mechanisms affecting fat deposition. For example, it could be possible to speculatively suggest a relationship between dysbindin (*DTNBP1*), involved in the modulation of glutamatergic neurotransmission in the brain, schizophrenia and obesity [46].

Comparison with other studies in pigs

We compared our GWA results with results obtained in our previous candidate gene studies for BFT in pigs [4-10] and those obtained by other GWA [25] and QTL mapping studies. In our previous studies [5,10], the *IGF2* intron3-g.3072G>A mutation [11] was the most significant marker ($P < 1.00E-50$ by selective genotyping [10]). As the *IGF2* gene is not assembled in the Sscrofa10.2 genome version, it was not possible to obtain a direct comparison with results obtained for SNPs mapped on SSC2 included in the Illumina PorcineSNP60 BeadChip. However, no SNP in the region where *IGF2* is likely to be found (0–10 Mb) reached the significance level of $P < 5.0E-05$ (Additional file 1: Table S1). Only one SNP (ASGA0008884, position 9139348; $P=2.12E-04$) was included in the list of markers with $P_{FDR} < 0.05$. Several other SSC2 SNPs were suggestively significant (Additional file 1: Table S1) indicating that they might pick up other regions affecting fatness as already reported by QTL studies (e.g. [47,48]) or candidate gene studies [5,10,49,50].

The second most significant marker of our previous candidate gene investigation was the *MC4R* p.Asp298Asn substitution [10]. In the current GWA study, no significant or suggestively significant SNPs were located in the SSC1 region around the *MC4R* gene, even if a few markers had $P < 1.0E-3$ (data not shown). The GWA study by Fan et al. [25], conducted on gilts of a commercial breeding stock, showed that markers around *MC4R* were significantly associated with 10th rib and last rib backfat. These slight differences in terms of level of significance of the markers between the two studies might be due to different *MC4R* haplotype structures in the two pig populations (Fontanesi et al. submitted) or to different positions in the pig body where BFT measurements were taken. However, in general, few results we obtained confirmed those previously obtained by Fan et al. [25] in their GWA study on BFT. This could be due to different experimental designs, incomplete power in the two studies, and/or to differences between the investigated populations. Other results we previously obtained in candidate gene studies (i.e. [10]) could be confirmed if we relaxed the significance threshold up to $FDR < 0.05$ (data not shown).

QTLs for fat deposition traits can be found over all pig chromosomes. Many different studies have repeatedly reported the presence of complex QTL patterns for fat related traits in SSC1, SSC2, SSC4, SSC6 and SSC7 [18]. In the present GWA study, SSC4, SSC6, SSC7, and also SSC9 resulted to be rich in significant or suggestively significant markers (SSC4: expected proportion = 0.068, observed = 0.110; SSC6: expected = 0.059, observed = 0.127; SSC7: expected = 0.063, observed = 0.102; SSC9: expected = 0.061, observed = 0.127). These results seem to indicate these chromosomes to support an important

Table 2 Results of the Gene Ontology (GO) analysis including closest genes to SNP with $P < 5.0E-05$

Go sub-ontology	GO term accession	GO term description	Number of involved genes	Involved genes	DAVID P-value
Biological Process	GO:0050767	regulation of neurogenesis	5	<i>ACTR3, LINGO1, IRX3, XRCC2, DLL1</i>	0.015
Biological Process	GO:0006928	cell motion	8	<i>ACTR3, CXCR4, SPOCK1, SCNN1G, IL12B, APBB2, ELMO1, CTNNA2</i>	0.018
Biological Process	GO:0051960	regulation of nervous system development	5	<i>ACTR3, LINGO1, IRX3, XRCC2, DLL1</i>	0.024
Biological Process	GO:0060284	regulation of cell development	5	<i>ACTR3, LINGO1, IRX3, XRCC2, DLL1</i>	0.029
Biological Process	GO:0030182	neuron differentiation	7	<i>LINGO1, CXCR4, MTPN, APBB2, OLFM3, NTM, CTNNA2</i>	0.038
Biological Process	GO:0045664	regulation of neuron differentiation	4	<i>ACTR3, LINGO1, IRX3, DLL1</i>	0.040
Biological Process	GO:0048666	neuron development	6	<i>LINGO1, CXCR4, APBB2, OLFM3, NTM, CTNNA2</i>	0.044
Molecular Function	GO:0031420	alkali metal ion binding	5	<i>KCNK9, KCNT2, ATP1B3, SLC22A4, SCNN1G</i>	0.031
Molecular Function	GO:0000166	nucleotide binding	19	<i>RBM24, XRCC2, SUCLG2, PKN2, ABCB1, ACTR3, MAP3K5, KCNT2, HIPK1, ASCC3, PDE1C, CELF4, SLC22A4, CELF2, DPYD, RAB38, DOCK10, ARL4C, MOCS1</i>	0.038
Cellular Component	GO:0043005	neuron projection	7	<i>NUMA1, CXCR4, MTPN, PARK2, APBB2, DTNBP1, CTNNA2</i>	0.009
Cellular Component	GO:0031252	cell leading edge	4	<i>ACTR3, CXCR4, APBB2, CTNNA2</i>	0.037

proportion of genetic variability for BFT in the Italian Large White breed. In particular, two markers below the suggestive significance threshold were located both on *IGSF3* or close to *PKN2* on SSC4 and a few close blocks of SNPs with $P < 5.0E-05$ (from about 65.1 - 65.4, 70.6 - 72.5, and 100.7 - 101.8 Mb) were located on SSC6 (Additional file 1: Table S1). As mentioned above, *FTO* is close to the marker at position 28215213 on SSC6. Single marker analysis using a few *FTO* SNPs in our previous large association study with BFT in Italian Large White pigs did not produce significant results [10]. However, subsequent haplotype analysis at this locus tended to confirm *FTO* as an important locus affecting fat deposition also in this pig breed [51].

Conclusions

This study is the first genome wide association analysis for BFT in Italian heavy pigs. The targeted trait is of paramount importance for the Italian pig breeding industry that is devoted to the production of high quality dry-cured hams for which fat coverage is a key factor during the processing and curing steps [2,3]. The genetic dissection of BFT could open new perspectives to improve selection efficiency. In this study we applied a selective genotyping approach within the Italian Large White pig population to reduce the cost of genotyping without losing much power [26-31]. We took advantage

of the large number of pigs that have been performance tested and genetically evaluated under the National selection program for this breed. The association analysis that compared SNP genotype frequencies between low BFT-EBV vs. high BFT-EBV groups identified 123 SNPs with $P < 5.0E-5$ that were more densely represented in a few chromosomes known to harbor important QTLs for fat deposition traits. The quite large number of markers below this threshold (spread in different chromosome regions) might indirectly support the fact that many genes, each with a small-medium contribution, are involved in determining BFT, according to the classical definition of a quantitative trait.

Several significant or suggestively significant SNPs were close to genes whose function might be directly or indirectly related to energy metabolism and fat deposition. Many other cannot be easily linked to the targeted trait and might provide, if confirmed in following up studies, new evidence on this matter. Even if the annotation available in Pre-Ensembl for Sscrofa10.2 is preliminary, GO enrichment analysis indicated that neuronal genes might affect fat deposition in pig confirming and enlarging previous indications reported in humans [44].

Summarizing, as more information is becoming available in pigs on biological aspects of fat metabolism and deposition, it is more and more clear that this species could represent an attractive biomedical model for

human obesity and associated diseases. Data here reported could give an insight over genetic mechanisms of fat metabolism and deposition that could be helpful in understanding also biology aspects of human obesity.

Methods

Animals and phenotypic traits

All animals used in this study were kept according to Italian and European legislation for pig production and all procedures described were in compliance with national and European Union regulations for animal care and slaughtering.

The national selection program of the Italian Large White breed is based on triplets of siblings from the same litter (two females and one castrated male) that are individually performance tested at the Central Test Station of the National Pig Breeder Association (ANAS) for the genetic evaluation of a boar from the same litter (sib-testing). Performance evaluation starts when the pigs are 30 to 45 days of age and it ends when the animals reach 155 ± 5 kg live weight. The nutritive level is *quasi ad libitum*, meaning that about 60% of the pigs are able to ingest the entire supplied ration. At the end of test, animals are transported to a commercial abattoir where they are slaughtered following standard procedures [52]. Then, backfat thickness is measured on the carcasses at the level of *Musculus gluteus medius*.

The association study was conducted following a selective genotyping approach (e.g. [26-31]). In this study we genotyped two extreme and divergent groups of Italian Large White gilts of these triplets (one female per triplet), performance tested in the period 1996–2007. Two-generation unrelated females (i.e. gilts with different and unique parents) were chosen according to their EBV for BFT (152 with most negative and 152 with most positive EBV) within a performance tested population of ~12000 pigs (details of EBV calculation are reported below). The two extreme groups were chosen ranking the animals according to their BFT EBV and then taking only the first unrelated gilts in the list (with the most positive or the most negative BFT EBV). BFT EBV used to choose the animals were recalculated for the whole performance tested population in 2007. Average BFT EBV in the negative and positive selected groups of pigs were -9.8 ± 1.6 mm and $+6.6 \pm 2.3$ mm, respectively. Genotyped pigs were a subset of the 560 two-generation unrelated pigs used in our previous candidate gene association study [10].

Genotyping

Genomic DNA was extracted from dried-blood by standard protocols. Based on quality control, all animals were used for genotyping using the PorcineSNP60

BeadChip [19] developed by Illumina according to manufacturer's protocol [53].

Data analyses

Estimated breeding values for BFT were calculated in the population using a BLUP-Multiple Trait-Animal Model that included the fixed effect of sex (considering the triplets of pigs from the same litter), batch on trial, inbreeding coefficient of the animal, interaction of sex by age at slaughtering, date of slaughtering and random effect of litter and animal. Three criteria were used to filter animals and SNP before association analysis: call rate >0.9 both at the 1) animal and 2) SNP level, and 3) $MAF > 0.05$. Animals and SNPs that passed these filters were taken for association analysis treating the two groups as cases and controls. Full pedigree information available was used to obtain a kinship matrix. In order to correct for possible family-based stratification (see Additional file 2: Figure S1), the EIGENSTRAT method [54] was applied including the kinship matrix, and association tests were performed. All analyses were performed in R [55], using an option of the package GenABEL [56] for computing the test-statistics according to the EIGENSTRAT method, and the package kinship [57] for building the pedigree kinship matrix.

For n animals, the first $K < n$ principal components, c_1, \dots, c_K , of the kinship matrix among the animals were used as axes of genetic variation. Let g_{ij} and p_j be the genotype at SNP i ($g_{ij} = 0, 1$ or 2) and the phenotype of animal j , respectively, a PC-based adjustment was performed on genotypes and phenotypes according to the following formulas:

$$g_{ij}^* = g_{ij} - \beta_{1i}c_{1j} - \dots - \beta_{Ki}c_{Kj}$$

$$p_j^* = p_j - \gamma_1c_{1j} - \dots - \gamma_Kc_{Kj}$$

where c_{kj} is the score of the k -th component on animal j , β_{ki} and γ_k are the partial regression coefficients for predicting the i -th genotype and the phenotype, respectively, on the basis of the k -th component (with $k = 1, \dots, K$).

The association test-statistic is computed as $(n - K - 1)r_i^2$, where

$$r_i^2 = \frac{\left(\sum_{j=1}^n p_j^* g_{ij}^* \right)^2}{\sum_{j=1}^n (p_j^*)^2 \sum_{j=1}^n (g_{ij}^*)^2}$$

is the squared correlation coefficient between the i -th PC-adjusted genotype and PC-adjusted phenotype. As noted by Price et al. [54], this statistic is a generalization of the Armitage trend statistic for discrete genotypes and phenotypes.

Wellcome Trust Case Control Consortium significance thresholds, whose definition depends on the prior odds and power, were used in this study [30]. In addition, correction for multiple testing was achieved by using a False Discovery Rate approach [58]. For each chromosome, the expected proportion of SNPs with $P < 5.0E-5$ was computed under the assumption of uniform distribution from the informative SNPs over the chromosome. This proportion was compared to the proportion of significant or suggestively significant markers actually observed on the same chromosome.

Bioinformatics analyses

Mapping of the PorcineSNP60 BeadChip SNPs was obtained by using BWA [59] on the Sscrofa9.2 and Sscrofa10.2 genome assemblies as previously described [10] and confirmed using the BLAT analysis available at the Ensembl (http://www.ensembl.org/Sus_scrofa/Info/Index) and Pre-Ensembl (http://pre.ensembl.org/Sus_scrofa/Info/Index) databases (February 2012). Coordinates for the Sscrofa10 genome preliminary version (September 2010) were downloaded from the Animal Genome repository web site <http://www.animalgenome.org/repository/>. Identification of the closest genes to SNPs with $P < 5.0E-05$ was obtained using Pre-Ensembl annotation of Sscrofa10.2 genome version and verified using Ensembl Sscrofa9.2 genome version (February 2012). Starting from the corresponding protein sequences retrieved from these databases, the corresponding gene symbols were extracted from NCBI Gene section (<http://www.ncbi.nlm.nih.gov/>) and/or Uniprot (<http://www.uniprot.org/>) databases (February 2012). Gene annotation was verified by BLAST analysis (<http://blast.ncbi.nlm.nih.gov/>). Gene Ontology analysis was carried out using DAVID Bioinformatics Resources 6.7 (<http://david.abcc.ncifcrf.gov/> [60]).

Additional files

Additional file 1: Table S1. Suggestively significant SNPs ($5.0E-07 < P \leq 5.0E-05$), their chromosome positions and their closest genes in Sscrofa10.2 (Pre-Ensembl). Notes are the same as those reported for Table 1.

Additional file 2: Figure S1. Two-dimensional graphical representation of relatedness among animals based on a multidimensional scaling representation of pedigree-based kinship matrix. Different symbols are used to denote pigs with positive or negative backfat thickness estimated breeding values.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LF conceived and coordinated the study, analysed data and drafted the manuscript. GS, GG, DGC, and PLM performed statistical and bioinformatics analyses. ES carried out laboratory activities. LB, RC and VR coordinated and

conceived the study. All authors reviewed and contributed to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank ANAS for providing samples and data. This work was supported by University of Bologna RFO and FAGenomicH projects and by Italian MiPAAF INNOVAGEN funds.

Author details

¹Department of Agricultural and Food Science and Technology, University of Bologna, Viale Fanin 46, Bologna 40127, Italy. ²Centre for Genome Biology, University of Bologna, Bologna 40126, Italy. ³Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Via delle Belle Arti 41, Bologna 40126, Italy. ⁴Biocomputing Group, Department of Biology, Geology and Environmental Science, University of Bologna, Via San Giacomo, Bologna 40126, Italy. ⁵Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico (CRA-PCM), Via Salaria 31, Monterotondo Scalo, Roma 00015, Italy.

Received: 20 February 2012 Accepted: 25 October 2012

Published: 15 November 2012

References

1. Lonergan SM, Huff-Lonergan E, Rowe LJ, Kuhlers DL, Jungst SB: Selection for lean growth efficiency in Duroc pigs influences pork quality. *J Anim Sci* 2001, **79**:2075–2085.
2. Russo V, Nanni Costa L: Suitability of pig meat for salting and the production of quality processed products. *Pig News Inform* 1995, **16**:7N–26N.
3. Bosi P, Russo V: The production of the heavy pig for high quality processed products. *Ital J Anim Sci* 2004, **3**:309–321.
4. Fontanesi L, Scotti E, Buttazzoni L, Davoli R, Russo V: The porcine fat mass and obesity associated (*FTO*) gene is associated with fat deposition in Italian Duroc pigs. *Anim Genet* 2009, **40**:90–93.
5. Fontanesi L, Speroni C, Buttazzoni L, Scotti E, Dall'Olio S, Nanni Costa L, Davoli R, Russo V: The insulin-like growth factor 2 (*IGF2*) gene intron-3-g.3072G>A polymorphism is not the only *Sus scrofa* chromosome 2p mutation affecting meat production and carcass traits in pigs: evidence from the effects of a cathepsin D (*CTSD*) gene polymorphism. *J Anim Sci* 2010, **88**:2235–2245.
6. Fontanesi L, Colombo M, Scotti E, Buttazzoni L, Bertolini F, Dall'Olio S, Davoli R, Russo V: The porcine *tribbles homolog 3* (*TRIB3*) gene: identification of a missense mutation and association analysis with meat quality and production traits in Italian heavy pigs. *Meat Sci* 2010, **86**:806–813.
7. Fontanesi L, Scotti E, Speroni C, Buttazzoni L, Russo V: A selective genotyping approach identifies single nucleotide polymorphisms in porcine chromosome 2 genes associated with production and carcass traits in Italian heavy pigs. *Ital J Anim Sci* 2011, **10**:e15.
8. Fontanesi L, Colombo M, Tognazzi L, Scotti E, Buttazzoni L, Dall'Olio S, Davoli R, Russo V: The porcine *TBC1D1* gene: mapping, SNP identification, and association study with meat, carcass and production traits in Italian heavy pigs. *Mol Biol Rep* 2011, **38**:1425–1431.
9. Fontanesi L, Speroni C, Buttazzoni L, Scotti E, Dall'Olio S, Davoli R, Russo V: Association between polymorphisms in cathepsin and cystatin genes with meat production and carcass traits in Italian Duroc pigs: confirmation of the effects of a cathepsin L (*CTSL*) gene marker. *Mol Biol Rep* 2012, **39**:109–115.
10. Fontanesi L, Galimberti G, Calò DG, Fronza R, Martelli PL, Scotti E, Colombo M, Schiavo G, Casadio R, Buttazzoni L, Russo V: Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *J Anim Sci* 2012, **90**:2450–2464.
11. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L: A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 2003, **425**:832–836.
12. Kim KS, Larsen N, Short T, Plastow G, Rothschild MF: A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. *Mamm Genome* 2000, **11**:131–135.
13. Fan B, Onteru SK, Mote BE, Serenius T, Stalder KJ, Rothschild MF: Large-scale association study for structural soundness and leg locomotion traits in the pig. *Genet Sel Evol* 2009, **41**:14.

14. Fontanesi L, Bertolini F, Scotti E, Trevisi P, Buttazzoni L, Dall'Olio S, Davoli R, Bosi P, Russo V: **Polymorphisms in an obesity related gene (PCSK1) are associated with fat deposition and production traits in Italian heavy pigs.** *Animal* 2012, **6**:1913–1924.
15. Fan B, Du ZQ, Rothschild MF: **The fat mass and obesity-associated (FTO) gene is associated with intramuscular fat content and growth rate in the pig.** *Anim Biotechnol* 2009, **20**:58–70.
16. Switonski M, Stachowiak M, Cieslak J, Bartz M, Grzes M: **Genetics of fat tissue accumulation in pigs: a comparative approach.** *J Appl Genet* 2010, **51**:153–168.
17. Rothschild MF, Hu Z-L, Jiang Z: **Advances in QTL mapping in pigs.** *Int J Biol Sci* 2007, **3**:192–197.
18. Hu Z-L, Reecy JM: **Animal QTLdb: Beyond a repository – A public platform for QTL comparisons and integration with diverse types of structural genomic information.** *Mamm Genome* 2007, **18**:1–4.
19. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MA: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.** *PLoS One* 2009, **4**:e6524.
20. Uimari P, Sironen A, Sevón-Aimonen ML: **Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed.** *Genet Sel Evol* 2011, **43**:42.
21. Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF: **A whole-genome association study for pig reproductive traits.** *Anim Genet* 2012, **43**:18–26.
22. Grindflek E, Lien S, Hamland H, Hansen MH, Kent M, van Son M, Meuwissen TH: **Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids.** *BMC Genomics* 2011, **12**:362.
23. Gregersen VR, Conley LN, Sorensen KK, Guldbrandtsen B, Velander IH, Bendixen C: **Genome-wide association scan and phased haplotype construction for quantitative trait loci affecting boar taint in three pig breeds.** *BMC Genomics* 2012, **13**:22.
24. Boddicker N, Waide EH, Rowland RR, Lunney JK, Garrick DJ, Reecy JM, Dekkers JC: **Evidence for a major QTL associated with host response to Porcine Reproductive and Respiratory Syndrome virus challenge.** *J Anim Sci* 2012, doi:10.2527/jas.2011-4464.
25. Fan B, Onteru SK, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF: **Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs.** *PLoS One* 2011, **6**:e14726.
26. Darvasi A, Soller M: **Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus.** *Theor Appl Genet* 1992, **85**:353–359.
27. Van Gestel S, Houwing-Duistermaat JJ, Adolfsson R, van Duijn CM, Van Broeckhoven C: **Power of selective genotyping in genetic association analyses of quantitative traits.** *Behav Genet* 2000, **30**:141–146.
28. Zhang G, Nebert DW, Chakraborty R, Jin L: **Statistical power of association using the extreme discordant phenotype design.** *Pharmacogenet Genomics* 2006, **16**:401–413.
29. Paternoster L, Evans DM, Nohr EA, Holst C, Gaborieau V, Brennan P, Gjesing AP, Grarup N, Witte DR, Jørgensen T, Linneberg A, Lauritzen T, Sandbaek A, Hansen T, Pedersen O, Elliott KS, Kemp JP, St Pourcain B, McMahon G, Zelenika D, Hager J, Lathrop M, Timpson NJ, Smith GD, Sørensen TI: **Genome-wide population-based association study of extremely overweight young adults - the GOYA study.** *PLoS One* 2011, **6**:e24303.
30. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447**:661–683.
31. Lakics V, Karran EH, Boess FG: **Quantitative comparison of phosphodiesterase mRNA distribution in human brain and peripheral tissues.** *Neuropharmacology* 2010, **59**:367–374.
32. Chakrabarti SK, Wen Y, Dobrian AD, Cole BK, Ma Q, Pei H, Williams MD, Bevard MH, Vandenhoff GE, Keller SR, Gu J, Nadler JL: **Evidence for activation of inflammatory lipoxigenase pathways in visceral adipose tissue of obese Zucker rats.** *Am J Physiol Endocrinol Metab* 2011, **300**:E175–E187.
33. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Mägi R, Workalemahu T, White CC, Bouatia-Naji N, Harris TB, Berndt SI, Ingelsson E, Willer CJ, Weedon MN, Luan J, Vedantam S, Esko T, Kilpeläinen TO, Kutalik Z, Li S, Monda KL, Dixon AL, Holmes CC, Kaplan LM, Liang L, Min JL, et al: **Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution.** *Nat Genet* 2010, **42**:949–960.
34. Foucaud-Vignault M, Soayfane Z, Ménez C, Bertrand-Michel J, Martin PG, Guillou H, Collet X, Lespine A: **P-glycoprotein dysfunction contributes to hepatic steatosis and obesity in mice.** *PLoS One* 2011, **6**:e23614.
35. Ichihara S, Yamada Y, Kato K, Hibino T, Yokoi K, Matsuo H, Kojima T, Watanabe S, Metoki N, Yoshida H, Satoh K, Aoyagi Y, Yasunaga A, Park H, Tanaka M, Nozawa Y: **Association of a polymorphism of ABCB1 with obesity in Japanese individuals.** *Genomics* 2008, **91**:512–516.
36. Kim A, Park T: **Diet-induced obesity regulates the galanin-mediated signalling cascade in the adipose tissue of mice.** *Mol Nutr Food Res* 2010, **54**:1361–1370.
37. Choquette AC, Bouchard L, Drapeau V, Lemieux S, Tremblay A, Bouchard C, Vohl MC, Pérusse L: **Association between olfactory receptor genes, eating behaviour traits and adiposity: results from the Quebec Family Study.** *Physiol Behav* 2012, **105**:772–776.
38. Kim KY, Stevens MV, Akter MH, Rusk SE, Huang RJ, Cohen A, Noguchi A, Springer D, Bocharov AV, Egggerman TL, Suen DF, Youle RJ, Amar M, Remaley AT, Sack MN: **Parkin is a lipid-responsive regulator of fat uptake in mice and mutant human cells.** *J Clin Invest* 2011, **121**:3701–3712.
39. Lee KT, Byun MJ, Kang KS, Park EW, Lee SH, Cho S, Kim H, Kim KW, Lee T, Park JE, Park W, Shin D, Park HS, Jeon JT, Choi BH, Jang GW, Choi SH, Kim DW, Lim D, Park HS, Park MR, Ott J, Schook LB, Kim TH, Kim H: **Neuronal genes for subcutaneous fat thickness in human and pig are identified by local genomic sequencing and combined SNP association study.** *PLoS One* 2011, **6**:e16356.
40. Balikova I, Lehesjoki AE, de Ravel TJ, Thienpont B, Chandler KE, Clayton-Smith J, Traskelin AL, Fryns JP, Vermeesch JR: **Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome.** *Hum Mutat* 2009, **30**:E845–E854.
41. Dankel SN, Fadnes DJ, Stavrum AK, Stansberg C, Holdhus R, Hoang T, Veum VL, Christensen BJ, Våge V, Sagen JV, Steen VM, Mellgren G: **Switch from stress response to homeobox transcription factors in adipose tissue after profound fat loss.** *PLoS One* 2010, **5**:e11033.
42. Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes O, Engström PG, Alonso ME, de la Calle Mustienes E, Gómez Skarmeta JL, Tavares MJ, Casares F, Manzanares M, van Heyningen V, Molven A, Nijlstad PR, Almgren P, Lenhard B, Becker TS: **Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3.** *Proc Natl Acad Sci USA* 2010, **107**:775–780.
43. Dina C, Meyre D, Gallina S, Durand E, Körner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Herberg S, Le Stunff C, Bougnères P, Kovacs P, Marre M, Balkau B, Cauchi S, Chèvre JC, Froguel P: **Variation in FTO contributes to childhood obesity and severe adult obesity.** *Nat Genet* 2007, **39**:724–726.
44. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, Lettre G, Lim N, Lyon HN, McCarrroll SA, Papadakis K, Qi L, Randall JC, Rocaasecca RM, Sanna S, Scheet P, Weedon MN, Wheeler E, Zhao JH, Jacobs LC, Prokopenko I, Soranzo N, Tanaka T, Timpson NJ, Almgren P, Bennett A, Genetic Investigation of Anthropometric Traits Consortium, et al: **Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.** *Nat Genet* 2009, **41**:25–34.
45. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SF, Polychronakos C, Hakonarson H: **A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci.** *PLoS Genet* 2011, **7**:e1002293.
46. Numakawa T, Yagasaki Y, Ishimoto T, Okada T, Suzuki T, Iwata N: **Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia.** *Hum Mol Genet* 2004, **13**:2699–26708.
47. Lee SS, Chen Y, Moran C, Cepica S, Reiner G, Bartenschlager H, Moser G, Geldermann H: **Linkage and QTL mapping for *Sus scrofa* chromosome 2.** *J Anim Breed Genet* 2003, **120**(Suppl. 1):11–19.
48. Tortoreau F, Gilbert H, Heuven HC, Bidanel JP, Groenen MA, Riquet J: **Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in Sino-European pig pedigrees.** *Genet Sel Evol* 2011, **43**:11.
49. Cepica S, Ovilo C, Masopust M, Knoll A, Fernandez A, Lopez A, Rohrer GA, Nonneman D: **Four genes located on a SSC2 meat quality QTL region are associated with different meat quality traits in Landrace × Chinese-European crossbred population.** *Anim Genet* 2012, **43**:333–336.

50. Fontanesi L, Buttazzoni L, Scotti E, Russo V: **Confirmation of the association between a single nucleotide polymorphism in the porcine *LDHA* gene and average daily gain and correlated traits in Italian Large White pigs.** *Anim Genet* 2012, **43**:649–650.
51. Fontanesi L, Russo V: **Nucleotide variability and haplotype heterogeneity at the porcine *fat mass and obesity associated (FTO)* gene.** *Anim Genet*. doi:10.1111/j.1365-2052.2012.02352.x.
52. Fontanesi L, Davoli R, Nanni Costa L, Beretti F, Scotti E, Tazzoli M, Tassone F, Colombo M, Buttazzoni L, Russo V: **Investigation of candidate genes for glycolytic potential of porcine skeletal muscle: association with meat quality and production traits in Italian Large White pigs.** *Meat Sci* 2008, **80**:780–787.
53. Steemers FJ, Gunderson KL: **Whole genome genotyping technologies on the BeadArray platform.** *Biotechnol J* 2007, **2**:41–49.
54. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38**:904–909.
55. R Development Core Team: *R: A language and environment for statistical computing* <http://www.R-project.org/>.
56. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM: **GenABEL: an R library for genome-wide association analysis.** *Bioinformatics* 2007, **23**:1294–1296.
57. Atkinson B, Therneau T: *kinship: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees.* R package version 1.1.0-23, 2009: <http://CRAN.R-project.org/package=kinship>.
58. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Royal Stat Soc Series B* 1995, **57**:289–300.
59. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754–1760.
60. Huang DW, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**:1–13.

doi:10.1186/1471-2164-13-583

Cite this article as: Fontanesi *et al.*: A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* 2012 **13**:583.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

