



Genome-wide screening for genetic variants in polyadenylation signal (PAS) sites in mouse selection lines for fatness and leanness

Martin Šimon¹ · Špela Mikec¹ · Nicholas M. Morton² · Santosh S. Atanur^{3,4} · Janez Konc⁵ · Simon Horvat¹ · Tanja Kunej¹

Received: 29 April 2022 / Accepted: 31 October 2022 / Published online: 21 November 2022
© The Author(s) 2022

Abstract

Alternative polyadenylation (APA) determines mRNA stability, localisation, translation and protein function. Several diseases, including obesity, have been linked to APA. Studies have shown that single nucleotide polymorphisms in polyadenylation signals (PAS-SNPs) can influence APA and affect phenotype and disease susceptibility. However, these studies focussed on associations between single PAS-SNP alleles with very large effects and phenotype. Therefore, we performed a genome-wide screening for PAS-SNPs in the polygenic mouse selection lines for fatness and leanness by whole-genome sequencing. The genetic variants identified in the two lines were overlapped with locations of PAS sites obtained from the PolyASite 2.0 database. Expression data for selected genes were extracted from the microarray expression experiment performed on multiple tissue samples. In total, 682 PAS-SNPs were identified within 583 genes involved in various biological processes, including transport, protein modifications and degradation, cell adhesion and immune response. Moreover, 63 of the 583 orthologous genes in human have been previously associated with human diseases, such as nervous system and physical disorders, and immune, endocrine, and metabolic diseases. In both lines, PAS-SNPs have also been identified in genes broadly involved in APA, such as *Polr2c*, *Eif3e* and *Ints11*. Five PAS-SNPs within 5 genes (*Car*, *Col4a1*, *Itga7*, *Lat*, *Nmnat1*) were prioritised as potential functional variants and could contribute to the phenotypic disparity between the two selection lines. The developed PAS-SNPs catalogue presents a key resource for planning functional studies to uncover the role of PAS-SNPs in APA, disease susceptibility and fat deposition.

Introduction

Various cellular mechanisms determine both the outcome of transcription and the function of the protein encoded by the same gene. Alternative polyadenylation (APA) has recently

Martin Šimon and Špela Mikec have contributed equally to this work.

✉ Simon Horvat
simon.horvat@bf.uni-lj.si

✉ Tanja Kunej
tanja.kunej@bf.uni-lj.si

Martin Šimon
martin.simon@bf.uni-lj.si

Špela Mikec
spela.mikec@bf.uni-lj.si

Nicholas M. Morton
nik.morton@ed.ac.uk

Santosh S. Atanur
santosh.atanur@imperial.ac.uk

Janez Konc
janez.konc@ki.si

¹ Biotechnical Faculty, Department of Animal Science, University of Ljubljana, Domžale, Slovenia

² University of Edinburgh, The Queen's Medical Research Institute, Centre for Cardiovascular Science, Edinburgh, UK

³ Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

⁴ Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK

⁵ Laboratory for Molecular Modeling, National Institute of Chemistry, Ljubljana, Slovenia

attracted a great deal of attention (Gebauer and Hentze 2004; Zhang et al. 2021) as it critically affects mRNA stability, localization, translation, protein coding and localization, and is also instrumental in the regulation of gene expression and gene function (Yuan et al. 2021a). It is estimated that more than a third of mouse and two-thirds of human genes undergo alternative polyadenylation events, resulting in various APA transcripts of a single gene (Yuan et al. 2021a). While 80% of APA events occur in the 3' UTR, resulting in transcripts with different 3' UTR lengths (Nourse et al. 2020) and consequently with preserved or lost interaction sites for regulators such as miRNAs, lncRNAs and RNA-binding proteins (RBPs) (Tian and Manley 2017), 20% of APA events occur upstream of the last terminal exon, often in an alternatively spliced intron (Nourse et al. 2020). The latter can lead to mRNA decay pathways or the production of truncated proteins (Yuan et al. 2021a).

Cleavage at a site where poly(A) tail is attached to the pre-mRNA (APA site) is regulated by adjacent *cis*-regulatory RNA elements, among which the polyadenylation signal (PAS) motif AAUAAA and its main variant AUUAAA, which are typically located approximately 20-nt upstream of the APA site, are the main components (Shulman and Elkon 2020). These are recognised by cleavage and polyadenylation specificity factors (CPSFs), of which CPSF73 performs the cleavage process (Mandel et al. 2006). Other regulatory sequences and protein complexes include UGUA accessory elements recognised by members of the mammalian cleavage factor Im complex (CFIm), and the GU-/U-rich sequence downstream of APA targeted by cleavage stimulation factors (CSTFs). In addition, the mammalian cleavage factor IIm complex (CFIIm) and the single factors poly(A) polymerase (PAPOL), poly(A)-binding proteins (PABPs) and symplekin scaffold protein (SYMPK) are also required for the cleavage and polyadenylation process (Yuan et al. 2021a).

The important role of APA in cell growth, proliferation and differentiation is well known. However, recent studies have also shown that APA is involved in various abnormal physiological conditions, including endocrine, haematological, oncological, immunological and neurological diseases (Chang et al. 2017). Surprisingly, although both *in vitro* and *in vivo* studies have shown that single nucleotide changes (SNP) in the PAS hexamer or its complete removal severely impair cleavage efficiency (Neve et al. 2017), a limited number of studies have focussed on the relationship between the polymorphic genetic code in PAS and the aetiology of a particular disease. For example, the 3' UTR variant rs78378222 in *TP53*, which changes the canonical PAS AATAAA to AATACA, increases the risk of various cancers (Wang et al. 2016). The recessive 3' UTR mutation c.*59A > G in PAS of the gene *INS* leads to reduced mRNA stability

and, consequently, neonatal diabetes through reduced insulin biosynthesis (Garin et al. 2010). A mutation in the PAS signal of the BMP1 short transcript decreases its expression and causes bone fragility in children (Fahiminiya et al. 2015), and the PAS mutation (AAUAAA to AAUGAA) in the *FOXP3* gene leads to the IPEX syndrome (Bennett et al. 2001).

Obesity, considered by many a 21st-century epidemic, has been widely believed to result from a disequilibrium between energy intake and expenditure. However, the aetiology of obesity is more complex, resulting from various factors, including genetic predispositions (González-Muniesa et al. 2017). Depending on the genes involved, three types of obesity have been described, of which polygenic obesity, in which a number of genes, each having a small effect, contribute to the phenotype, is the most common clinical manifestation. The other two types are monogenic and syndromic obesities. The latter is associated with mental retardation, dysmorphic features and organ-specific developmental anomalies (Huvenne et al. 2016). Recent studies have shown that APA may also be involved in the development of obesity. For example, targeting the cytoplasmic polyadenylation element-binding protein CPEB4 protects against diet-induced obesity (Pell et al. 2021). Also, in a high-fat-diet-induced obesity rat model, ~89,000 unique alternatively polyadenylated transcripts were identified in the hypothalamus, including those from protein-coding genes, miRNAs and lncRNAs (Brutman et al. 2018), indicating the importance of APA in the regulation of obesity. Moreover, APA events have been associated with tail fat deposition in sheep (Yuan et al. 2021b). Furthermore, while APA-induced truncation of heme oxygenase 1 (HO1) 3' UTR has a strong inhibitory effect on preadipocyte differentiation of 3T3-L1 cells (Cui et al. 2021), 3' UTR shortening of the adipogenesis-associated Mth938 domain containing (AAMDC) protein has a positive effect on the differentiation process of bovine preadipocytes (Xiao et al. 2019).

Based on the above studies, we hypothesised that DNA polymorphisms in PAS hexamers might be involved in the regulation of APA, ultimately influencing fat deposition and leading to an obese or lean phenotype. However, to our knowledge, no study has yet been conducted to investigate the association between SNPs in PAS (PAS-SNPs) and obesity (Fig. 1).

In the present study, we used unique mouse models established over 60 generations of divergent selection for increased (Fat line—FLI) and decreased (Lean line—FHI) body fat percentage (Sharp et al. 1984) that may best represent the polygenic type of human obesity and leanness (Horvat et al. 2000; Bünger et al. 2001; Simončič et al. 2008, 2011). Apart from using a unique genetic model in which long-term selection

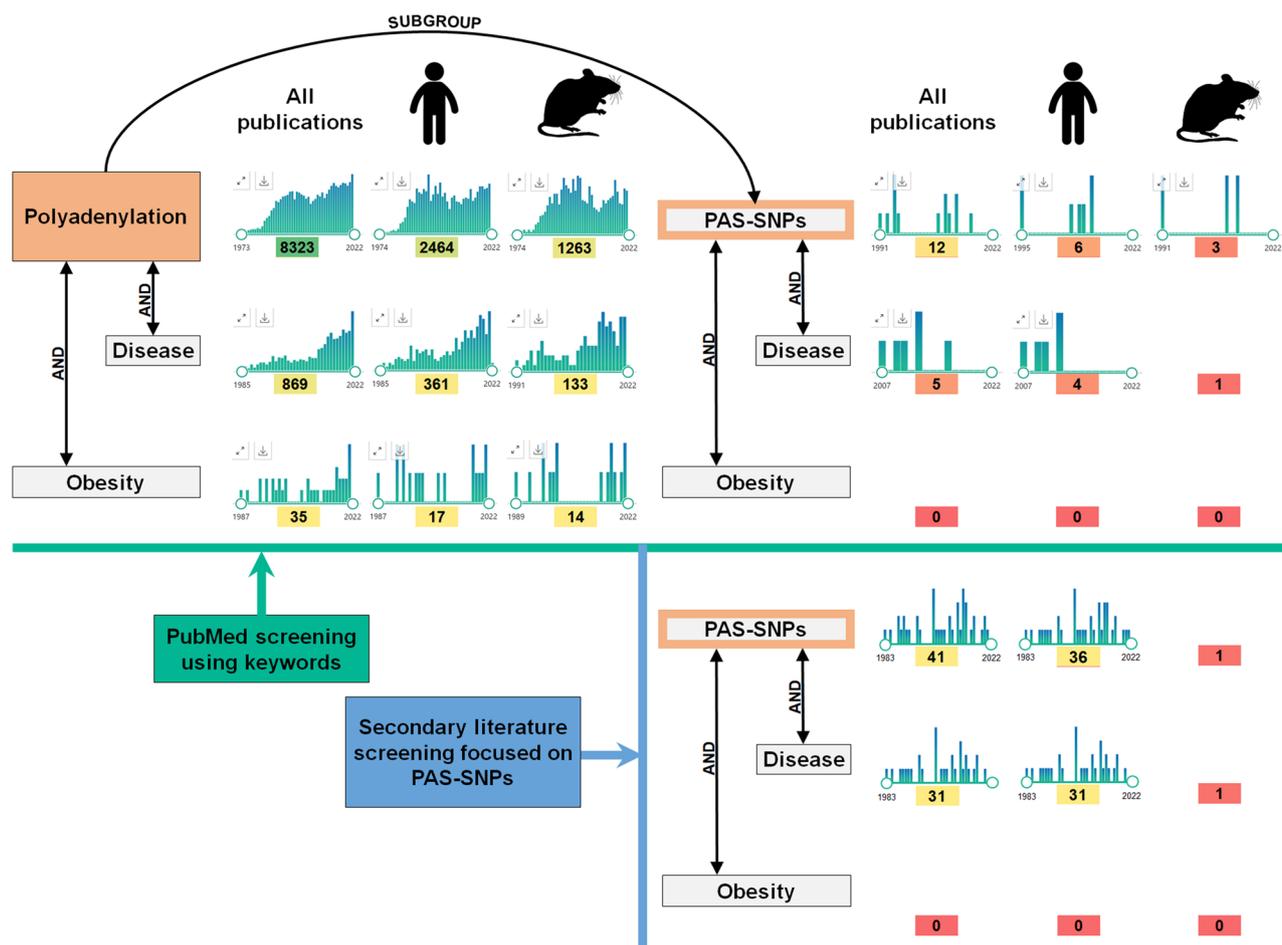


Fig. 1 The number of publications extracted from PubMed using keywords related to polyadenylation, PAS-SNPs and their relationship with diseases and obesity, and those found by manual screening. The used keywords are summarized in Supplementary Table S1

fixed opposing obesity or obesity resistance alleles in the Fat and Lean lines, respectively, our study is also novel in terms of the analytical approach. In contrast to a limited number of previous studies in which PAS-SNPs were uncovered by examining single SNP associations between a disease and a PAS-SNP allele with very large effects, our study is designed to systematically search for genome-wide PAS candidates including alleles with medium and small effects.

To indicate the importance of PAS-SNPs in the development of obesity and to promote this field of research, the aims of the present study were (1) to identify PAS-SNPs in the divergent mouse models for body fat percentage, (2) to analyse the biological processes of genes with PAS-SNPs using bioinformatics databases, (3) to identify disease-, obesity- and

APA-related genes with PAS-SNPs and (4) to identify PAS-SNPs within candidate genes with potential functional impact on their expression.

Materials and methods

Literature search

Literature related to polyadenylation, PAS-SNPs and their relationship with diseases and obesity was screened on 12/01/2022 in the PubMed database (<https://pubmed.ncbi.nlm.nih.gov/>) using search terms described in Supplementary Table S1.

Mouse selection lines

Starting with a base population of the inbred (JU, CBA) and outbred (CFLP) mouse lines, obese (FLI, Fat line) and lean (FHL, Lean line) mouse lines have been established by divergent selection for over sixty generations for increased or decreased body fat percentage (Sharp et al. 1984). The initial body fat of approximately 10% increased to 22% and decreased to 4% in the Fat and Lean lines, respectively, which resulted from the gradual accumulation of “obese” or “lean” alleles (Bünger and Hill 1999).

Identifying SNPs from whole-genome sequencing

The Illumina NextSeq 500 platform was used for the whole-genome sequencing (WGS) of DNA samples from the Fat and Lean mouse lines. Sequencing reads were first preprocessed according to the FastQC report and then mapped to the mouse reference genome (version GRCm38.86) using the Burrows-Wheeler Aligner (BWA) alignment tool (Li and Durbin 2009). Base quality score recalibration, indel realignment, duplicate removal, variant calling and hard filtering were performed according to the Genome Analysis Toolkit (GATK) (McKenna et al. 2010) Best Practices recommendations (Depristo et al. 2011; Van der Auwera et al. 2013). Variants were annotated using the Ensembl Variant Effect Predictor (<https://www.ensembl.org/Tools/VEP>) (McLaren et al. 2016).

Identification of SNPs overlapping PAS motif

The locations of SNPs identified by WGS that differ between the lines were overlapped with locations of PAS motifs obtained from the PolyASite 2.0 portal (<https://polyasite.unibas.ch/>) (Herrmann et al. 2020), considering strand-specific PAS and APA genomic positions to identify the genes with PAS-SNPs. An example is provided in Supplementary Fig. S1. Only PAS motifs within genes were included in the analysis.

Identification of SNPs overlapping PAS motif in obesity- and polyadenylation-related genes

Obesity-related genes were obtained from the literature (Mlinar et al. 2006; Swami 2009; Rao et al. 2014; da Fonseca et al. 2017; D’Angelo et al. 2018; Flores-Dorantes et al. 2020; Agrawal et al. 2021; Loos and Yeo 2022). In total, the list contains 431 genes (Supplementary Table S2). Genes (281) associated with polyadenylation were retrieved from various databases: AmiGO 2 (<http://amigo.geneontology.org/amigo>), Reactome (<https://reactome.org/>), QuickGO (<https://www.ebi.ac.uk/QuickGO/annotations>), KEGG (<https://www.kegg.jp/>), and Mouse Genome Informatics

(<http://www.informatics.jax.org>). A set of genes is given in Supplementary Table S3.

In silico functional analysis

To get a potential functional biological impact of PAS-SNPs, first, GO enrichment analysis of PAS-SNP containing genes was done using MonaGO (<https://monago.erc.monash.edu/>) (Xin et al. 2020), a visualization tool that uses GO annotation information from DAVID (<https://david.ncifcrf.gov/home.jsp>).

Disease-associated genes were identified using Mouse Genome Informatics (<http://www.informatics.jax.org>), and disease–gene interactions were visualized in Cytoscape (Version 3.9.0) (Shannon 2003). The miRNA target sites were obtained from the miRTarBase (Huang et al. 2019).

Multiple-species alignment of the genes with PAS-SNPs was performed using the Ensembl alignment tool (Howe et al. 2021). For the human SNPs at orthologous positions, their location relative to known APA sites and PAS motifs was analysed using the PolyASite 2.0 portal (<https://polyasite.unibas.ch/>) (Herrmann et al. 2020).

Expression of selected genes using microarray and prioritisation of candidate PAS-SNPs

To examine the expression of common (genes containing PAS-SNPs in both lines), disease-associated, and obesity- and APA-related genes carrying PAS-SNPs identified in either line, the expression data were obtained from the microarray transcriptome profiling performed on various mouse tissues, including white adipose tissue (WAT) [subcutaneous (sWAT), epididymal (eWAT), and mesenteric (mWAT)], brown adipose tissue (BAT), liver, muscle, adrenal gland, thymus and kidney. After RNA preparation, samples were hybridized to the Affymetrix Mouse Genome 430–2.0 GeneChip. The obtained data were processed as described previously (Morton et al. 2011; Pedroni et al. 2014). The expression of the genes was considered differential when the expression between Fat and Lean mouse lines differed by at least 1.5-fold at $p < 0.05$ (DEGs). The p value rather than adjusted p value was used to avoid losing the potential candidate PAS-SNPs; however, for the DEGs carrying candidate PAS-SNPs, their expression was also checked at the adjusted $p < 0.05$. The expressions of differentially expressed disease-related, common, obesity-related and APA-related genes carrying PAS-SNPs are given in Supplementary table S4.

Then, Affymetrix probes (locations were obtained from the Ensembl database) of DEGs were along with PAS-SNPs mapped to their corresponding genes using the Golden Helix GenomeBrowse® v3.0.0 visualization tool (<http://www.goldenhelix.com>) (Golden Helix, Inc, Bozeman, MT) to

identify candidate PAS-SNPs. The PAS-SNPs were prioritised considering their effect on PAS and locations relative to the Affymetrix probes and their expressions. For the candidate PAS-SNPs, the DNA sequence within the APA cluster and 60 bp upstream was analysed to explore whether other PAS are present and whether other SNPs in the two lines may create de-novo PAS of the corresponding APA site.

Results

The PolyASite 2.0 database [Release *Mus musculus*: v2.0 (GRCm38.96)] describes 301,006 APA sites in the mouse genome, with the AATAAA PAS motif being the most abundant, followed by ATTAAA (Supplementary Fig. S2). In the present study, we identified whole-genome PAS-SNPs specific for the Fat (309) and Lean (373) mouse selection lines for body fat percentage, located in 265 and 326 genes. Details (including PAS-SNPs identified in both lines) of PAS-SNPs are found in Supplementary Table S5. The GO enrichment analysis revealed that these genes are involved in various biological processes, but in both lines they predominantly participate in cellular transport. The two lines share 8 common genes containing PAS-SNPs that affect different APA sites within a single gene. Also, 30 (Fat line) and 33 (Lean line) genes have already been linked to human diseases. A large proportion of these genes in both lines are related to the nervous system (Fat: 36.7%, Lean: 39.4%) and physical disorders (Fat: 20.0%, Lean: 12.1%). Moreover, in the Fat line, several genes are involved in musculoskeletal disorders (Fat: 26.7%), while in the Lean line they are implicated in various syndromes (Lean: 30.3%). Furthermore, a group of genes participating in immune, endocrine and metabolic diseases were identified in both lines. In addition, a total of 14 PAS-SNPs within 6 obesity- and 7 APA-related genes were identified. 22 candidate genes carrying PAS-SNPs were differentially expressed between the two lines. Finally, manual examination prioritised 5 PAS-SNPs within *Car8*, *Itga7*, *Lat*, *Nmnat1* and *Col4a1*. The workflow and main results are shown in Fig. 2.

Genotyping

Genotyping revealed that 309 and 373 SNPs specific for either the Fat or Lean line are located in PAS within 265 and 326 genes. Interestingly, only 8 of the total 583 genes are shared between the two lines, carrying PAS-SNPs at different APA sites within the genes (*Trim11*, *Yars*, *Mrpl3*, *Arhgef4*, *Tenn4*, *Creb5*, *Smyd3*, *Gm37240*) (Supplementary Fig. S3).

GO enrichment analysis revealed that genes with PAS-SNPs are involved in various biological processes (BP), some of them even in several BP. In both lines, a

large number of these genes participate in cellular transport. Moreover, the genes in the Fat line also play a role in cytoskeletal organization and in protein modifications (protein glycosylation, protein peptidyl-prolyl isomerization). Meanwhile, in the Lean line, they participate in cell adhesion, cellular respiration and response to stimuli (Supplementary Fig. S4).

To investigate whether the sequence variants identified in the Fat and Lean mouse lines are mouse-specific or whether SNPs exist at orthologous positions in other species, we performed cross species sequence alignment using the Ensemble alignment tool. 32 PAS-SNPs are located at orthologous positions in other species (Table 1).

Of note, PAS-SNP rs1132123542 in *Snx24* of the Fat line is located at the orthologous position of human PAS-SNP rs943743662. Interestingly, the *Snx24* transcripts in both species (mouse and human) may be longer than annotated, as suggested by the APA sites located downstream of the genes. Alignment examples of parts of the mouse and human *Cxadr*, *Snx24*, *Car8*, *Abca5* and *Dner* are given in Fig. 3.

Disease-related genes with PAS-SNPs

30 (Fat line) and 33 (Lean line) human orthologous genes have already been associated with diseases. A large percentage of these genes are linked with nervous system and physical disorders in both lines. Moreover, several genes are involved in musculoskeletal disorders in the Fat line, while they are associated with syndromic diseases in the Lean line. Furthermore, a cluster of genes participating in immune, endocrine and metabolic diseases have been identified in both lines. Considering the involvement of a particular gene in various diseases, *Col4a1* (linking nervous, musculoskeletal and urinary system diseases and physical disorders), *Cttna2* (linking immune, endocrine, and metabolic diseases), and *Nmnat1* (linking nervous and musculoskeletal diseases and physical disorders) might be core genes containing PAS-SNPs in the Fat line (Fig. 4A). Meanwhile, the core genes in the Lean line might be *Nf2* (linking syndromic, nervous, urinary and cellular proliferative diseases), *Lrp1* (linking cardiovascular and nervous system diseases and physical disorders), and *Ifngr2* (linking immune, endocrine and metabolic diseases) (Fig. 4B).

Common genes with PAS-SNPs in Fat and Lean lines

There are eight common genes with PAS-SNPs identified in both lines but influencing different APA sites: *Trim11*, *Yars*, *Mrpl3*, *Arhgef4*, *Tenn4*, *Creb5*, *Smyd3*, *Gm37240*. The location of the SNPs and APA sites with the corresponding PAS hexamer within the eight genes is given in Table 2.

Moreover, not only are PAS-SNPs located in the two lines in PAS of different APAs within a given gene, PAS-SNPs

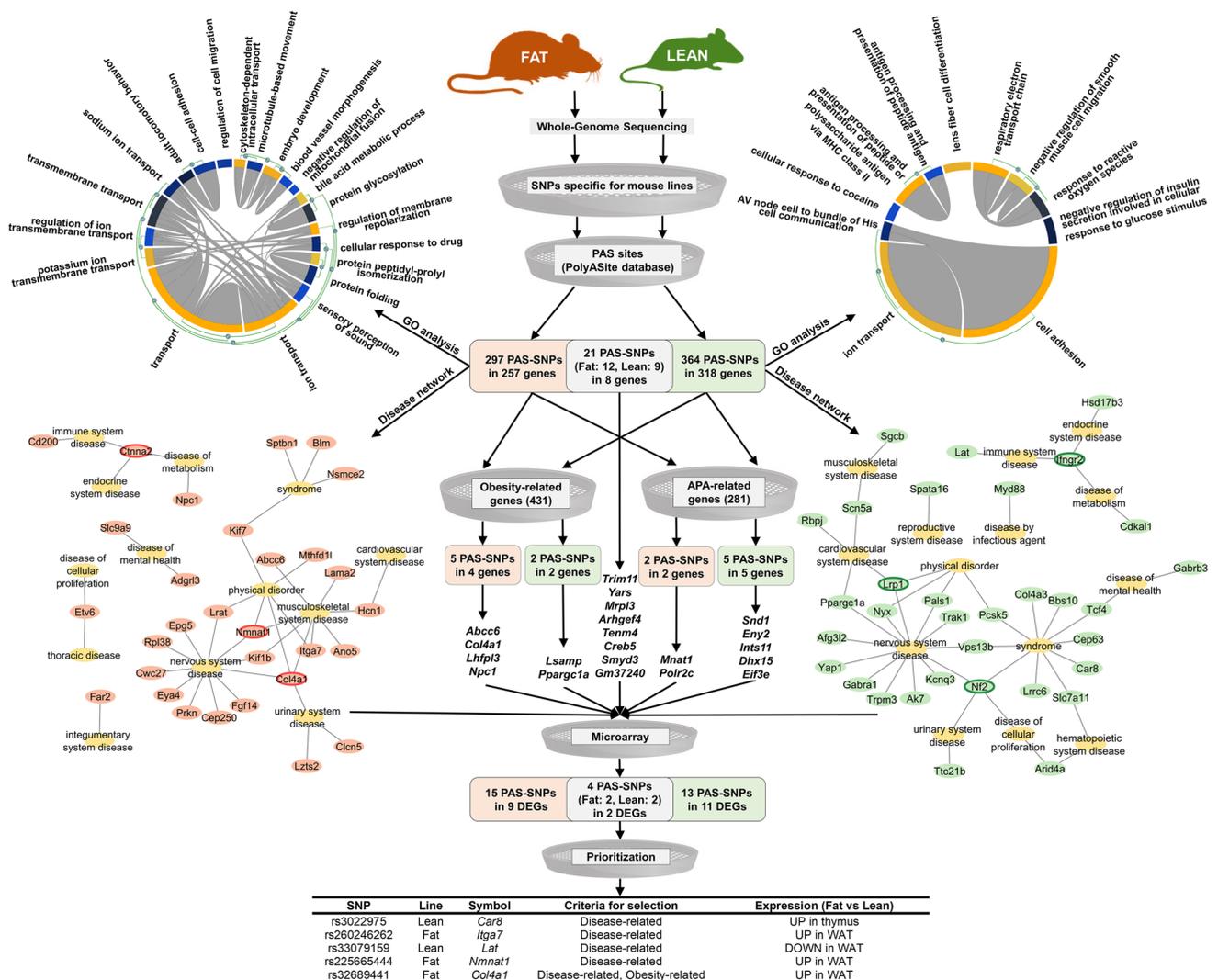


Fig. 2 Workflow and main results. Steps include Genotyping of Fat and Lean mouse selection lines, identification of PAS-SNPs, GO enrichment analysis by MonaGO, identification of disease-related genes (MGI) and construction of gene-disease interaction network

also alter the PAS sites to more or less abundant PAS hexamers or cause their complete loss according to the previously annotated PAS (Supplementary Fig. S2). For example, in the *Tenn4* gene, a PAS-SNP rs13459552 of the Fat line located in the 3' UTR alters the AGTAAA to a more important canonical AATAAA. Meanwhile, in the Lean line, an intronic PAS-SNP rs51720083 causes a complete loss of PAS (Supplementary Fig. S5).

Another interesting example is *Mrpl3*. While the SNP rs259336677 in the Fat line causes loss in PAS (AAGAAA > AAGAAG), the SNP rs238373229 alters AATATA to the canonical AATAAA of the APA site 9:105,077,566: +. Meanwhile, in the Lean line, SNP rs50804946 causes loss of one of the two PASs (TAT AAA > TATAAG) of the APA site 9:105,077,462: +,

visualized by Cytoscape, identification of common (genes containing PAS-SNPs in both lines), and obesity and APA-related genes having PAS-SNPs identified in Fat or Lean line

which is the most frequently observed APA site of *Mrpl3*, occurring in 79.5% of *Mrpl3* alternative polyadenylation events, according to PolyASite (Supplementary Fig. S6).

Even more PAS-SNPs were identified in *Smyd3* (Fat: 4, Lean: 1). Interestingly, in the Fat line, SNP rs241932144 in 3' UTR causes loss of PAS (ATTAAA > ATAAAA) of mostly used APA site (1:178,957,259: -) within *Smyd3*, occurring in 86.3% of alternative polyadenylation events (PolyASite). Meanwhile, the SNP rs220155151 alters a less used PAS AATACA to a more used AATATA of APA site located downstream (1:178,956,062: -). At the same time, other SNPs detected in the Fat line alter PAS to a less used hexamer or cause its loss. Changes in PAS hexamers in 3' UTR may influence the length of the *Smyd3* transcripts and consequently affect miRNA binding (Fig. 5).

Table 1 SNPs identified in the Fat and Lean mouse lines at orthologous positions of sequence variants in other species

Mouse		Other species				Human	
Variant ID	Gene symbol	Species	Variant ID	Variant type	Gene symbol	APA ID (PolyASite)	Variant location relative to PAS and APA (PolyASite)
rs36810866	<i>A830018L16Rik</i>	Human	rs574671612	SNP	<i>C8orf34</i> (ortholog of <i>A830018L16Rik</i>)		
rs30860267	<i>Kdm5b</i>	Human	rs921113807	SNP	<i>KDM5B</i>		
rs250253373	<i>Dner</i>	Human	rs1283834778	SNP	<i>DNER</i>		
rs30937553	<i>Gm40770</i>	Human	rs918141202	SNP	/		
rs52039424	<i>1700025F24Rik</i>	Human	rs887567006	SNP	Upstream of <i>ENSG00000257083</i>		
rs218957174	<i>Mthfd11</i>	Human	rs1360996180	SNP	<i>MTHFD1L</i>		
rs25557469	<i>Abca5</i>	Human	rs1281346561	SNP	<i>ABCA5</i>	17:69,247,483:+	Upstream of known PAS (45 bp from APA) in antisense direction
rs224932127	<i>Asic2</i>	Human	rs1400720214	SNP	<i>ASIC2</i>		
rs48386684	<i>Wfikkn2</i>	Human	rs1017487641	SNP	<i>WFIKKN2</i>	17:50,842,346:+	Between 2 PAS
rs4208259	<i>Cxadr</i>	Human	rs943743662	SNP	<i>CXADR</i>	21:17,569,608+	Between 2 PAS
rs1132123542	<i>Snx24</i>	Human	rs557040637	SNP	<i>SNX24</i>	5:123,009,226:+	In PAS
rs47475774	<i>Ccna2</i>	Human	rs755179571	SNP	<i>CCNA2</i>		
rs250712171	<i>Gm45604</i>	Pig	rs703127080	SNP	<i>SLC16A7</i>		
rs26901910	<i>Nf2</i>	Sheep	rs599786180	SNP	<i>NF2</i>		
rs51031107	<i>1700060O08Rik</i>	Horse	rs68603952	SNP	/		
rs30155538	<i>Mcur1</i>	Horse	rs782875901	SNP	Downstream of <i>MCURI</i>		
rs243180722	<i>Lsamp</i>	Goat	rs671877994	SNP	/		
rs30463516	<i>Tspan2os</i>	Human	rs1169250871	SNP	/		
rs51812848	<i>Grid1</i>	Human	rs111881471	indel	<i>GRID1</i>		
rs32419944	<i>Slc38a2</i>	Human	rs1351333839	indel	<i>SLC38A2</i>	12:46,360,463:–	Over part of PAS
rs29870198	<i>Fbxo38</i>	Human	rs1422927536	indel	<i>FBXO38</i>	5:148,442,330:+	Over parts of 2 PAS
rs27462803	<i>Ckap2l</i>	Human	rs943325762	indel	<i>CKAP2L</i>		
rs30152485	<i>Wdr3</i>	Human	rs754569327	indel	<i>WDR3</i>	1:117,959,396:+	Within APA cluster
			rs1192320348	indel			
rs31269486	<i>Gm36569</i>	Human	rs1232100051	indel	<i>GASK1B</i> and <i>GASK1B-AS1</i>		
rs3022975	<i>Car8</i>	Human	rs1208110829	indel	<i>CA8</i>	8:60,189,629:–	Upstream of known PAS (31 bp from APA)
rs37020130	<i>Dhx15</i>	Human	rs1027076262	indel	<i>DHX15</i>	4:24,539,852:–	Within APA cluster
rs585906802	<i>Ctnna2</i>	Human	rs546967032	indel	<i>CTNNA2</i>		
rs52130811	<i>Slc35f5</i>	Human	rs554688324	indel	<i>SLC35F5</i>		
rs32088144	<i>Brinp3</i>	Human	rs1343625029	indel	<i>BRINP3</i>		
rs213215053	<i>Col4a3</i>	Human	rs1457057611	indel	<i>COL4A3</i>		
rs244556382	<i>5730522E02Rik</i>	Human	rs1464977908	indel	<i>LINC01122</i>		
rs231153192	<i>Ggh</i>	Human	rs111869558	indel	<i>GGH</i>		
			rs1421773187	indel			
			rs1429142810	indel			
			rs1474924870	indel			
			rs898026262	deletion			

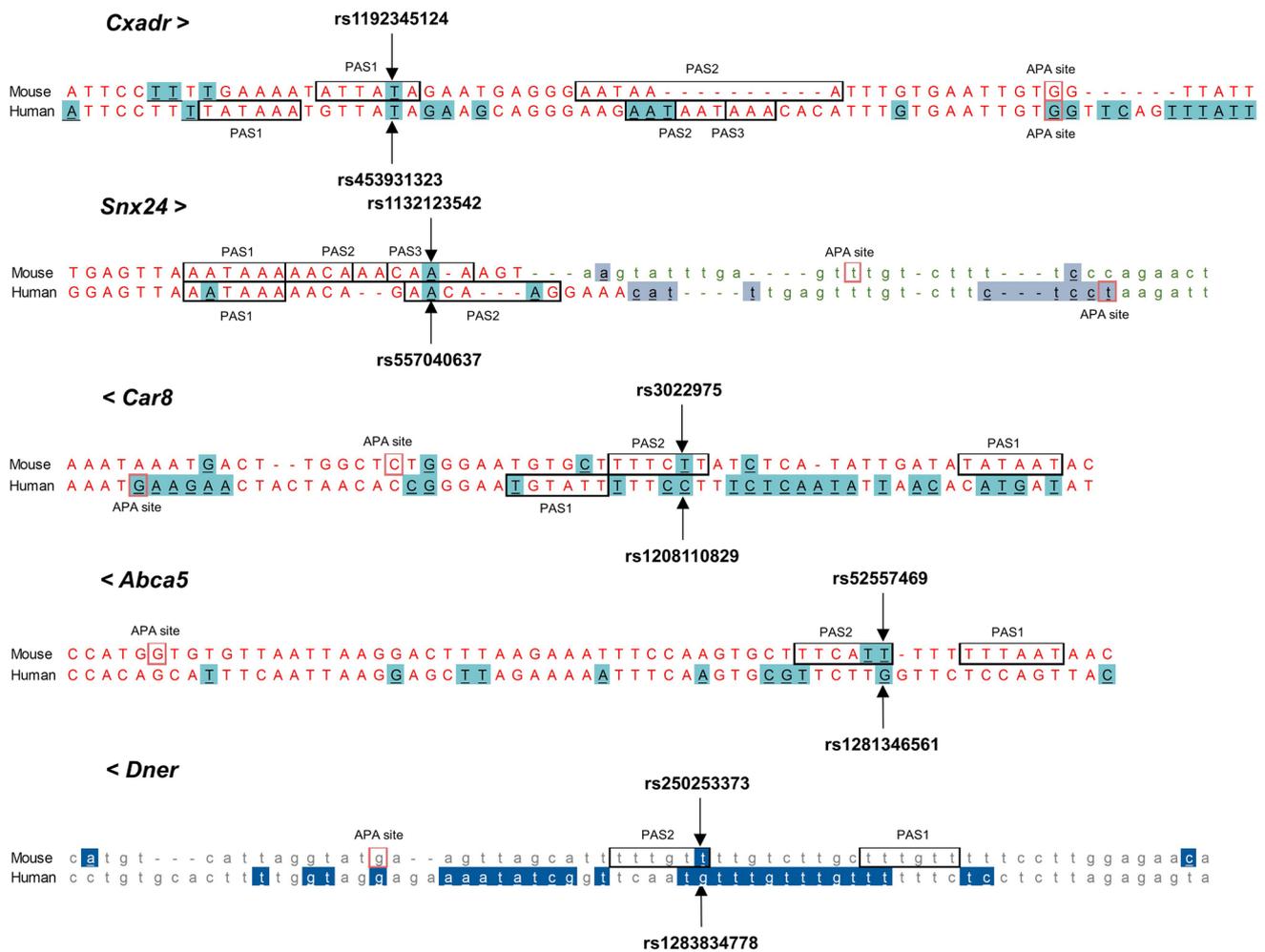


Fig. 3 Multiple-species alignment of parts of the selected genes with PAS-SNPs with marked genetic variants, PAS motifs, and APA sites. Legend: capital red letters—nucleotides in 3' UTR, capital black letters shaded light blue—3' UTR variants, small green letters—nucleo-

tides downstream of the gene, small black letters shaded grey—downstream variants, small grey letters—nucleotides in the intron, small white letters shaded blue—intronic variants

Obesity-related genes with PAS-SNPs

We then selected 431 genes that are associated with obesity and identified 5 PAS-SNPs (rs32753534, rs32689441, rs47853609, rs48552886, rs587469149) in 4 obesity-related genes in the Fat line (*Abcc6*, *Col4a1*, *Lhfp13*, *Npc1*) and 2 PAS-SNPs (rs243180722, rs38383450) in 2 obesity-related genes in the Lean line (*Lsamp*, *Ppargc1a*). All the SNPs are intronic variants of their corresponding genes. Considering the representation and usage of PAS motifs in the mouse genome described in the Supplementary Fig. S2, rs32753534 and rs38383450 change the PAS to a more used PAS hexamer. Meanwhile, the opposite was found for the remaining PAS-SNPs (rs32689441, rs47853609, rs48552886, rs587469149, rs243180722)

that cause a complete loss of previously identified PAS motifs (Table 3).

An example of PAS-SNP rs38383450 (T > A) in the first intron of *Ppargc1a* transcript ENSMUST00000127135 identified in the Lean line is given in Supplementary Fig. S7. The SNP changes TATAAA PAS hexamer to the canonical AATAAA.

APA-related genes with PAS-SNPs

We then selected 281 genes that are involved in various aspects of alternative polyadenylation. 2 SNPs (rs29182020, rs579356626) in the Fat line are in PAS of two APA-related genes (*Mnat1*, *Polr2c*), and 5 SNPs (rs33836529, rs33289253, rs227466545, rs37020130, rs33250559) in the Lean line are in PAS of 5 APA-related

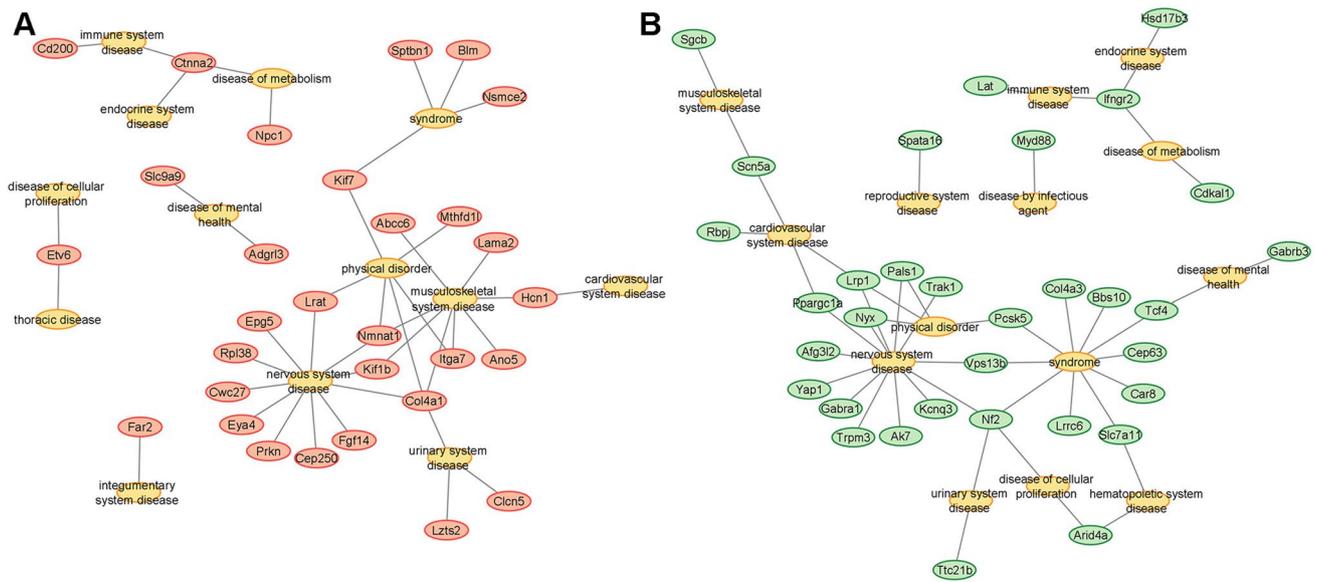


Fig. 4 Disease network of genes with PAS-SNPs identified in Fat **A** and Lean **B** mouse selection lines for body fat percentage

Table 2 Location, APA site, PAS hexamer and its change due to the SNP within common genes of Fat and Lean lines

Variant				Polyadenylation region				
Mouse line	Variant ID	Location	Gene symbol	APA ID	Type	PAS	PAS with alternative allele	Consequence on PAS
Fat	rs252244088	11:58,980,172	<i>Trim11</i>	11:58,980,198:+	IN	AATAAA	AATACA	MA → LA motif
Lean	rs29473166	11:58,979,030	<i>Trim11</i>	11:58,979,039:+	IN	TATAAA	TGTAAA	Loss
Fat	rs27517013	4:129,219,177	<i>Yars</i>	4:129,219,188:+	TE	CATAAA	CATAAG	Loss
Lean	rs3167104	4:129,197,453	<i>Yars</i>	4:129,197,470:+	IN	AATATA	AATACA	MA → LA motif
Fat	rs259336677	9:105,077,523	<i>Mrpl3</i>	9:105,077,566:+	TE	AAGAAA	AAGAAG	Loss
Fat	rs238373229	9:105,077,540	<i>Mrpl3</i>	9:105,077,566:+	TE	AATATA	AATAAA	LA → MA motif
Lean	rs50804946	9:105,077,439	<i>Mrpl3</i>	9:105,077,462:+	TE	TATAAA	TATAAG	Loss
Fat	rs581302908	1:34,809,203	<i>Arhgef4</i>	1:34,809,218:+	TE	AATAAA	AATAGA	MA → LA motif
Lean	rs46757124	1:34,691,617	<i>Arhgef4</i>	1:34,691,633:+	IN	AATAAA	AGTAAA	MA → LA motif
Fat	rs13459552	7:96,910,302	<i>Tenn4</i>	7:96,910,339:+	TE	AGTAAA	AATAAA	LA → MA motif
Lean	rs51720083	7:96,237,422	<i>Tenn4</i>	7:96,237,473:+	IN	ATTAAA	ATCAAA	Loss
Fat	rs29969996	6:53,639,550	<i>Creb5</i>	6:53,639,571:+	IN	AATGAA	AATAAA	LA → MA motif
Lean	rs29943631	6:53,297,572	<i>Creb5</i>	6:53,297,588:+	IN	ATTAAA	ATTAAC	Loss
Fat	rs220155151	1:178,956,075	<i>Smyd3</i>	1:178,956,062:-	TE	AATACA	AATATA	LA → MA motif
Fat	rs241932144	1:178,957,273	<i>Smyd3</i>	1:178,957,259:-	TE	ATTAAA	ATAAAA	Loss
Fat	rs236503723	1:179,168,920	<i>Smyd3</i>	1:179,168,907:-	IN	ATTAAA	ATTATA	MA → LA motif
Fat	rs235763389	1:179,508,325	<i>Smyd3</i>	1:179,508,286:-	TE	ATTAAA	ATTTAA	Loss
Lean	rs48210465	1:179,394,422	<i>Smyd3</i>	1:179,394,400:-	IN	AATACA	AATATA	LA → MA motif
Fat	rs30706788	3:84,809,818	<i>Gm37240</i>	3:84,809,806:-	IN	ATTAAA	GTAAA	Loss
Lean	rs258005769	3:85,755,470	<i>Gm37240</i>	3:85,755,450:-	IN	AGTAAA	AATAAA	LA → MA motif
Lean	rs30399416	3:85,789,619	<i>Gm37240</i>	3:85,789,600:-	IN	ACTAAA	ACCAAA	Loss

Legend: column Type, *IN* intronic APA site, *TE* terminal exon APA site; column Consequence, *MA* more abundant, *LA* less abundant

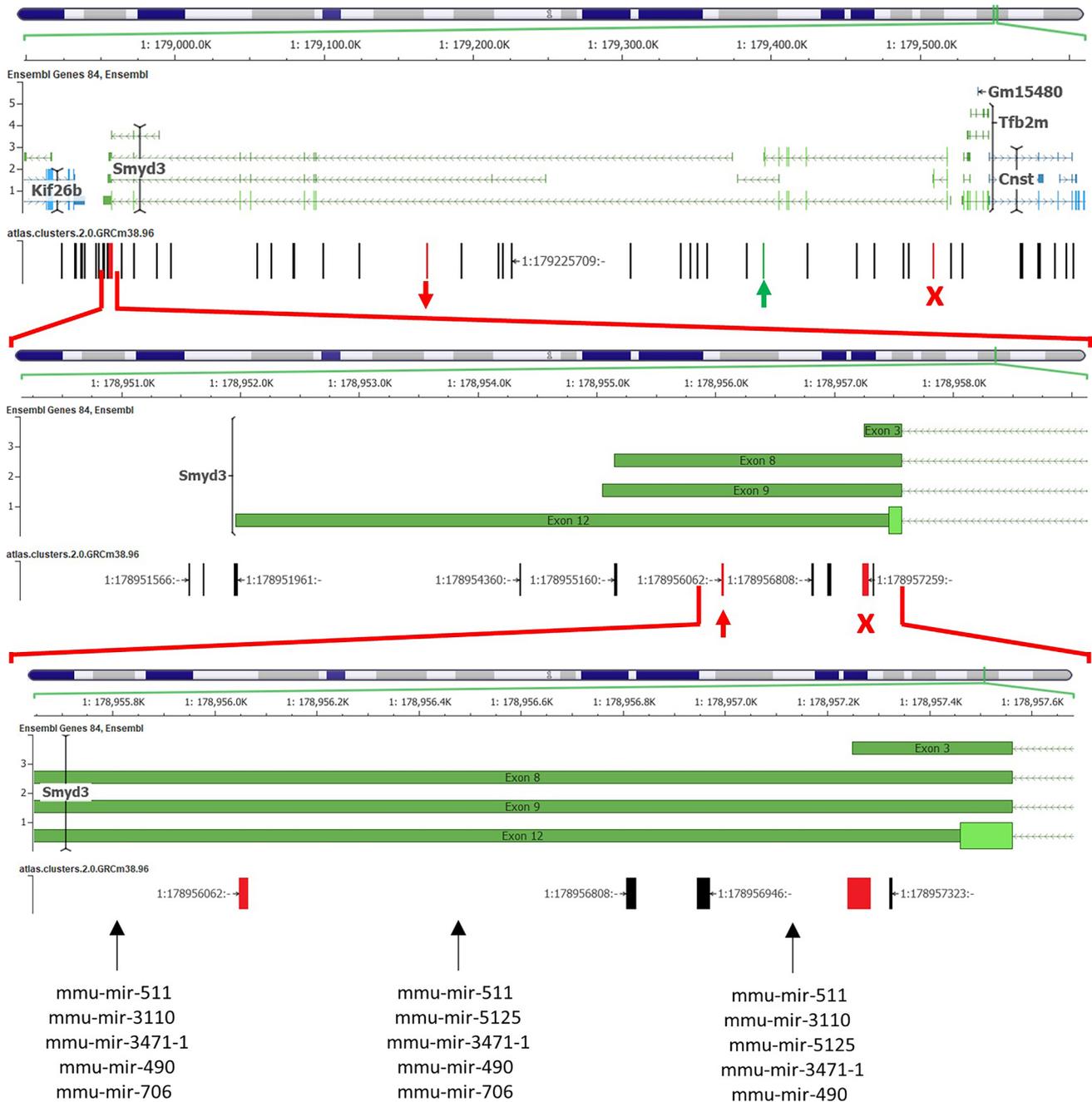


Fig. 5 PAS-SNPs within *Smyd3*. ↑—altered PAS hexamer to a more used PAS, ↓—altered PAS hexamer to a less used PAS, X—lost PAS by SNP. Red—Fat line, Green—Lean line. Approximate miRNA target locations are also indicated (miRTarBase—predicted by miRanda)

genes (*Snd1*, *Eny2*, *Ints11*, *Dhx15*, *Eif3e*). While rs579356626, rs33289253 and rs227466545 are 3' UTR variants, the remaining four SNPs are in the introns of the corresponding genes. Considering the representation and usage of PAS motifs in the mouse genome described in Supplementary Fig. S2, rs579356626, rs33836529, rs227466545 and rs37020130 change the most used PAS

to a less used PAS hexamer. Meanwhile, rs29182020, rs33289253 and rs33250559 cause a complete loss of previously identified PAS motifs (Table 4).

An example of PAS-SNP rs33289253 (A > T) in 3' UTR of *Eny2* identified in the Lean line is given in Supplementary Fig. S8. The SNP overlaps with AATAAT PAS hexamer, resulting in lost PAS.

Table 3 Location, APA site, PAS hexamer and its change due to the SNP within obesity-related genes of Fat and Lean lines

Variant				Polyadenylation region				
Mouse line	Variant ID	Location	Gene symbol	APA ID	Type	PAS	PAS with alternative allele	Consequence on PAS
Fat	rs32753534	7:45,971,020	<i>Abcc6</i>	7:45,970,994:–	IN	ATTACA	ATTAAA	LA → MA motif
Fat	rs32689441	8:11,302,646	<i>Col4a1</i>	8:11,302,636:–	IN	AAGAAA	AAGGAA	Loss
Fat	rs47853609	5:22,926,791	<i>Lhfpl3</i>	5:22,926,806:+	IN	AATGAA	AATGGA	Loss
Fat	rs48552886	5:22,926,797	<i>Lhfpl3</i>	5:22,926,806:+	IN	GATAAA	GATGAA	Loss
Fat	rs587469149	18:12,210,261	<i>Npc1</i>	18:12,210,214:–	IN	AATAGA	TATAGA	Loss
Lean	rs243180722	16:41,720,764	<i>Lsamp</i>	16:41,720,799:+	IN	AATATA	AATATT	Loss
Lean	rs38383450	5:51,561,938	<i>Ppargc1a</i>	5:51,561,914:–	IN	TATAAA	AATAAA	LA → MA motif

Legend: column Type, *IN* intronic APA site, *TE* terminal exon APA site; column Consequence, *MA* more abundant, *LA* less abundant

Table 4 Location, APA site, PAS hexamer and its change due to the SNP within obesity-related genes of Fat and Lean lines

Variant				Polyadenylation region				
Mouse line	Variant ID	Location	Gene symbol	APA ID	Type	PAS	PAS with alternative allele	Consequence on PAS
Fat	rs29182020	12:73,176,649	<i>Mnat1</i>	12:73,176,668:+	IN	AATAAA	AATTAA	Loss
Fat	rs579356626	8:94,873,520	<i>Polr2c</i>	8:94,873,526:+	TE	AATAAA	AATAAG	MA → LA motif
Lean	rs33836529	6:28,799,697	<i>Snd1</i>	6:28,799,710:+	IN	AATAAA	AATACA	MA → LA motif
Lean	rs33289253	15:44,436,556	<i>Eny2</i>	15:44,436,573:+	TE	AATAAT	AATATT	Loss
Lean	rs227466545	4:155,889,082	<i>Ints11</i>	4:155,889,098:+	TE	AATAAA	AATACA	MA → LA motif
Lean	rs37020130	5:52,159,959	<i>Dhx15</i>	5:52,159,940:–	IN	AATAAA	AGTAAA	MA → LA motif
Lean	rs33250559	15:43,268,200	<i>Eif3e</i>	15:43,268,181:–	IN	AGTAAA	AGTGAA	Loss

Legend: column Type, *IN* intronic APA site, *TE* terminal exon APA site; column Consequence, *MA* more abundant, *LA* less abundant

Expression of selected genes using microarray

In order to identify the PAS-SNPs that may influence the alternative polyadenylation and thus the expression and variability of gene transcripts, we examined the expression of candidate genes with PAS-SNPs (Disease-related: 63, Common: 8, Obesity-related: 6, APA-related: 7) using microarray analyses of various tissues. Note that *Col4a1* is found in both groups: as a core gene in the disease-related network and as an obesity-related gene. Out of 83 candidate genes with PAS-SNPs, 22 genes were differentially expressed between the Fat and Lean lines: *Car8*, *Cdkal1*, *Col4a1*, *Creb5*, *Cttna2*, *Eny2*, *Fgf14*, *Gabrb3*, *Ifngr2*, *Itga7*, *Kif1b*, *Lat*, *Lhfpl3*, *Nmnat1*, *Pcsk5*, *Ppargc1a*, *Sgcb*, *Slc9a9*, *Sptbn1*, *Tcf4*, *Tenm4* and *Trak1* (Table 5).

Prioritisation of candidate PAS-SNPs

Affymetrix probes from differentially expressed genes were then mapped to the corresponding genes along with PAS-SNPs to identify candidate PAS-SNPs. Among the 22 DEGs

carrying PAS-SNPs, PAS-SNPs of five genes were prioritised, three within *Col4a1*, *Itga7* and *Nmnat1* of the Fat line and two within *Car8* and *Lat* of the Lean line. PAS-SNPs within *Col4a1*, *Itga7* and *Lat* may affect the protein lengths encoded by these genes. Meanwhile, PAS-SNPs rs3022975 and rs225665444 within *Car8* and *Nmnat1* may result in the genes having different 3' UTR lengths in the two lines (Fig. 6A and Supplementary Fig. S9).

For the five genes, the DNA sequences within the APA cluster and 60 bp upstream were analysed to explore whether other PAS are present and if other SNPs in the two lines may create de-novo PAS of the corresponding APA site. The rs260246262 and rs33079159 within *Itga7* and *Lat* of the Fat and Lean line cause loss of less abundant motifs, whereas the canonical AATAAA and ATTAAA are still present. Similarly, rs225665444 within *Nmnat1* of the Fat line damages the second most important PAS, ATTAAA, but two canonical motifs remain untouched. In contrast, the rs3022975 within Lean line *Car8* and rs32689441 within Fat line *Col4a1* disturb the most abundant signals in the examined regions (Fig. 6B).

Table 5 Differentially expressed candidate genes between Fat and Lean lines carrying PAS-SNPs

Variant	Gene		Criteria for gene selection	Expression (Fat vs Lean)	
Variant ID	Mouse line	Gene symbol			Gene name
rs3022975	Lean	<i>Car8</i>	carbonic anhydrase 8	Disease-related ¹	UP in thymus*
rs264180954	Lean	<i>Cdkal1</i>	CDK5 regulatory subunit associated protein 1-like 1	Disease-related ²	UP in adrenal gland*, kidney*, thymus*, BAT*, WAT*, sWAT*, mWAT*, eWAT*
rs585906802	Fat	<i>Cttna2</i>	catenin (cadherin associated protein), alpha 2	Disease-related ^{2,3,4}	UP in WAT, sWAT*, eWAT*
rs46962113, rs234950047, 14_124175608_T/A	Fat	<i>Fgf14</i>	fibroblast growth factor 14	Disease-related ⁵	DOWN in thymus*, BAT*, sWAT*
rs45740960, rs47131593	Lean	<i>Gabrb3</i>	gamma-aminobutyric acid (GABA) A receptor, subunit beta 3	Disease-related ⁵	DOWN in liver*
16_91565577_A/T	Lean	<i>Ifngr2</i>	interferon gamma receptor 2	Disease-related ^{2,3,4}	UP in WAT
rs260246262	Fat	<i>Itga7</i>	integrin alpha 7	Disease-related ^{7,8}	UP in WAT*
rs32587514	Fat	<i>Kif1b</i>	kinesin family member 1B	Disease-related ^{5,7}	UP in WAT*
rs33079159	Lean	<i>Lat</i>	linker for activation of T cells	Disease-related ³	DOWN in WAT
rs225665444	Fat	<i>Nmnat1</i>	nicotinamide nucleotide adenylyltransferase 1	Disease-related ^{3,7,8}	UP in WAT
rs36933769	Lean	<i>Pcsk5</i>	proprotein convertase subtilisin/kexin type 5	Disease-related ^{1,8}	UP in adrenal gland*
rs38392118	Lean	<i>Sgcb</i>	sarcoglycan, beta (dystrophin-associated glycoprotein)	Disease-related ⁷	DOWN in kidney* and BAT*
rs36243838, rs30381123, rs234371181	Fat	<i>Slc9a9</i>	solute carrier family 9 (sodium/hydrogen exchanger), member 9	Disease-related ⁶	DOWN in WAT
rs13473736, rs262769988	Fat	<i>Sptbn1</i>	spectrin beta, non-erythrocytic 1	Disease-related ¹	UP* and DOWN in WAT*, UP in mWAT* and sWAT*, DOWN in kidney*
rs29580819	Lean	<i>Tcf4</i>	transcription factor 4	Disease-related ^{1,6}	UP in WAT
rs29593481, rs219766009	Lean	<i>Trak1</i>	trafficking protein, kinesin binding 1	Disease-related ⁵	UP in WAT*
rs32689441	Fat	<i>Col4a1</i>	collagen type IV alpha 1	Disease-related ^{5,7,8,9} , Obesity-related	UP in WAT
rs47853609, rs48552886	Fat	<i>Lhfp13</i>	lipoma HMGIC fusion partner-like 3	Obesity-related	UP in adrenal gland*
rs38383450	Lean	<i>Ppargc1a</i>	peroxisome proliferative activated receptor, gamma, coactivator 1-alpha	Obesity-related	DOWN in adrenal gland*
rs29969996, rs29943631	Fat, Lean	<i>Creb5</i>	cAMP responsive element-binding protein 5	Common	DOWN in adrenal gland*
rs13459552, rs51720083	Fat, Lean	<i>Tenn4</i>	teneurin transmembrane protein 4	Common	UP in WAT* and mWAT*, DOWN in adrenal gland*
rs33289253	Lean	<i>Eny2</i>	ENY2 transcription and export complex 2 subunit	APA-related	UP in mWAT*

Legend: column Criteria for gene selection, ¹syndrome, ²disease of metabolism, ³immune system disease, ⁴endocrine system disease, ⁵nervous system disease, ⁶disease of mental health, ⁷musculoskeletal system disease, ⁸physical disorder, ⁹urinary system disease; column Expression (Fat vs Lean), UP—higher expression of Affymetrix probe(s) in the Fat compared to the Lean line, DOWN—lower expression of Affymetrix probe(s) in the Fat compared to the Lean line, *DEG also at adjusted $p < 0.05$



Fig. 6 Visual identification of potentially functional PAS-SNPs. **A** Location of PAS-SNPs, Affymetrix probes and their expressions within *Nmnat1*. 3rd track, black rectangles—Affymetrix probes with no expression difference between the lines, red rectangles – the expression being higher in the Fat line compared to the Lean line, red vertical line denotes PAS-SNPs identified in the Fat line. **B** Location of SNPs, PAS motifs and APA sites. Legend: coloured 6 bp hex-

amers—PAS motifs (the colour denotes their abundance in mouse genome), bigger underlined nucleotides in bold—locations of SNPs, nucleotides coloured black and written white—APA site ID (most often cleavage site in this region according to PolyASite 2.0 portal), nucleotides in a black box—APA cluster (bigger the box more variable the cleavage site)

Discussion

Alternative polyadenylation (APA) has recently been established as one of the key mechanisms regulating information transfer from the genome to the phenome (Zhang et al. 2021). APA is regulated by various protein complexes that recognise *cis*-regulatory elements, among which the polyadenylation signal (PAS) is the most important component (Shulman and Elkon 2020). Studies have shown that alterations in PAS can affect the cleavage efficiency, resulting in mRNA transcripts of different lengths and thus different coding potential and proteins with altered functionality (Gebauer and Hentze 2004; Zhang et al. 2021). However, few studies have focussed on whole-genome single

nucleotide changes in PAS (PAS-SNPs) and their possible association with disease susceptibility and phenotype.

In the present study, genome-wide identification of PAS-SNPs in two divergent mouse selection lines for body fat percentage was performed as an indication of association with their phenotypic divergence. A total of 309 and 373 PAS-SNPs specific to either the Fat or Lean line were identified within 297 and 325 genes involved in various biological pathways. Pathways of cell adhesion, antigen processing, immune response, and bile acid and glucose metabolism were previously enriched in transcriptome analysis of the two mouse lines (Morton et al. 2011; Simončič et al. 2011). The other large portion of the genes with PAS-SNPs is involved in the intra- and inter-cellular

transport of ions and organic compounds. The involvement of various transport channels in obesity has also been previously reported (Vasconcelos et al. 2016; Duong et al. 2018; McCauley et al. 2020). For example, a combined effect of sodium, potassium and calcium channels remodelling on atrial fibrosis was demonstrated in a diet-induced obese mouse model (McCauley et al. 2020). In the same study, mitochondrial antioxidant therapy abrogated the ion channel and structural remodelling and alleviated atrial fibrosis (McCauley et al. 2020). Interestingly, in the present study, the genes with PAS-SNPs of the Lean line were also enriched in the respiratory electron transport chain and the response to ROS.

Since obesity can be accompanied by various diseases, we next investigated whether the human orthologous genes with PAS-SNPs identified in the two lines had been previously associated with human diseases. 30 and 33 genes were identified in the Fat and Lean lines, respectively, representing more than 10% of the genes with PAS-SNPs. In both lines, a large percentage of these genes are linked to the nervous system and physical disorders, and a cluster of genes participating in immune, endocrine and metabolic diseases were also identified. The interplay between the nervous, endocrine and immune systems in obesity has already been reviewed previously (Schwartz et al. 2017; Guarino et al. 2017), as has the relationship between musculoskeletal diseases (several genes with PAS-SNPs identified in the Fat line), metabolic syndrome and obesity (Collins et al. 2018).

For all three core genes (*Nmnat1*, *Ctnna2*, *Col4a1*) in the Fat line disease network, we found that they were more highly expressed in WAT compared to the Lean line, and rs32689441 within *Nmnat1* and especially rs225665444 within *Col4a1* were identified as priority candidate PAS-SNPs. NMNAT1 (nicotinamide nucleotide adenylyltransferase 1) synthesizes NAD(+), which is required by various enzymes. NAD(+) has been demonstrated to promote preadipocyte differentiation (Okabe et al. 2020), which might be via SIRT1 (sirtuin 1) that coordinates adipogenesis (Majeed et al. 2021). Increased adipogenesis in the Fat line may also be further linked to the higher expression level of *Ctnna2* [catenin (cadherin associated protein), alpha 2], which is consistent with Greene et al. (2021), who found that *CTNNA2* expression was up-regulated in obese individuals (Greene et al. 2021). Alpha-catenin promotes adipogenesis by suppressing Wnt signalling (Laudes 2011; Sun et al. 2014). Moreover, the enlargement of obese adipocytes due to increased Fat storage is accompanied by changes in the cytoskeleton and extracellular matrix as well as in tissue structure (Pérez-Pérez et al. 2012). At the same time, the rapidly expanding adipose tissue becomes hypoxic because the vascular system cannot develop at the same rate (Halberg et al. 2009). The extracellular matrix protein COL4A1 (collagen, type IV, alpha 1) regulates angiogenesis

(Maragoudakis et al. 1993) and was found to be up-regulated in vascular endothelial cells and subcutaneous white adipose tissue by hypoxia and the dominant-active form of the human hypoxia-inducible factor HIF1A (hypoxia-inducible factor 1 subunit alpha) (Manalo et al. 2005; Halberg et al. 2009). HIF1A is one of the master regulators of the cellular response to hypoxia (Kunej 2021), which has been proposed as one of the key reasons for adipose tissue dysfunction in obese individuals and the resulting inflammation and metabolic disorders (Trayhurn 2013). While the expression levels of *Hif1a* and *Hif2a* were significantly higher in the subcutaneous adipose tissue of the Fat line compared to those measured in the Lean line (Manalo et al. 2005; Morton et al. 2011), several potential regulatory variants in *Hif3a* have recently been identified in our mouse models (Mikec et al. 2022). Regarding angiogenesis, both lines carry PAS-SNPs in *Yars*, which encodes tyrosyl-tRNA synthetase that could act as an angiogenic factor depending on the splicing variant (Wakasugi et al. 2002). More importantly, a prioritised PAS-SNP rs260246262 has been identified within *Itga7* (integrin alpha 7), which expression is higher in WAT of the Fat line. ITGA7 transmits signals from extracellular matrix deposition and activates phosphorylation of intracellular FAK—JNK/ERK1/2 signals, promoting adipogenesis in WAT (Chen et al. 2022).

Meanwhile, in the Lean line, the core genes of the disease network are *Nf2* (neurofibromin 2), *Lrp1* (low-density lipoprotein receptor-related protein 1) and *Ifngr2* (interferon gamma receptor 2), which we found here to be differentially expressed (higher in the WAT of Fat line), but the PAS-SNPs do not seem to have any influence here on the APA. Nevertheless, IFNGR2 is part of the interferon- γ receptor complex involved in the JAK (Janus kinase)/STAT (signal transducer and activator of transcription) signalling pathway, which may contribute to adipocyte dysfunction and insulin resistance (Gurzov et al. 2016), the latter possibly occurring in our Fat line (Pirman et al. 2021). In addition, the expression level of *Lat* (linker for activation of T cells), which carries a priority candidate PAS-SNP rs33079159 in the Lean line, was lower in the Fat line WAT. Depletion of T cells has recently been observed in the adipose tissue of obese individuals (Porsche et al. 2021). It is possible that differential expression of *Ifngr2* and *Lat* contributes to the chronic, low-grade inflammation of adipose tissue that has been reported previously and links obesity to type II diabetes (Zatterale et al. 2020). Consistent with this, other disease-related genes that were found differentially expressed in WAT, although their regulation does not appear to be under the control of PAS-SNPs, were also found: *Kif1b* (kinesin family member 1B) and *Cdkal1* (CDK5 regulatory subunit associated protein 1-like 1), both of which have been previously associated with insulin sensitivity, diabetes and

obesity in humans (Steinthorsdottir et al. 2007; Palsgaard et al. 2009; Kang et al. 2020; Maruyama et al. 2022).

Also worth mentioning are four other disease-related genes carrying PAS-SNPs with differential expression in adipose tissue: *Fgf14* (fibroblast growth factor 14), *Slc9a9* [solute carrier family 9 (sodium/hydrogen exchanger), member 9], *Tcf4* (transcription factor 4) and *Trak1* (trafficking protein, kinesin binding 1), although PAS-SNPs may not be involved in regulating their expression here. Importantly, all of these genes have been linked to neurological disorders (van Spronsen et al. 2013; Forrest et al. 2014; Di Re et al. 2017; Patak et al. 2020). In addition, *Tcf4* is a member of the microphthalmia-associated transcription factor family involved in nutrient sensing and energy homeostasis (Martina et al. 2014). According to Blaszkiwicz et al. (2019), the differential expression levels of these genes may indicate alterations in the communication between brain and adipose tissue and information transfer about the adipose tissue energy status of the Fat line.

Both lines carry PAS-SNPs in *Trim11* (tripartite motif-containing 11). TRIM11 negatively regulates interferon- β which plays a crucial role in innate immunity (Lee et al. 2013). *Trim11* was not found to be differentially expressed in our study, but the tissues where this gene is mainly expressed (embryonic stages, immune cells, testis) were not analysed here. The PAS-SNPs in *Trim11* of the two lines may therefore differentially affect the transcript translational potential in these other tissues not examined yet. PAS-SNP rs36827001 in the 3' UTR of *Myd88* (myeloid differentiation primary response gene 88) of the Lean line was also identified. MYD88 signalling is critical for the control of both innate and adaptive immune responses to various central nervous system infections (Butchi et al. 2015). Recent studies have shown that MYD88 participates in and influences the COVID-19 disease severity (Cuevas et al. 2021; Mabrey et al. 2021).

There is a considerable interplay between hormones from the adrenal glands and adipose tissue (Kargi and Iacobellis 2014). Dysregulation of adrenal cortex contributes to insulin resistance and obesity onset (Roberge et al. 2007). In the present study, PAS-SNPs were identified in *Creb5* (cAMP responsive element-binding protein 5) and *Ppargc1a* (peroxisome proliferator-activated receptor gamma coactivator 1-alpha) in both lines and in the Fat line, respectively. We found that the expression levels of both transcripts were lower in the adrenals of the Fat line, but this was probably not due to the PAS-SNPs. It was found that cAMP-CREB-PGC1a signalling is involved in mitochondrial biogenesis (Xing et al. 2017), suggesting a reduced number of mitochondria in the Fat line adrenal glands. PAS-SNPs were also identified in *Mnat1*, *Mief1* and *Mrpl3* of the Fat, Lean and both lines (no difference in expression). MNAT1 (menage a trois 1) physically associates with PGC1a and is required for

its transcriptional function (Sano et al. 2007). Meanwhile, *Mief1* (mitochondrial elongation factor 1), which lies in the dominant Fob2 (Fat line obesity QTL 2) confidence interval identified in our previous study (Horvat et al. 2000), and *Mrpl3* (mitochondrial ribosomal protein L3) encode proteins involved in mitochondrial translation and biogenesis. The reduced number of mitochondria has been observed, for example, in steroidogenic tissue (Leydig cells) from mice lacking insulin and IGF1 (insulin-like growth factor 1) receptors (Radovic et al. 2019).

We also found three other genes with PAS-SNPs that are differentially expressed in the adrenal glands: *Pcsk5* (Lean line), *Tenm4* (both lines) and *Lhfp13* (Fat line), with higher, lower and higher expressions in the Fat line, again likely without the functional effect coming from their PAS-SNPs. While PCKS5 has been associated with cholesterol metabolism (Iatan et al. 2009), TENM4 (teneurin transmembrane protein 4) has been implicated in neurite development via activation of the focal adhesion kinase (FAK) signalling pathway (Suzuki et al. 2014). According to Chuang et al. (2019), decreased FAK activation can lead to cellular senescence of the adrenal gland in the Fat line. In addition, both teneurins and LHFPL3 (lipoma HMGIC fusion partner-like 3) have been shown to regulate GABA_A receptors (γ -Aminobutyric acid type A) (Yamasaki et al. 2017; Li et al. 2020), which are involved in catecholamine secretion in chromaffin cells of the adrenal medulla (Harada et al. 2016). Impaired catecholamine-mediated signalling has already been linked to obesity (Zouhal et al. 2013; Kargi and Iacobellis 2014). In short, these findings suggest adrenal dysfunction in the Fat line.

Furthermore, PAS-SNPs within *Arhgef4* [Rho guanine nucleotide exchange factor (GEF) 4] and *Smyd3* (SET and MYND domain containing 3) identified in both lines might have a differential effect on pre-mRNA processing and/or protein function as they are located in PAS hexamers of different APA sites within a given gene. In addition, PAS-SNPs have also been identified in other obesity-related genes: *Abcc6* and *Npc1* in the Fat line and *Lsamp* (limbic system-associated membrane protein) in the Lean line, as well as in *Arhgap8* (Rho GTPase activating protein 8) in the Fat line, which is localised in the dominant Fob2 (Horvat et al. 2000). However, their expression was not different between the two lines. Nevertheless, ARHGEF4, ARHGAP8 and LSAMP are involved in nervous system development and have been associated with neurobehavioral disorders (Koido et al. 2012; Gimelli et al. 2014; Cuellar Barboza et al. 2017). *Abcc6* (ATP-binding cassette, sub-family C (CFTR/MRP), member 6) and *Npc1* (NPC intracellular cholesterol transporter 1) are genes that encode membrane proteins involved in cholesterol transport and homeostasis (Fletcher et al. 2014; Kuzaj et al. 2014). For example, mutations in *NPC1* have been

linked to early-onset and morbid obesity in adulthood (Fletcher et al. 2014). The discrepancy between our mouse models in the expression of genes participating in cholesterol metabolism and transport has been noted previously (Stylianou et al. 2005; Simončič et al. 2008, 2011). In addition, studies indicate that 65% of the variation in obesity is genetic. However, genes associated with obesity risk are easy targets for epigenetic mutations (Lavebratt et al. 2012). SMYD3 (SET and MYND domain containing 3), which methylates various histone and non-histone targets (Bottino et al. 2020), is an obesity-related gene and has been proposed as one of the candidates for body mass index in pigs and humans (Zhou et al. 2016), suggesting that PAS-SNPs in *Smyd3* may influence differential epigenetic regulation between the two lines.

Finally, we investigated whether PAS-SNPs are also found in the genes involved in APA. In addition to *Mnat1* mentioned above, seven other genes (Fat line: 2, Lean line:5) with PAS-SNPs were identified: *Rbms2* (RNA-binding motif, single-stranded interacting protein 2), *Polr2c* [polymerase (RNA) II (DNA directed) polypeptide C] of the Fat line and *Snd1* (staphylococcal nuclease and tudor domain containing 1), *Eny2* (ENY2 transcription and export complex 2 subunit), *Ints11* (integrator complex subunit 11), *Dhx15* [DEAH (Asp-Glu-Ala-His) box polypeptide 15] and *Eif3e* (eukaryotic translation initiation factor 3, subunit E) in the Lean line. However, we only found that *Eny2* was differentially expressed (higher in the mWAT of the Fat line), but with an unlikely effect of PAS-SNP. ENY2 (ENY2 transcription and export complex 2 subunit) is a multifunctional transcription factor that links transcription to mRNA export (Gurskiy et al. 2010) and participates in processing the 3' end of transcripts (Kopytova et al. 2010). Recently, ENY2 was demonstrated to regulate the activities of multiple deubiquitinating enzymes (Atanassov et al. 2016). Interestingly, defective regulation of the ubiquitin/proteasome system in the hypothalamus of obese male mice has been proposed as an important mechanism for the progression and self-perpetuation of obesity (Ignacio-Souza et al. 2014). It is also worth highlighting two more genes with PAS-SNPs, notwithstanding the fact that their expression is not differential between the lines: *Ints11* (integrator complex subunit 6) and *Eif3e* (eukaryotic translation initiation factor 3, subunit E). INTS11 (along with INSTS6) is essential for adipocyte differentiation (Otani et al. 2013), and EIF3 has been demonstrated to associate with INTS6, which is involved in the degradation of certain mRNAs. More importantly, INTS11, a paralogue of CPSF73, interacts with INTS9, which is critical for integrator complex-mediated snRNA processing at the 3' end (Albrecht and Wagner 2012; Wu et al. 2017), suggesting, according to the findings by Otani et al. (2013), differential snRNAs-mediated cellular regulation between

the lines, which may partially explain their discrepancy in body fat percentage.

Taken together, PAS-SNPs within a substantial number of genes identified in our Fat and Lean mouse models may affect various psychological aspects of obesity, including angiogenesis, adipogenesis and adrenal gland dysfunction. Based on the PAS-SNPs locations within the corresponding genes, differences in their expression between the lines and biological functions, we propose PAS-SNPs rs225665444 (*Nmant1*), rs32689441 (*Col41a*) and rs260246262 (*Itga7*) in the Fat line and rs3022975 (*Car8*) and rs33079159 (*Lat*) in the Lean line as potential candidate functional PAS-SNPs that can affect the length of 3' UTR (*Car8*, *Nmnt1*) and the lengths and thus the functions of the encoded proteins (*Col4a1*, *Itga7* and *Lat*). In detail, the rs260246262 within *Itga7* of the Fat line disrupts one of two motifs in the region, potentially allowing the entire ITGA7 protein to be encoded. A similar thing could happen with *Lat* and *Col4a1* of the Lean and Fat lines due to the presence of rs33079159 and rs32689441, respectively. Meanwhile, the 3' UTR variant rs225665444 in the Fat line could lead to distinct 3' UTR lengths of the *Nmnt1* transcript between the lines, and thus different abundance of RBPs- and miRNAs-target motifs. The rs225665444 causes the loss of ATTAAA (second most important PAS) of the most distal and most used APA site within *Nmnt1*. Although two overlapping canonical AATAAA motifs remain in the region, the loss of ATTAAA may result in an overall decreased attractiveness of the region to the polyadenylation machinery. Looking at the expressions of Affymetrix probes within the 3' UTR of *Nmnt1*, the overall gene abundance is higher in the Fat line, but the lost ATTAAA motif by rs225665444 may favour a shorter transcript isoform. In contrast, rs3022975 destroys PAS within *Car8* 3' UTR of the Lean line, which may lead to increased expression of the shorter transcript in the Fat line.

Although our study identified PAS-SNPs in numerous genes and proposed several candidate functional PAS-SNPs within genes that likely contribute to the divergent phenotypes between the lines, we here list some limitations of the study. The lists of APA- and obesity-related genes are not complete. There are probably more studies on PAS-SNPs, but the outdated MeSH terminology potentially may not allowed us to find all the studies included in the PubMed database. We focussed on SNPs within PAS sites of the reference mouse genome. However, SNPs in the Fat and Lean lines may introduce de-novo PAS within the same and other genes. We analysed only the location of PAS-SNPs along with Affymetrix probes for the selected candidate genes to prioritise PAS-SNPs.

Based on the results of this study, some future directions could include

1. Systematic whole-genome analyses of SNPs along with Affymetrix probes could be performed to identify other candidate PAS-SNPs.
2. The possible effect of PAS-SNPs on protein lengths could be analysed by Western blot.
3. Changes in transcript lengths or abundances of transcript isoforms could be analysed by RT-qPCR.
4. The results should be combined with RNA sequencing data (e. g. whole transcriptome termini site sequencing) and proteomics to identify functional PAS-SNPs with the potential phenotypic impact.

Conclusion

In the present study, we identified 583 genes (Fat: 257, Lean: 318, both lines: 8) with PAS-SNPs in mouse selection lines for body fat percentage. A considerable portion of these genes is involved in various diseases, including obesity. In addition, PAS-SNPs were also identified in genes broadly involved in polyadenylation and 3'-end processing. The developed PAS-SNPs catalogue presents a key resource for designing future functional studies to uncover their role in APA, disease susceptibility and fat deposition in mouse.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00335-022-09967-8>.

Acknowledgements The authors also want to thank Dr Tim J. Aitman (Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom) for the help with WGS.

Author contributions MŠ and ŠM: formal analysis, Writing—original draft preparation. NMM: formal analysis, Writing—review & editing. SSA: formal analysis. JK: resources. SH and TK: conceptualization, Writing—review & editing, Supervision.

Funding The authors acknowledge the study was financially supported by the Slovenian Research Agency under the postgraduate research program Young researchers (ŠM), research program P4-0220 and research project J4-2548.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval The FLI (Fat) and FHI (Lean) selection lines have been maintained in our laboratory for more than 70 generations. All mice used in this study were maintained according to local ethical and EU regulatory guidelines under the Veterinary Administration of Republic of Slovenia permit No. U34401-23/2020/6.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in

the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agrawal N, Lawler K, Davidson CM et al (2021) Predicting novel candidate human obesity genes and their site of action by systematic functional screening in *Drosophila*. *PLoS Biol* 19:e3001255. <https://doi.org/10.1371/journal.pbio.3001255>
- Albrecht TR, Wagner EJ (2012) snRNA 3' end formation requires heterodimeric association of integrator subunits. *Mol Cell Biol* 32:1112–1123. <https://doi.org/10.1128/mcb.06511-11>
- Atanassov BS, Mohan RD, Lan X et al (2016) ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. *Mol Cell* 62:558–571. <https://doi.org/10.1016/j.molcel.2016.03.030>
- Bennett C, Brunkow M, Ramsdell F et al (2001) A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA→AAUGAA) leads to the IPEX syndrome. *Immunogenetics* 53:435–439. <https://doi.org/10.1007/s002510100358>
- Blażkiewicz M, Willows JW, Johnson CP, Townsend KL (2019) The importance of peripheral nerves in adipose tissue for the regulation of energy balance. *Biology* 8:10. <https://doi.org/10.3390/biology8010010>
- Bottino C, Peserico A, Simone C, Caretti G (2020) SMYD3: an oncogenic driver targeting epigenetic regulation and signaling pathways. *Cancers* 12:142. <https://doi.org/10.3390/cancers12010142>
- Bozeman, M. T. Golden Helix GenomeBrowse® visualization tool (version 2.X) [software]. Golden Helix, Inc. Available from <http://www.goldenhelix.com>
- Brutman JN, Zhou X, Zhang Y et al (2018) Mapping diet-induced alternative polyadenylation of hypothalamic transcripts in the obese rat. *Physiol Behav* 188:173–180. <https://doi.org/10.1016/j.physbeh.2018.01.026>
- Bünger L, Hill WG (1999) Inbred lines of mice derived from long-term divergent selection on fat content and body weight. *Mamm Genome* 10:645–648. <https://doi.org/10.1007/s003359901063>
- Bünger L, Laidlaw A, Bulfield G et al (2001) Inbred lines of mice derived from long-term growth selected lines: unique resources for mapping growth genes. *Mamm Genome* 12:678–686. <https://doi.org/10.1007/s00335001-3018-6>
- Butchi N, Kapil P, Puntambekar S et al (2015) Myd88 initiates early innate immune responses and promotes CD4 T cells during coronavirus encephalomyelitis. *J Virol* 89:9299–9312. <https://doi.org/10.1128/JVI.01199-15>
- Chang J-W, Yeh H-S, Yong J (2017) Alternative polyadenylation in human diseases. *Endocrinol Metab* 32:413. <https://doi.org/10.3803/EnM.2017.32.4.413>
- Chen H, Yan X, Sun A et al (2022) High-fat-diet-induced extracellular matrix deposition regulates integrin—FAK signals in adipose tissue to promote obesity. *Mol Nutr Food Res* 66:2101088. <https://doi.org/10.1002/mnfr.202101088>
- Chuang H-H, Wang P-H, Niu S-W et al (2019) Inhibition of FAK signaling elicits lamin A/C-associated nuclear deformity and cellular senescence. *Front Oncol* 9:22. <https://doi.org/10.3389/fonc.2019.00022>
- Collins KH, Herzog W, MacDonald GZ et al (2018) Obesity, metabolic syndrome, and musculoskeletal disease: common inflammatory

- pathways suggest a central role for loss of muscle integrity. *Front Physiol* 9:112. <https://doi.org/10.3389/fphys.2018.00112>
- Cuellar Barboza A, McElroy S, Winham S et al (2017) 448. Evidence for a role of binge eating and obesity in bipolar disorder genetic risk: genome-wide associations in PRR5-ARHGAP8 and TCF7L2. *Biol Psychiatry*. <https://doi.org/10.1016/j.biopsych.2017.02.932>
- Cuevas AM, Clark JM, Potter JJ (2021) Increased TLR/MyD88 signaling in patients with obesity: is there a link to COVID-19 disease severity? *Int J Obes* 45:1152–1154. <https://doi.org/10.1038/s41366-021-00768-8>
- Cui J, Li C, Cui X et al (2021) Shortening of HO1 3'UTRs by alternative polyadenylation suppresses adipogenesis in 3T3-L1. *J Agric Food Chem* 69:8038–8049. <https://doi.org/10.1021/acs.jafc.1c01822>
- D'Angelo CS, Varela MC, de Castro CIE et al (2018) Chromosomal microarray analysis in the genetic evaluation of 279 patients with syndromic obesity. *Mol Cytogenet* 11:14. <https://doi.org/10.1186/s13039-018-0363-7>
- da Fonseca ACP, Mastrorandi C, Johar A et al (2017) Genetics of non-syndromic childhood obesity and the use of high-throughput DNA sequencing technologies. *J Diabetes Complic* 31:1549–1561. <https://doi.org/10.1016/j.jdiacomp.2017.04.026>
- Di Re J, Wadsworth PA, Laezza F (2017) Intracellular fibroblast growth factor 14: emerging risk factor for brain disorders. *Front Cell Neurosci* 11:103. <https://doi.org/10.3389/fncel.2017.00103>
- Depristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–501. <https://doi.org/10.1038/ng.806>
- Duong M, Uno K, Nankivell V et al (2018) Induction of obesity impairs reverse cholesterol transport in ob/ob mice. *PLoS ONE* 13:e0202102. <https://doi.org/10.1371/journal.pone.0202102>
- Fahiminiya S, Al-Jallad H, Majewski J et al (2015) A polyadenylation site variant causes transcript-specific BMP1 deficiency and frequent fractures in children. *Hum Mol Genet* 24:516–524. <https://doi.org/10.1093/hmg/ddu471>
- Fletcher R, Gribben C, Ma X et al (2014) Correction: the role of the Niemann-Pick disease, type C1 protein in adipocyte insulin action. *PLoS ONE* 9:e116042. <https://doi.org/10.1371/journal.pone.0116042>
- Flores-Dorantes MT, Díaz-López YE, Gutiérrez-Aguilar R (2020) Environment and gene association with obesity and their impact on neurodegenerative and neurodevelopmental diseases. *Front Neurosci* 14:863. <https://doi.org/10.3389/fnins.2020.00863>
- Forrest MP, Hill MJ, Quantock AJ et al (2014) The emerging roles of TCF4 in disease and development. *Trends Mol Med* 20:322–331. <https://doi.org/10.1016/j.molmed.2014.01.010>
- Garin I, Edghill EL, Akerman I et al (2010) Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. *Proc Natl Acad Sci* 107:3105–3110. <https://doi.org/10.1073/pnas.0910533107>
- Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 5:827–835. <https://doi.org/10.1038/nrm1488>
- Gimelli S, Stathaki E, Béna F et al (2014) Recurrent microdeletion 2q21.1: report on a new patient with neurological disorders. *Am J Med Genet Part A* 164:801–805. <https://doi.org/10.1002/ajmg.a.36357>
- González-Muniesa P, Martínez-González M-A, Hu FB et al (2017) Obesity. *Nat Rev Dis Prim* 3:17034. <https://doi.org/10.1038/nrdp.2017.34>
- Greene MW, Abraham PT, Kuhlers PC et al (2021) Consensus molecular subtype differences linking colon adenocarcinoma and obesity revealed by a cohort transcriptomic analysis. medRxiv. <https://doi.org/10.1101/2021.08.31.21262900>
- Guarino D, Nannipieri M, Iervasi G et al (2017) The role of the autonomic nervous system in the pathophysiology of obesity. *Front Physiol* 8:665. <https://doi.org/10.3389/fphys.2017.00665>
- Gurskiy DY, Orlova AV, Kopytova DV et al (2010) Multifunctional factor ENY2 couples different stages of gene expression. *Russ J Genet* 46:1500–1503. <https://doi.org/10.1134/S1022795410120148>
- Gurzov EN, Stanley WJ, Pappas EG et al (2016) The JAK/STAT pathway in obesity and diabetes. *FEBS J*. <https://doi.org/10.1111/febs.13709>
- Halberg N, Khan T, Trujillo ME et al (2009) Hypoxia-inducible factor 1 α induces fibrosis and insulin resistance in white adipose tissue. *Mol Cell Biol* 29:4467–4483. <https://doi.org/10.1128/MCB.00192-09>
- Harada K, Matsuoka H, Fujihara H et al (2016) GABA signaling and neuroactive steroids in adrenal medullary chromaffin cells. *Front Cell Neurosci* 10:100. <https://doi.org/10.3389/fncel.2016.00100>
- Herrmann CJ, Schmidt R, Kanitz A et al (2020) PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res* 48:D174–D179. <https://doi.org/10.1093/nar/gkz918>
- Horvat S, Bünger L, Falconer VM et al (2000) Mapping of obesity QTLs in a cross between mouse lines divergently selected on fat content. *Mamm Genome* 11:2–7. <https://doi.org/10.1007/s003350010002>
- Howe KL, Achuthan P, Allen J et al (2021) Ensembl 2021. *Nucleic Acids Res* 49:D884–D891. <https://doi.org/10.1093/nar/gkaa942>
- Huang H-Y, Lin Y-C-D, Li J et al (2019) miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res* 48:D148–D154. <https://doi.org/10.1093/nar/gkz896>
- Huvenne H, Dubern B, Clément K, Poitou C (2016) Rare genetic forms of obesity: clinical approach and current treatments in 2016. *Obes Facts* 9:158–173. <https://doi.org/10.1159/000445061>
- Iatan I, Dastani Z, Do R et al (2009) Genetic variation at the proprotein convertase subtilisin/kexin type 5 gene modulates high-density lipoprotein cholesterol levels. *Circ Cardiovasc Genet* 2:467–475. <https://doi.org/10.1161/CIRCGENETICS.109.877811>
- Ignacio-Souza LM, Bombassaro B, Pascoal LB et al (2014) Defective regulation of the ubiquitin/proteasome system in the hypothalamus of obese male mice. *Endocrinology* 155:2831–2844. <https://doi.org/10.1210/en.2014-1090>
- Kang J, Guan R-C, Zhao Y, Chen Y (2020) Obesity-related loci in TMEM18, CDKAL1 and FAIM2 are associated with obesity and type 2 diabetes in Chinese Han patients. *BMC Med Genet* 21:65. <https://doi.org/10.1186/s12881-020-00999-y>
- Kargi AY, Iacobellis G (2014) Adipose tissue and adrenal glands: novel pathophysiological mechanisms and clinical applications. *Int J Endocrinol* 2014:1–8. <https://doi.org/10.1155/2014/614074>
- Koido K, Traks T, Balõtšev R et al (2012) Associations between LSAMP gene polymorphisms and major depressive disorder and panic disorder. *Transl Psychiatry* 2:e152–e152. <https://doi.org/10.1038/tp.2012.74>
- Kopytova DV, Orlova AV, Krasnov AN et al (2010) Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. *Genes Dev* 24:86–96. <https://doi.org/10.1101/gad.550010>
- Kunej T (2021) Integrative map of HIF1A regulatory elements and variations. *Genes* 12:1526. <https://doi.org/10.3390/genes12101526>
- Kuzaj P, Kuhn J, Dabisch-Ruthe M et al (2014) ABCC6- a new player in cellular cholesterol and lipoprotein metabolism? *Lipids Health Dis* 13:118. <https://doi.org/10.1186/1476-511X-13-118>
- Laudes M (2011) Role of WNT pathway in the determination of human mesenchymal stem cells into preadipocytes. *J Mol Endocrinol* 46:65–72. <https://doi.org/10.1530/JME-10-0169>

- Lavebratt C, Almgren M, Ekström TJ (2012) Epigenetic regulation in obesity. *Int J Obes* 36:757–765. <https://doi.org/10.1038/ijo.2011.178>
- Lee Y, Song B, Park C, Kwon K-S (2013) TRIM11 negatively regulates IFN β production and antiviral activity by targeting TBK1. *PLoS ONE* 8:e63255. <https://doi.org/10.1371/journal.pone.0063255>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li J, Xie Y, Cornelius S et al (2020) Alternative splicing controls teneurin-latrophilin interaction and synapse specificity by a shape-shifting mechanism. *Nat Commun* 11:2140. <https://doi.org/10.1038/s41467-020-16029-7>
- Loos RJF, Yeo GSH (2022) The genetics of obesity: from discovery to biology. *Nat Rev Genet* 23:120–133. <https://doi.org/10.1038/s41576-021-00414-z>
- Mabrey FL, Morrell ED, Wurfel MM (2021) TLRs in COVID-19: how they drive immunopathology and the rationale for modulation. *Innate Immun* 27:503–513. <https://doi.org/10.1177/17534259211051364>
- Majeed Y, Halabi N, Madani AY et al (2021) SIRT1 promotes lipid metabolism and mitochondrial biogenesis in adipocytes and coordinates adipogenesis by targeting key enzymatic pathways. *Sci Rep* 11:8177. <https://doi.org/10.1038/s41598-021-87759-x>
- Manalo DJ, Rowan A, Lavoie T et al (2005) Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. *Blood* 105:659–669. <https://doi.org/10.1182/blood-2004-07-2958>
- Mandel CR, Kaneko S, Zhang H et al (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444:953–956. <https://doi.org/10.1038/nature05363>
- Maragoudakis ME, Missirlis E, Karakiulakis GD et al (1993) Basement membrane biosynthesis as a target for developing inhibitors of angiogenesis with anti-tumor properties. *Kidney Int* 43:147–150. <https://doi.org/10.1038/ki.1993.24>
- Martina JA, Diab HI, Li H, Puertollano R (2014) Novel roles for the MiTF/TFE family of transcription factors in organelle biogenesis, nutrient sensing, and energy homeostasis. *Cell Mol Life Sci* 71:2483–2497. <https://doi.org/10.1007/s00018-014-1565-8>
- Maruyama T, Kobayashi T, Sugiura Y et al (2022) Association between serum miRNAs and gingival gene expression in an obese rat model. *J Periodontol Res* 57:502–509. <https://doi.org/10.1111/jre.12979>
- McCauley MD, Hong L, Sridhar A et al (2020) Ion channel and structural remodeling in obesity-mediated atrial fibrillation. *Circ Arrhythmia Electrophysiol* 13:e008296. <https://doi.org/10.1161/CIRCEP.120.008296>
- McKenna A, Hanna M, Banks E et al (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:12971303. <https://doi.org/10.1101/gr.107524.110>
- McLaren W, Gil L, Hunt SE et al (2016) The ensembl variant effect predictor. *Genome Biol* 17:122. <https://doi.org/10.1186/s13059-016-0974-4>
- Mikec Š, Šimon M, Morton NM et al (2022) Genetic variants of the hypoxia-inducible factor 3 alpha subunit (Hif3a) gene in the fat and lean mouse selection lines. *Mol Biol Rep*. <https://doi.org/10.1007/s11033-022-07309-0>
- Mlinar B, Marc J, Pfeifer M (2006) Molecular mechanisms of insulin resistance obesity and metabolic syndrome. *Biochem Medica*. <https://doi.org/10.11613/bm.2006.003>
- Morton NM, Nelson YB, Michailidou Z et al (2011) A stratified transcriptomics analysis of polygenic Fat And Lean mouse adipose tissues identifies novel candidate obesity genes. *PLoS ONE* 6:e23944. <https://doi.org/10.1371/journal.pone.0023944>
- Neve J, Patel R, Wang Z et al (2017) Cleavage and polyadenylation: ending the message expands gene regulation. *RNA Biol* 14:865–890. <https://doi.org/10.1080/15476286.2017.1306171>
- Nourse J, Spada S, Danckwardt S (2020) Emerging roles of rna 3'-end cleavage and polyadenylation in pathogenesis, diagnosis and therapy of human disorders. *Biomolecules* 10:915. <https://doi.org/10.3390/biom10060915>
- Okabe K, Nawaz A, Nishida Y et al (2020) NAD⁺ metabolism regulates preadipocyte differentiation by enhancing α -ketoglutarate-mediated histone H3K9 demethylation at the PPAR γ promoter. *Front Cell Dev Biol* 8:586179. <https://doi.org/10.3389/fcell.2020.586179>
- Otani Y, Nakatsu Y, Sakoda H et al (2013) Integrator complex plays an essential role in adipose differentiation. *Biochem Biophys Res Commun* 434:197–202. <https://doi.org/10.1016/j.bbrc.2013.03.029>
- Palsgaard J, Brøns C, Friedrichsen M et al (2009) Gene expression in skeletal muscle biopsies from people with type 2 diabetes and relatives: differential regulation of insulin signaling pathways. *PLoS ONE* 4:e6575. <https://doi.org/10.1371/journal.pone.0006575>
- Patak J, Faraone SV, Zhang-James Y (2020) Sodium hydrogen exchanger 9 <scp>NHE9</scp> (<scp> *SLC9A9* </scp>) and its emerging roles in neuropsychiatric comorbidity. *Am J Med Genet Part B Neuropsychiatr Genet* 183:289–305. <https://doi.org/10.1002/ajmg.b.32787>
- Pedroni SMA, Turban S, Kipari T et al (2014) Pregnancy in obese mice protects selectively against visceral adiposity and is associated with increased adipocyte estrogen signalling. *PLoS ONE* 9:e94680. <https://doi.org/10.1371/journal.pone.0094680>
- Pell N, Garcia-Pras E, Gallego J et al (2021) Targeting the cytoplasmic polyadenylation element-binding protein CPEB4 protects against diet-induced obesity and microbiome dysbiosis. *Mol Metab* 54:101388. <https://doi.org/10.1016/j.molmet.2021.101388>
- Pérez-Pérez R, García-Santos E, Ortega-Delgado FJ et al (2012) Attenuated metabolism is a hallmark of obesity as revealed by comparative proteomic analysis of human omental adipose tissue. *J Proteomics* 75:783–795. <https://doi.org/10.1016/j.jpro.2011.09.016>
- Pirman T, Mrak V, Fonseca R, Horvat S (2021) Protein digestibility and bioavailability in an F2 mouse cross between the selected fat mouse line and an M2 congenic line carrying the anti-obesity and anti-diabetic Tst allele. *J Anim Physiol Anim Nutr*. <https://doi.org/10.1111/jpn.13650>
- Porsche CE, Delproposto JB, Geletka L et al (2021) Obesity results in adipose tissue T cell exhaustion. *JCI Insight* 6:8. <https://doi.org/10.1172/jci.insight.139793>
- Radovic SM, Starovlah IM, Capo I et al (2019) Insulin/IGF1 signaling regulates the mitochondrial biogenesis markers in steroidogenic cells of prepubertal testis, but not ovary. *Biol Reprod* 100:253–267. <https://doi.org/10.1093/biolre/foy177>
- Rao KR, Lal N, Giridharan NV (2014) Genetic & epigenetic approach to human obesity. *Indian J Med Res* 140:589–603
- Roberge C, Carpentier AC, Langlois MF et al (2007) Adrenocortical dysregulation as a major player in insulin resistance and onset of obesity. *Am J Physiol - Endocrinol Metab* 293:E1465–E1478. <https://doi.org/10.1152/ajpendo.00516.2007>
- Sano M, Izumi Y, Helenius K et al (2007) Ménage-à-Trois 1 Is critical for the transcriptional function of PPAR γ coactivator 1. *Cell Metab* 5:129–142. <https://doi.org/10.1016/j.cmet.2007.01.003>
- Schwartz MW, Seeley RJ, Zeltser LM et al (2017) Obesity pathogenesis: an endocrine society scientific statement. *Endocr Rev* 38:267–296. <https://doi.org/10.1210/er.2017-00111>

- Shannon P (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sharp GL, Hill WG, Robertson A (1984) Effects of selection on growth, body composition and food intake in mice I. Responses in selected traits. *Genet Res* 43:75–92. <https://doi.org/10.1017/S0016672300025738>
- Shulman ED, Elkon R (2020) Systematic identification of functional SNPs interrupting 3'UTR polyadenylation signals. *PLOS Genet* 16:e1008977. <https://doi.org/10.1371/journal.pgen.1008977>
- Simončič M, Režen T, Juvan P et al (2008) Transcriptome analysis revealed association of some P450 genes with obesity in a polygenic obese mouse model. *Acta Chim Slov* 55:101–110
- Simončič M, Režen T, Juvan P et al (2011) Obesity resistant mechanisms in the Lean polygenic mouse model as indicated by liver transcriptome and expression of selected genes in skeletal muscle. *BMC Genomics* 12:96. <https://doi.org/10.1186/1471-2164-12-96>
- Steinhorsdóttir V, Thorleifsson G, Reynisdóttir I et al (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 39:770–775. <https://doi.org/10.1038/ng2043>
- Stylianou IM, Clinton M, Keightley PD et al (2005) Microarray gene expression analysis of the Fob3b obesity QTL identifies positional candidate gene Sqle and perturbed cholesterol and glycolysis pathways. *Physiol Genomics* 20:224–232. <https://doi.org/10.1152/physiolgenomics.00183.2004>
- Sun Y, Zhang J, Ma L (2014) α -catenin. *Cell Cycle* 13:2334–2339. <https://doi.org/10.4161/cc.29765>
- Suzuki N, Numakawa T, Chou J et al (2014) Teneurin-4 promotes cellular protrusion formation and neurite outgrowth through focal adhesion kinase signaling. *FASEB J* 28:1386–1397. <https://doi.org/10.1096/fj.13-241034>
- Swami M (2009) A modular approach to signalling. *Nat Rev Genet* 10:348–349. <https://doi.org/10.1038/nrg2595>
- Tian B, Manley JL (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* 18:18–30. <https://doi.org/10.1038/nrm.2016.116>
- Trayhurn P (2013) Hypoxia and adipose tissue function and dysfunction in obesity. *Physiol Rev* 93:1–21. <https://doi.org/10.1152/physrev.00017.2012>
- Van der Auwera GA, Carneiro MO, Hartl C et al (2013) From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma* 43:483–492. <https://doi.org/10.1002/0471250953.bi1110s43>
- van Spronsen M, Mikhaylova M, Lipka J et al (2013) TRAK/Milton motor-adaptor proteins steer mitochondrial trafficking to axons and dendrites. *Neuron* 77:485–502. <https://doi.org/10.1016/j.neuron.2012.11.027>
- Vasconcelos LHC, Souza ILL, Pinheiro LS, Silva BA (2016) Ion channels in obesity: pathophysiology and potential therapeutic targets. *Front Pharmacol* 7:58. <https://doi.org/10.3389/fphar.2016.00058>
- Wakasugi K, Slike BM, Hood J et al (2002) Induction of angiogenesis by a fragment of human tyrosyl-tRNA synthetase. *J Biol Chem* 277:20124–20126. <https://doi.org/10.1074/jbc.C200126200>
- Wang Y, Wu X-S, He J et al (2016) A novel TP53 variant (rs78378222 A > C) in the polyadenylation signal is associated with increased cancer susceptibility: evidence from a meta-analysis. *Oncotarget*. <https://doi.org/10.18632/oncotarget.9056>
- Wu Y, Albrecht TR, Baillat D et al (2017) Molecular basis for the interaction between integrator subunits IntS9 and IntS11 and its functional importance. *Proc Natl Acad Sci* 114:4394–4399. <https://doi.org/10.1073/pnas.1616605114>
- Xiao R, Li C, Wang C et al (2019) Adipogenesis associated Mth938 domain containing (AAMDC) protein expression is regulated by alternative polyadenylation and microRNAs. *FEBS Lett* 593:1724–1734. <https://doi.org/10.1002/1873-3468.13449>
- Xin Z, Cai Y, Dang LT et al (2020) MonaGO: a novel gene ontology enrichment analysis visualisation system. *bioRxiv*. <https://doi.org/10.1101/2020.09.27.316067>
- Xing F, Luan Y, Cai J et al (2017) The anti-warburg effect elicited by the cAMP-PGC1 α pathway drives differentiation of glioblastoma cells into astrocytes. *Cell Rep* 18:468–481. <https://doi.org/10.1016/j.celrep.2016.12.037>
- Yamasaki T, Hoyos-Ramirez E, Martenson JS et al (2017) GARLH family proteins stabilize GABAA receptors at synapses. *Neuron* 93:1138–1152.e6. <https://doi.org/10.1016/j.neuron.2017.02.023>
- Yuan F, Hankey W, Wagner EJ et al (2021a) Alternative polyadenylation of mRNA and its role in cancer. *Genes Dis* 8:61–72. <https://doi.org/10.1016/j.gendis.2019.10.011>
- Yuan Z, Ge L, Sun J et al (2021b) Integrative analysis of Iso-Seq and RNA-seq data reveals transcriptome complexity and differentially expressed transcripts in sheep tail fat. *PeerJ* 9:e12454. <https://doi.org/10.7717/peerj.12454>
- Zatterale F, Longo M, Naderi J et al (2020) Chronic adipose tissue inflammation linking obesity to insulin resistance and type 2 diabetes. *Front Physiol*. <https://doi.org/10.3389/fphys.2019.01607>
- Zhang Y, Liu L, Qiu Q et al (2021) Alternative polyadenylation: methods, mechanism, function, and role in cancer. *J Exp Clin Cancer Res* 40:51. <https://doi.org/10.1186/s13046-021-01852-7>
- Zhou L, Ji J, Peng S et al (2016) A GWA study reveals genetic loci for body conformation traits in Chinese Laiwu pigs and its implications for human BMI. *Mamm Genome* 27:610–621. <https://doi.org/10.1007/s00335-016-9657-4>
- Zouhal H, Lemoine-Morel S, Mathieu M-E et al (2013) Catecholamines and obesity: effects of exercise and training. *Sport Med* 43:591–600. <https://doi.org/10.1007/s40279-013-0039-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.