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. Atanur3,4 · 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
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 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 IIIm complex (CFIIIm) 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 AAUAAA to AAU GAA 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 adipogene- sis-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
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 (FHI, 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 AATTTA PAS motif being the most abundant, followed by ATTTAA (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, Tenm4, 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), Ctnna2 (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, Tenm4, 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
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 Tenm4 gene, a PAS-SNP rs13459552 of the Fat line located in the 3′UTR alters the AGT AAA to a more important canonical AAT AAA. 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 (TATAAA > TATAAG) of the APA site 9:105,077,462: +, 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 AAT ACA to a more used AAT ATA 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).

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 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.
| Mouse Variant ID | Gene symbol | Mouse Species | Other species Variant ID | Variant type | Other species Gene symbol | Human APA ID (PolyASite) | Variant location relative to PAS and APA (PolyASite) | Human Gene symbol |
|-----------------|-------------|---------------|-------------------------|--------------|---------------------------|--------------------------|--------------------------------------------------|------------------|
| rs36810866      | A830018L16Rik | Human         | rs574671612              | SNP          | C8orf34 (ortholog of A830018L16Rik) |                         |                                                                  |                  |
| rs30860267      | Kdm5b       | Human         | rs921113807              | SNP          | KDM5B                     |                         |                                                                  |                  |
| rs250253373     | Dmer        | Human         | rs1283834778             | SNP          | DNER                      |                         |                                                                  |                  |
| rs30937553      | Gm40770     | Human         | rs918141202              | SNP          | /                         |                         |                                                                  |                  |
| rs52039424      | 1700025F24Rik | Human       | rs887567006              | SNP          | Upstream of ENSG00000257083 |                         |                                                                  |                  |
| rs21895714      | Mthfd1l     | Human         | rs1360996180             | SNP          | MTHFD1L                   |                         | Upstream of known PAS (45 bp from APA) in antisense direction |                  |
| rs52557469      | Abca5       | Human         | rs1281346561             | SNP          | ABCA5                     | 17:69,247,483:+         |                                                                  |                  |
| rs224932127     | Asic2       | Human         | rs1400720214             | SNP          | ASIC2                     |                         |                                                                  |                  |
| rs483386684     | Wjikk2      | Human         | rs1017487641             | SNP          | WFIKKN2                   | 17:50,842,346:+         | Between 2 PAS                                                   |                  |
| rs4208259       | Cxadr       | Human         | rs943743662              | SNP          | CXADR                     | 21:17,569,608:+         | Between 2 PAS                                                   |                  |
| rs1132123542    | Ssx24       | Human         | rs557040637              | SNP          | SNX24                      | 5:123,009,226:+         | In PAS                                                          |                  |
| rs47475774      | Ccna2       | Human         | rs755179571              | SNP          | CCNA2                      |                         |                                                                  |                  |
| rs250712171     | Gm45604     | Pig           | rs703127080              | SNP          | SLC16A7                    |                         |                                                                  |                  |
| rs26901910      | Nf2         | Sheep         | rs599786180              | SNP          | NF2                        |                         |                                                                  |                  |
| rs51031107      | 17000660008Rik | Horse       | rs68603952               | SNP          | /                         |                         |                                                                  |                  |
| rs30155538      | Mcur1       | Horse         | rs782875901              | SNP          | Downstream of MCUR1       |                         |                                                                  |                  |
| rs243180722     | Lsamp       | Goat          | rs671877994              | SNP          | /                         |                         |                                                                  |                  |
| rs30463516      | Tspan2os    | Human         | rs1169250871             | SNP          | /                         |                         |                                                                  |                  |
| rs51812848      | Grid1       | Human         | rs111881471              | indel        | GRID1                      |                         |                                                                  |                  |
| rs32419944      | Slc38a2     | Human         | rs1351333839             | indel        | SLC38A2                    |                         |                                                                  |                  |
| rs29870198      | Fbxo38      | Human         | rs1422927536             | indel        | FBXO38                     | 12:46,360,463:−         | Over part of PAS                                                  |                  |
| rs27462803      | Ckap2l      | Human         | rs943325762              | indel        | CKAP2L                     | 5:148,442,330:+         | Over parts of 2 PAS                                                |                  |
| rs30152485      | Wdr3        | Human         | rs754569327              | indel        | WDR3                       | 1:117,959,396:+         | Within APA cluster                                                 |                  |
| rs31269486      | Gm36569     | Human         | rs1232100051             | indel        | GASK1B and GASK1B-AS1     |                         |                                                                  |                  |
| rs3022975       | Car8        | Human         | rs1208110829             | indel        | CA8                        | 8:60,189,629:−          | Upstream of known PAS (31 bp from APA)                        |                  |
| rs37020130      | Dhx15       | Human         | rs1027076262             | indel        | DHX15                      | 4:24,539,852:−          | Within APA cluster                                                 |                  |
| rs585906802     | Ctnna2      | Human         | rs546967032              | indel        | CTNNA2                     |                         |                                                                  |                  |
| rs52130811      | Slc35f5     | Human         | rs554688324              | indel        | SLC35F5                    |                         |                                                                  |                  |
| rs32088144      | Brinp3      | Human         | rs1343625029             | indel        | BRINP3                     |                         |                                                                  |                  |
| rs213215053     | Col4a3      | Human         | rs1457057611             | indel        | COL4A3                     |                         |                                                                  |                  |
| rs244556382     | 5730522E02Rik | Human       | rs1464977908             | indel        | LINCO1122                  |                         |                                                                  |                  |
| rs231153192     | Ggh         | Human         | rs111869558              | indel        | GGH                        |                         |                                                                  |                  |
Genome-wide screening for genetic variants in polyadenylation signal (PAS) sites in mouse...

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, LhfpL3, 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 AAT AAA.

Obesity-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 | Mouse line | Variant ID | Location | Gene symbol | Polyadenylation region | Consequence on PAS |
|---------|------------|------------|----------|-------------|------------------------|-------------------|
| Fat     | rs252244088| 11:58,980,172| Trim11   | AATAA      | 11:58,980,198:+        | MA → LA motif     |
| Lean    | rs29473166 | 11:58,979,030| Trim11   | TATAA      | 11:58,979,039:+        | MA → LA motif     |
| Fat     | rs27517013 | 4:129,219,177| Yars     | AATAA      | 4:129,219,188:+        | MA → LA motif     |
| Lean    | rs3167104  | 4:129,197,453| Yars     | TATAA      | 4:129,197,440:+        | Loss              |
| Fat     | rs259336677| 9:105,077,523| Mrpl3    | ACTAAA     | 9:105,077,566:+        | Loss              |
| Fat     | rs238373229| 9:105,077,540| Mrpl3    | AATAA      | 9:105,077,566:+        | Loss              |
| Lean    | rs30804946 | 9:105,077,439| Mrpl3    | AATAA      | 9:105,077,462:+        | Loss              |
| Fat     | rs581302908| 1:34,809,203| Arhgef4  | AATAA      | 1:34,809,218:+         | MA → LA motif     |
| Lean    | rs46757124 | 1:34,691,617| Arhgef4  | TATAA      | 1:34,691,633:+         | MA → LA motif     |
| Fat     | rs13459552 | 7:96,910,302| Tenm4    | AGTAAA     | 7:96,910,339:+         | Loss              |
| Lean    | rs51720083 | 7:96,237,422| Tenm4    | ATAAA      | 7:96,237,473:+         | Loss              |
| Fat     | rs29969996 | 6:53,639,550| Creb5    | ATGAAA     | 6:53,639,571:+         | Loss              |
| Lean    | rs29943631 | 6:53,297,572| Creb5    | AATAA      | 6:53,297,577:+         | Loss              |
| Fat     | rs220155151| 1:178,957,273| Smyd3   | AATAA      | 1:178,957,273         | Loss              |
| Fat     | rs241932144| 1:178,957,273| Smyd3   | AATAA      | 1:178,957,259         | Loss              |
| Fat     | rs236503723| 1:179,168,920| Smyd3   | AATAA      | 1:179,168,907         | Loss              |
| Fat     | rs235763389| 1:179,508,325| Smyd3   | AATAA      | 1:179,508,286         | Loss              |
| Lean    | rs48214065 | 1:179,394,422| Smyd3   | AATAA      | 1:179,394,400         | Loss              |
| Fat     | rs30706788 | 3:84,809,818| Gm37240  | ATTAAA     | 3:84,809,860          | Loss              |
| Lean    | rs258005769| 3:85,755,470| Gm37240  | AATAA      | 3:85,755,450          | Loss              |
| Lean    | rs30399416 | 3:85,789,619| Gm37240  | AATAA      | 3:85,789,600          | Loss              |

Legend: column Type, IN intronic APA site, TE terminal exon APA site; column Consequence, MA more abundant, LA less abundant
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.
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, Cnma2, 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 rs26024626 and rs33079159 within Itga7 and Lat of the Fat and Lean line cause loss of less abundant motifs, whereas the canonical AATCAA and ATTTAA are still present. Similarly, rs225665444 within Nmnat1 of the Fat line damages the second most important PAS, ATTTAA, 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).
| Variant ID | Mouse line | Gene symbol | Gene name | Criteria for gene selection | Expression (Fat vs Lean) |
|-----------|------------|-------------|-----------|-----------------------------|--------------------------|
| rs3022975 | Lean       | Carb8       | carbonic anhydrase 8 | Disease-related\(^1\) | UP in thymus\(^*\) |
| rs264180954 | Lean | Cdkal1 | CDK5 regulatory subunit 1-like | Disease-related\(^2\) | UP in adrenal gland\(^*\), kidney\(^*\), BAT\(^*\), WAT\(^*\), sWAT\(^*\), mWAT\(^*\), eWAT\(^*\) |
| rs585906802 | Fat       | Ctnna2      | catenin (cadherin-associated protein), alpha 2 | Disease-related\(^2,3,4\) | UP in WAT, sWAT\(^*\), eWAT\(^*\) |
| rs46962113, rs234950047, 14_124175608_T/A | Fat | Fgf14 | fibroblast growth factor 14 | Disease-related\(^5\) | DOWN in thymus\(^*\), BAT\(^*\), sWAT\(^*\) |
| rs45740960, rs47131593 | Lean | Gabrb3 | gamma-aminobutyric acid (GABA) A receptor, subunit beta 3 | Disease-related\(^6\) | DOWN in liver\(^*\) |
| 16_91565577_A/T | Lean | Ifngr2 | interferon gamma receptor 2 | Disease-related\(^2,3,4\) | UP in WAT |
| rs260246262 | Fat | Itga7 | integrin alpha 7 | Disease-related\(^7,8\) | UP in WAT\(^*\) |
| rs32587514 | Fat | Kif1b | kinesin family member 1B | Disease-related\(^2,5,7\) | UP in WAT\(^*\) |
| rs33079159 | Lean | Lat | linker for activation of T cells | Disease-related\(^5\) | DOWN in WAT |
| rs225665444 | Fat | Nmnat1 | nicotinamide nucleotide adenyltransferase 1 | Disease-related\(^3,7,8\) | UP in WAT |
| rs36933769 | Lean | Pcsk5 | proprotein convertase subtilisin/kexin type 5 | Disease-related\(^7,8\) | UP in adrenal gland\(^*\) |
| rs38392118 | Lean | Sgcb | sarcoglycan, beta (dystrophin-associated glycoprotein) | Disease-related\(^7\) | DOWN in kidney\(^*\) and BAT\(^*\) |
| rs36243838, rs30381123, rs234371181 | Fat | Slc9a9 | solute carrier family 9 (sodium/hydrogen exchanger), member 9 | Disease-related\(^6\) | DOWN in WAT |
| rs13473736, rs262769988 | Fat | Sptbn1 | spectrin beta, non-erythrocytic 1 | Disease-related\(^7\) | UP* and DOWN in WAT\(^*\), UP in mWAT\(^*\) and sWAT\(^*\), DOWN in kidney\(^*\) |
| rs29580819 | Lean | Tcf4 | transcription factor 4 | Disease-related\(^1,6\) | UP in WAT |
| rs29593481, rs219766009 | Lean | Trak1 | trafficking protein, kinesin binding 1 | Disease-related\(^8\) | UP in WAT\(^*\) |
| rs32689441 | Fat | Col4a1 | collagen type IV alpha 1 | Disease-related\(^7,8,9\), Obesity-related | UP in WAT |
| rs47853609, rs48552886 | Fat | Lhpl3 | lipoma HMGIC fusion partner-like 3 | Obesity-related | UP in adrenal gland\(^*\) |
| rs38383450 | Lean | Ppargc1a | peroxisome proliferative activated receptor, gamma, coactivator 1-alpha | Obesity-related | DOWN in adrenal gland\(^*\) |
| rs29969996, rs29943631 | Fat, Lean | Creb5 | cAMP responsive element-binding protein 5 | Common | DOWN in adrenal gland\(^*\) |
| rs13459552, rs51720083 | Fat, Lean | Tenm4 | teneurin transmembrane protein 4 | Common | UP in WAT\(^*\) and mWAT\(^*\), DOWN in adrenal gland\(^*\) |
| rs33289253 | Lean | Eny2 | ENY2 transcription and export complex 2 subunit | APA-related | UP in mWAT\(^*\) |

Legend: column Criteria for gene selection,\(^1\)syndrome, \(^2\)disease of metabolism, \(^3\)immune system disease, \(^4\)endocrine system disease, \(^5\)nervous system disease, \(^6\)disease of mental health, \(^7\)musculoskeletal system disease, \(^8\)physical disorder, \(^9\)urinary 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 \(p < 0.05\)
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 (Nnnat1, 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 Nnnat1 and especially rs225665444 within Col4a1 were identified as priority candidate PAS-SNPs. NMNAT1 (nicotinamide nucleotide adenyltransferase 1) synthesizes NAD(+), which is required by various enzymes. NAD(+)-dependent enzymes have 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 Blaszkiewicz 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 Pparc1a (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-CREBP-GIC1a 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 Mmnt1, 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 Mrp13 (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), Tennm4 (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 LHFP3 (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 Muat1 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), Dlx15 (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 Eif2 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 (Gurski 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 parologue 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 (Nmnt1), rs326894441 (Col4a1a) and rs260246262 (Iga7) 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, Iga7 and Lat). In detail, the rs260246262 within Iga7 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 rs32689444, 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** MS and ŠM: formal analysis, Writing—original draft preparation. NMM: formal analysis, Writing—review & editing. SSA: formal analysis. JK: resources. SH and TK: conceptualization, draft preparation. NMM: formal analysis, Writing—review & editing.

**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, Brunkowski 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/s002510030358

Blaszkiewicz 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, 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/s0033501-3018-6

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 UTRs by alteration...
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 (2003) 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 (rs7837222 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, Baillart 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-PGC1α pathway drives differentiation of gobliostoma 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/10.1101/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

Zheng 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 conformations in sheep: evidence from a meta-analysis. Oncotarget. 13:e1608977. https://doi.org/10.1371/journal.pgen.100895

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