Homeodomain transcription factors in adipocyte thermogenesis: insights into the species-specific and conserved regulatory elements of human UCP1

Beáta B. Tóth[1*], Géza Hegedűs[1,2], Eszter Virág[1,3], Gergely Tamás Pethő[1], Levente Laczkó[1,4,] László Fésüs[5]

[1]"One Health" Institute, Faculty of Health Science, University of Debrecen, Debrecen, H-4032, Debrecen, Egyetem Tér 1., Hungary (toth.beata@etk.unideb.hu)

[2] Department of Information Technology and Its Applications, Faculty of Information Technology, University of Pannonia, Gasparich Márk Str 18/A, 8900 Zalaegerszeg, Hungary

[3] Research Institute for Medicinal Plants and Herbs Ltd., Budakalász, Hungary

[4] HUN-REN-DE Conservation Biology Research Group, University of Debrecen, Debrecen, H-4032, Egyetem Tér 1, Hungary

[5]Biochemistry and Molecular Biology Department, Faculty of Medical Science, University of Debrecen; H-4032, Debrecen, Egyetem Tér 1., Hungary

*Corresponding author

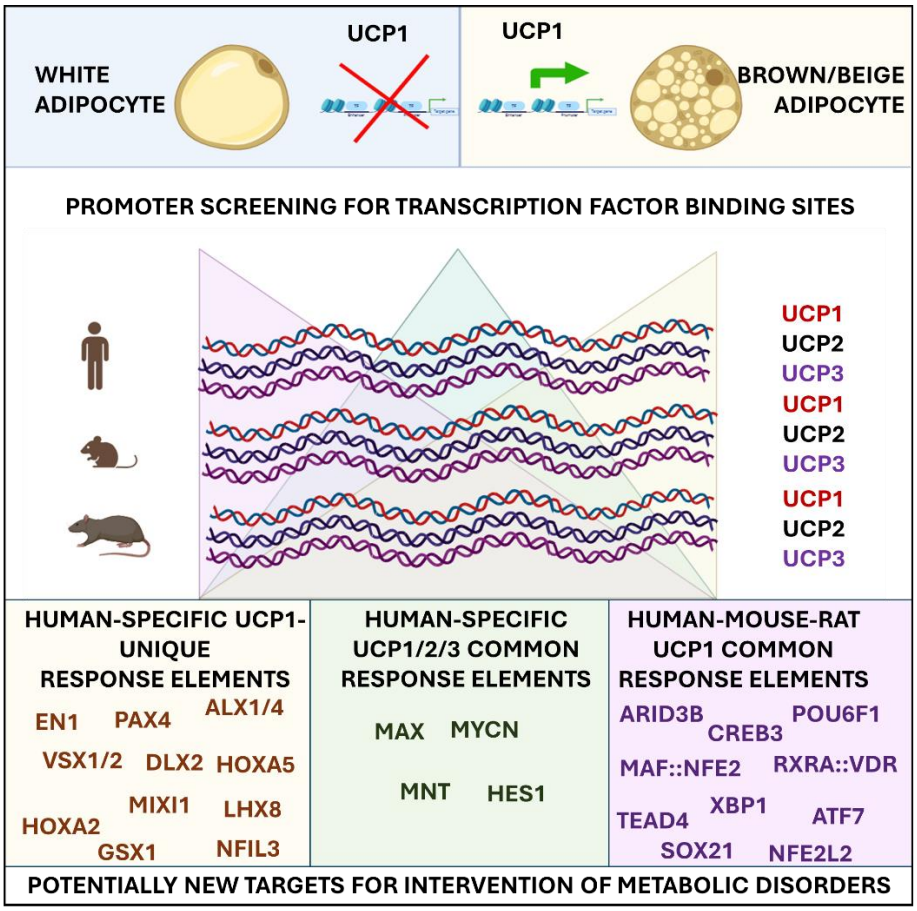
Abstract

Adipocyte thermogenesis is a promising target for treating metabolic disorders, but its regulatory mechanisms remain unclear. This study investigates transcription factors (TFs) and regulatory elements that may control the human *UCP1* gene, which is essential for thermogenesis and the formation of the adipocyte phenotype. Using the Eukaryota Promoter Database, we performed computational analyses of the *UCP1, UCP2* and *UCP3* promoter sequences in humans, mice and rats to identify conserved and species-specific elements. We also used transcriptome data from human neck-derived adipocytes and databases (Contra v3, ChiPBase, TFlink, AdipoNet, ISMARA, etc.) to narrow potential regulatory TFs. Our results show that mouse and rat *UCP1* enhancers lack large segments, primarily due to the insertion of repetitive elements that are already lost in some clades. We identified key TFs such as PPARA, PPARG, THR, RARE, RXR, JUN, TFAP2 and SREBF1 as general regulators of *UCPs.* Additionally, human-specific *UCP1* regulatory hotspots (e.g. 5'-TCTAATTAGA-3') recognised by homeodomain TFs (e.g. EN1, PAX4, HOXA5 and PRRX2) and NFIL3 were detected. Phylogenetically conserved regulatory elements suggest common TFs in human *UCP1*

paralogues (MAX, MYCN, MNT, HES1) and cross-species (POU6F1). These results improve our understanding of thermogenic adipocyte development and provide new therapeutic targets for metabolic diseases.

Keywords: Adipocyte browning, *UCP1* regulation, homeodomain Family Transcription Factors, Thermogenesis, Transposable Repeat Elements, Enhanceosome, Response element

Graphical Abstract



I. INTRODUCTION

Adipocytes are crucial for maintaining energy balance and a healthy metabolism in the human body. While white adipocytes primarily serve as energy reservoirs, brown and beige adipocytes are specialised in energy expenditure through thermogenesis. This thermogenic capacity is promising for influencing overall energy homeostasis, especially in the context of

42 metabolic disorders (Tayanloo-Beik et al., 2023; Seki et al., 2022; Wibmer et al., 2021; Becher
43 et al., 2021; Razzoli et al., 2016, Nedergaard et al., 2007). Non-shivering adipocyte
44 thermogenesis, a crucial component of the body's heat production mechanisms, can be
45 triggered by cold exposure and various environmental stimuli, however, the intricate
46 mechanisms are not yet fully understood (Hoang et al., 2013). Cold-induced adaptive
47 thermogenesis involves the upregulation of Uncoupling Protein 1 (UCP1) in the mitochondrial
48 membrane, which recycles protons to the matrix, thereby uncoupling oxidative
49 phosphorylation and ATP synthesis. This leads to controlled heat production and is an
50 attractive target for weight control strategies (Feldman et al., 2009; Ricquier 2011,
51 Nedergaard et al. 2001, Enerbäck 1997). In addition to UCP1-mediated thermogenesis, other
52 cyclic mechanisms may also contribute to adipocyte thermoregulation, such as endoplasmic
53 reticulum-associated Ca2+ (Ikeda et al. 2017), mitochondrial creatine (Kazak and Spiegelman
54 2020) and the triacylglycerol-fatty acid cycle (Cohen and Kajimura 2021). Despite extensive
55 research, a comprehensive understanding of the regulatory mechanisms controlling adipocyte
56 thermogenesis/browning remains elusive (Waldén et al. 2012; Nedergaard and Cannon 2014).
57 Several TFs have been proposed as direct regulators of *UCP1* gene expression, including
58 peroxisome proliferator-activated receptors (PPARs), retinoic acid receptors (RXRs, RARs),
59 thyroid hormone receptor (THR), cAMP response element-binding proteins (CREBPs), nuclear
60 factor erythroid response element 2 like 2 (NFE2L2), sterol regulatory element-binding protein
61 1 (SREBP1), early B-cell factor 1-2 (EBF1-2), the c-Jun::c-Fos complex and the CCAAT/enhancer-
62 binding proteins (CEBPA) as well as indirect regulators such as PPARG coactivator 1 alpha
63 (PGC-1α) and PR domain containing 16 (PRDM16) (Rosen et al. 2000; del Mar Gonzalez-
64 Barroso 2000; Rim and Kozak 2002; Seale et al. 2007; Timmons et al. 2007; Ricquier 2011;
65 Villaroya et al. 2017; Zhang et al. 2023). Recent reviews provide insights into the complex

epigenetic and transcriptional regulation underlying brown/beige adipocyte development (e.g. Li et al. 2019; Shapira and Seale, 2019; Ambele and Pepper, 2017; Villaroya et al., 2017; Inagaki et al., 2016; Harms et al., 2015; Bonet et al., 2013), but the intricate web of regulatory mechanisms is still to some extent unclear.

This study aims to use computational methods to identify transcription factors (TFs) that may regulate the expression of the human *UCP1* gene by identifying TFs that have putative functional DNA-binding motifs within the promoter regions of the human UCP1 gene in the range of -5Kb – +1Kb (6000 base pairs (BP); position: GRCh37/hg19; Chr. 4: 141,489,115-141,495,114). To simplify nomenclature, the term 'promoter' is used here for the entire sequence regions analysed, including the intergenic regions (1kb DNA segments in the positive direction from the transcription start site (TSS)), the proximal promoter (-1kb – TSS) and the distal enhancers (-1kb – - 5kb). By comparing *UCP1* orthologues and paralogues in humans, mice and rats, we aimed to uncover conserved and human-specific *UCP1* regulatory elements.

To determine whether a particular binding sequence is present within the human *UCP1* promoter, we used the eukaryotic promoter database (EPD) interface. This database uses a probability-based algorithm to identify TF response elements (REs) at four different p-value thresholds (p-value < 0.01, 0.001, 0.0001 and 0.00001). In this study, we focused on the two most stringent p-value thresholds ($p < 0.0001$ and $p < 0.00001$) because empirical evidence and computational results suggest that more permissive thresholds may lead to high false-positive rates (Wunderlich and Mirny, 2009; Benita et al., 2009). However, our exploratory studies have shown that using the most stringent threshold ($p < 1e-5$) can lead to high false-negative rates, as in the case of RE of HIF1A TF in gene promoters such as *VEGFA* and *GLUT1*, where the regulatory role of HIF1A is well known. By setting the significance level at $p < 1e-5$,

89 there would certainly be fewer false positives in the data, but the number of false negatives
90 would increase and be lost for further analysis, while false positives can be filtered out with
91 high efficiency by comparing paralogous and orthologous gene promoters. Therefore, to
92 achieve a balance between sensitivity and specificity, we primarily used the threshold $p < 1e-$
93 4, while occasionally referring to the results of $p < 1e-5$.

94 To exclude false results, we extended our analysis to *UCP2* and *UCP3* promoters from humans,
95 mice and rats. The UCP proteins belong to the solute carrier family (SLC), which consists of
96 different transporters such as ion channels, exchangers and passive transporters. *UCP2* and
97 *UCP3* are paralogues of *UCP1* but have different tissue distributions and do not share its
98 thermogenic function (Nedergaard and Canon 2003, Rousset et al. 2004, Nicholls 2021). Both
99 UCP2 and UCP3 contribute to the fine-tuning of mitochondrial function and energy
100 metabolism and are therefore important players in the maintenance of metabolic health and
101 protection against metabolic disorders, but their exact function is still unclear. UCP2 is widely
102 distributed in various tissues, including adipose tissue, brain, lung, pancreas and immune cells.
103 In contrast to UCP1, which is mainly found in brown adipose tissue and is primarily involved
104 in non-shivering thermogenesis by conducting protons back into the matrix, UCP2 has a more
105 general role. UCP2 is thought to regulate the production of reactive oxygen species (ROS) and
106 thus protect cells from oxidative stress (Hirschenson et al. 2022). UCP2 is also involved in the
107 regulation of insulin secretion by beta cells and in this way can also influence metabolic
108 efficiency and energy expenditure (Cadenas 2018). UCP3, on the other hand, is mainly
109 expressed in skeletal muscle and to a lesser extent in brown adipose tissue. One of the main
110 proposed functions of UCP3 is to promote the export of fatty acids from the mitochondria,
111 thereby preventing lipid-induced oxidative damage (Himms-Hagen and Harper 2001,

112 Schrauven et al. 2002). UCP3 may also be involved in maintaining energy balance and may
113 play a role in muscle thermogenesis, especially during periods of increased fatty acid intake or
114 muscle activity. By comparing these closely related genes, we aimed to identify TFs specifically
115 involved in the regulation of UCP1 and thereby adipocyte thermogenesis.

116 We also aimed to identify conserved and human-specific regulatory elements of the *UCP*
117 promoters by studying evolutionarily distant orthologues from mouse and rat. While mice and
118 rats are valuable as model organisms for thermogenic studies, the translation of results to
119 human physiology can be challenging due to inherent differences (Enerbäck et al. 1997). These
120 differences may affect metabolism and thermogenic processes, as variations in body size and
121 environmental temperature are among the factors that play an important role in shaping
122 evolutionary strategies to maintain energy balance (Tattersall and Milsom, 2009; Kottke et al.,
123 1948). Exploring potential regulatory differences in the expression of paralogous genes with
124 common origins but different tissue expression and protein functions, such as *UCP2* and *UCP3*,
125 in various species opens up the possibility of deciphering human *UCP1*-specific regulators and
126 providing insights into potentially more specific targets of intervention that may not be
127 apparent in traditional model organisms (Nedergaard and Canon 2003, Rousset et al. 2004,
128 Nicholls 2021, Hughes and Criscuolo 2008, Enerbäck et al. 1997, Warfel et al. 2023).

129 This study aims to provide a comprehensive understanding of the transcriptional regulation
130 of human *UCP1* by identifying key TFs with putative binding motifs. We take advantage of
131 functional differences between members of the UCP family and evolutionary conservation
132 between species. This could elucidate the specific regulatory processes that control the
133 phenotype of human adipocytes and improve our understanding of thermogenic fat cell
134 development, which would have significant therapeutic implications.

135 **RESULTS**

136 **1. Comparative analysis of the sequences of *UCP1*, *UCP2* and *UCP3* promoters, genes**
137 **and proteins in different species: structural and evolutionary aspects.**

138 To elucidate the structural features underlying the specific function of the three mitochondrial
139 solute carrier family 25 (SLC25) proteins, UCP1 (SLC25A7), UCP2 (SLC25A8) and UCP3
140 (SLC25A9), we performed a thorough comparative analysis encompassing the nucleotide
141 sequences of the genes and their promoters in the human, mouse and rat genomes. We also
142 aligned the amino acid sequences of the encoded proteins and analysed their complex three-
143 dimensional (3D) spatial structures.

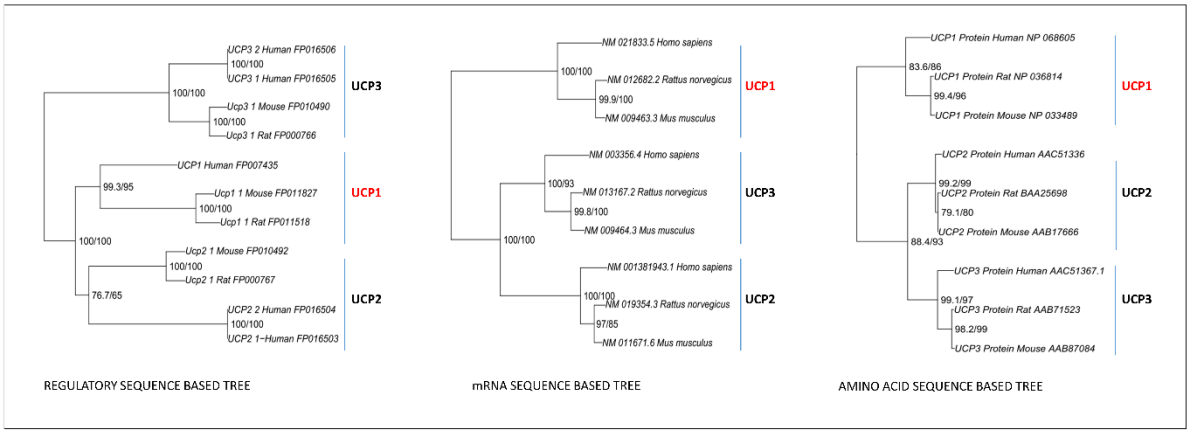
144 *1.1. Sequence similarity analysis shows distinct evolutionary trends between genes and*
145 *their regulatory regions.*

146 A phylogenetic tree reconstructed by merging the human, mouse and rat *UCP1, UCP2* and
147 *UCP3* data clearly shows that orthologous sequences were grouped, indicating a closer
148 relationship than to their paralogous counterparts when considering amino acid composition,
149 coding (mRNA) and regulatory DNA sequences, with high support values in all cases (Fig. 1A).
150 As expected, orthologous genes also show greater similarity between mouse and rat and are
151 more tightly clustered compared to human. In contrast, we obtain a different clustering of the
152 paralogous genes depending on the sequences analysed. While the reconstruction of the
153 mRNA and amino acid sequences identified UCP2 and UCP3 as a separate cluster from UCP1,
154 UCP1 and UCP2 were grouped based on the promoter sequence. These results are also
155 confirmed by pairwise distance analysis of the coding and promoter sequences of the *UCP*
156 genes (Suppl Fig. 1A).

157 **Figure 1.**

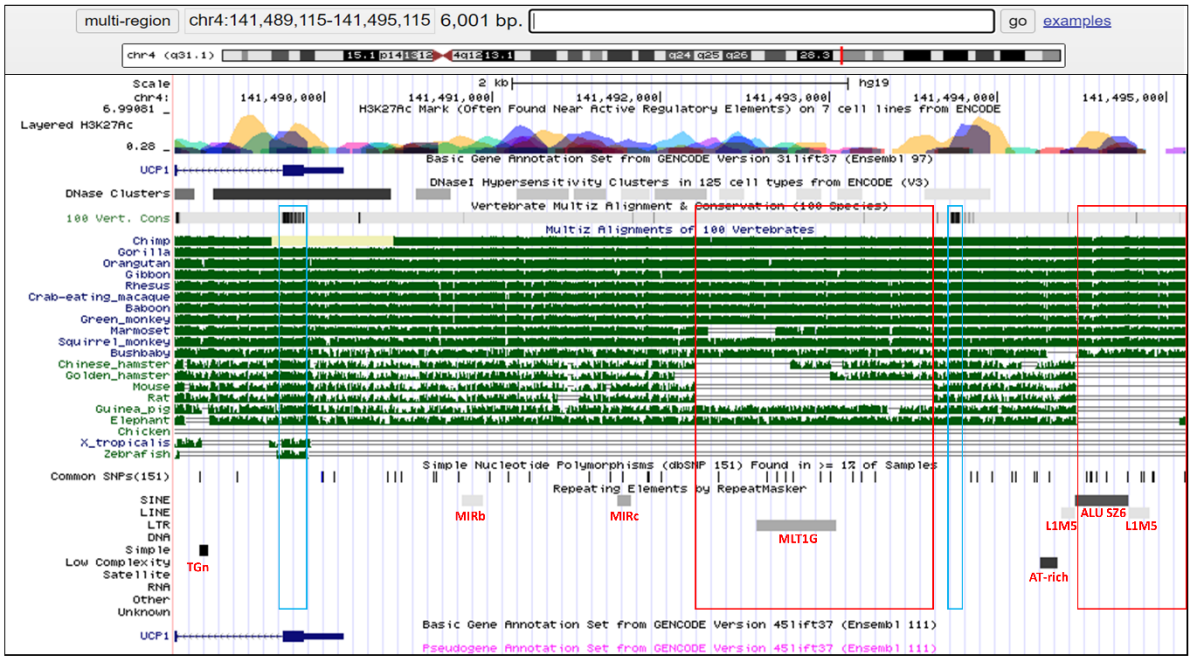
1A

Phylogenetic Tree Reconstruction



1B

Comparison of the nucleotide sequence of the human UCP1 promoter spanning from +1000bp to -5000bp with corresponding sequences from other species.



158

159

160 **Figure 1. 1A.** Phylogenetic tree reconstruction shown as a midpoint rooted phylogram based on the nucleotide
161 sequence of the *UCP1*, *UCP2* and *UCP3* upstream regulatory region, the mRNA and amino acid sequence of the
162 corresponding proteins in human, mouse and rat. Support values (aLRT/rapid bootstrap) are shown above the
163 branches (packages ape 5.7.1 in R environment). **1B** Comparison of the nucleotide sequence of the human UCP1
164 promoter between +1000bp – -5000bp with sequences from other species using the Multiz Alignment of 100
165 Vertebrates database in the UCSC genome browser. Red boxes highlight sequence segments absent from *UCP1*
166 promoter of mice and rats, blue boxes indicate segments displaying significant conservation across diverse
167 species. Red letters and grey lines indicate the abbreviation and position of the repetitive elements detected by
168 UCSC integrated Repeat Masker database.

169 *1.2 The presence of transposable and repetitive elements leads to considerable sequence*
170 *divergence within the distal enhancer regions of humans and rodents.*

171 The thermogenic *UCP1* gene evolved around 120-150 million years ago after the split of the
172 platypuses (monotremes) and marsupials (metatherians) from the lineage that gave rise to
173 the placental mammals (eutherians) (Gaudry et al. 2017). However, Jastroch et al. (2005)
174 demonstrated that all three members of the core *UCP* family were present before the
175 divergence of the ray-finned and lobe-finned vertebrate lineages around 420 million years
176 ago, however, were later lost or inactivated in several clades, such as birds. A cross-species
177 alignment of the analysed 6000 bp long nucleotide sequences of the *UCP1* promoter revealed
178 that a 1419 bp long sequence interval extending from -2083 bp to -3502 bp (GRCh37/hg19:
179 Chr4 141.492.198-141.493.617) in the distal enhancer region of the human *UCP1* promoter
180 and the last 658 bp long segment of the 6000 bp long sequence analysed (Chr4: 141.494.458-
181 141.495.115; at -4341 bp to -5000 bp) was missing in the UCP1 promoters of mice and rats
182 (Fig. 1B). This absence was confirmed by a BLAST search in the NCBI database
183 (https://blast.ncbi.nlm.nih.gov/). Examination of the presence of the 1419 bp-long sequence
184 in other species using the Multiz Alignment of 100 Vertebrate Tool in the UCSC Genome
185 Browser (https://genome.ucsc.edu/) suggests that it is an evolutionarily novel sequence that
186 arose around the divergence between placentals and marsupials with the appearance of the
187 thermogenic *UCP1* gene (Jastroch et. al. 2024), however, it seems to be secondarily lost in
188 certain clades, including mice and rats (Suppl. Fig. 1B). It is noteworthy that short palindromic
189 sequences or their residues can be detected at the beginning and end of these evolutionarily
190 novel DNA fragments in the human *UCP1* promoter (Suppl. Fig. 1C, 1D,1E, 1F, 1G; red boxes).
191 Eukaryotic Promoter Database analysis ($p < 1e-5$) shows that the RE of CREB3 at -3498 bp
192 corresponds to the palindromic "TGACGTCA" sequence at the end of the 1419 bp long

193 segment (Suppl. Fig. 1D). Notably, the 468 bp central region of this new sequence (Chr4:
194 141.492.573-141.493.040; at -2458 bp to -2925 bp), identified as a long terminal repeat (LTR)
195 retrotransposon repetitive element by RepeatMasker (http://www.repeatmasker.org) (Jurka
196 2000) named MLT1G from the endogenous retroviruses ERVL-MaLR family (Fig. 1B). This 468
197 bp sequence appears to be integrated into the human genome several times, as the BLAST
198 search shows (Suppl Fig. 1H). It is also interesting that at the beginning and the end of this
199 central segment, TF REs of the homeodomain family are identified, DLX2 RE at -2478 bp, while
200 at the end for HOXD8, HOXD9, HOXB5, POU6F1 at -2913 bp. The Eukaryotic Promoter
201 Database does not identify a TF RE in the middle part of this central segment with a probability
202 of $p < 1e-5$. In the 658 bp long sequence at the end of the UCP1 promoter segment which was
203 also missing in the rodent genome, a 317 bp (Chr4: 141.494.458-141.495.774; at -4343 bp to
204 -4659 bp) and a 124 bp long segment (Chr4: 141.494.775-141.494.898; at -4660 bp to -4783
205 bp) were also recognised by RepeatMasker as a short interspersed nuclear element (SINE) and
206 a long interspersed nuclear element (LINE), respectively (Fig. 1B). The SINE was classified as
207 an Alu sz6 repeat from the ALU family, and the following LINE was identified as an L1M5 repeat
208 which belongs to the L1 family. The BLAST searches have found similar repeat sequences at
209 over 1000 sites in the human genome. Using RepeatMasker, we saw several additional
210 repetitive elements in the human *UCP1* promoter, whose possible role requires further
211 investigation (Fig. 1B, grey lines). This observation may emphasise the dynamic nature of the
212 genetic elements and their potential role in regulating gene expression.

213 Nevertheless, the intronic regions of *UCP1* at around +250 bp and the distal enhancer region
214 at around -3600 bp show above-average sequence similarity across species and may indicate
215 conserved DNA regions with REs that are important for transcriptional control, such as the
216 myeloid zinc finger 1 (MZF1) at -3667 bp, and the glucocorticoid modulatory element binding

217 protein 1 (GMEB1) and GLIS family zinc finger 2 (GLIS2) at around +250 bp (Fig. 1B, blue boxes;
218 Suppl Table 2).

219 *1.3 Subtle structural differences between orthologous and paralogous UCP1 proteins may*
220 *have functional implications*

221 The primary 3D structures of human *UCP1, UCP2* and *UCP3* proteins show a high
222 degree of homology with their mouse and rat counterparts. This was established by predictive
223 modelling with AlphaFold 2 and structural alignment with the RCSB PDB interface (Fig. 2A, see
224 Methods). Supplementary Figure 2A shows these spatial protein structures in a colour-coded
225 format and provides insight into their conformational similarities. Remarkably, the model
226 reliably predicts the spatial structure of *UCP1* in all three species, whereas model confidence
227 tends to be lower for *UCP2* and *UCP3*. Interestingly, a short interdomain sequence shows a
228 particularly low model confidence within the UCP proteins of all three species (Suppl Fig. 2A,
229 red colour). The context-dependent variability in spatial structure may provide the
230 opportunity to influence functionality under certain conditions, such as modulating proton
231 conductance or influencing the binding of fatty acids and purine nucleotides. Detailed pairwise
232 structural alignment measurements of the analysed proteins, showing the extent of structural
233 similarity to human *UCP1*, are presented in Table 1. These analyses show a consistent
234 template modelling score (TM score) of 0.75 or higher compared to human UCP1 and
235 matching equivalent residue numbers in all cases (Table 1). In particular, the results show the
236 remarkable similarity between mouse and rat *UCP1* and human *UCP1* with an impressive TM
237 score of 0.97. The homology is slightly lower for the paralogous genes *UCP2* (0.84 TM score)
238 and *UCP3* (0.77 TM score). It is worth noting that the amino acid sequence similarities closely
239 match the result of the DNA sequence comparisons (Table 1.), both underscoring the greater

240 similarity in the orthologous protein structures. However, the relevance of amino acid
241 sequence homology to protein structure similarity is challenged by the fact that the presence
242 of the conserved UCP protein structure is a recurring feature within the SLC25 protein family
243 (Hoang et al., 2013; Crichton et al., 2017), regardless of the degree of homology between the
244 constituent amino acids. For example, the ADP/ATP translocase (SLC25A4 or ANT1) has a
245 similar predicted 3D folded structure (Fig. 2B) and a substantial TM score of 0.8, despite a
246 modest amino acid homology of 21% (Table 1). In summary, the phylogenetic analyses of
247 *UCP1, UCP2* and *UCP3* in the human, mouse and rat genomes, and their protein 3D structural
248 comparison show the closest relationships between the orthologues. The paralogues show
249 different clustering patterns based on gene and promoter sequences. The observed structural
250 similarities, independent of amino acid sequence homology, emphasise the conserved nature
251 of UCP protein structure within the SLC25 family and raise questions about the functional
252 impact of sequence variations in specific interdomain regions.

253 Table 1.

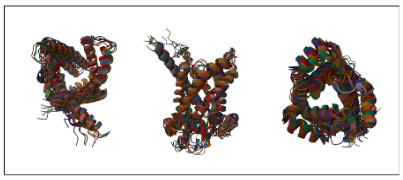
	Entry	Chain	RMSD	TM-score	Identity	Equivalent Residues	Sequence Length	Modelled Residues
UCP1-Human	AF-P25874-F1-model_v4.pdb	A	-	-	-	-	307	307
UCP1-Mouse	AF-P12242-F1-model_v4.pdb	A	1.14	**0.97**	79%	307	307	307
UCP1-Rat	AF-P04633-F1-model_v4.pdb	A	1.06	**0.97**	79%	305	307	307
UCP2-Human	AF-P55851-F1-model_v4.pdb	A	3.08	**0.84**	57%	304	309	309
UCP2-Mouse	AF-Q9R246-F1-model_v4.pdb	A	2.86	**0.84**	57%	301	309	309
UCP2-Rat	AF-P56500-F1-model_v4.pdb	A	2.2	**0.89**	57%	302	309	309
UCP3-Human	AF-P55916-F1-model_v4.pdb	A	3.51	**0.77**	48%	298	312	312
UCP3-Mouse	AF-P56501-F1-model_v4.pdb	A	4.01	**0.75**	54%	305	308	308
UCP3-Rat	AF-P56499-F1-model_v4.pdb	A	3.55	**0.78**	46%	299	308	308
ANT1-Human	AF-P12235-F1-model_v4.pdb	A	3.07	**0.8**	**21%**	283	298	298

Table 1. Results of pairwise structure alignments, measures describing the extent of structural similarity of the human, mouse and rat UCP1, UCP2 and UCP3 proteins to human UCP1. The structure alignments are based on protein structures predicted by AlphaFold 2 and aligned by RCSP-PBD. For interpretation of the measures, see Methodology.

Figure 2.

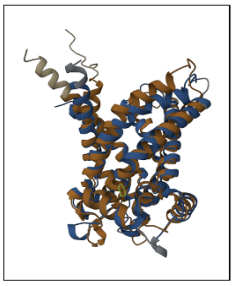
2A

Aligned predicted protein structure of UCP1, UCP2 and UCP3 in human, mouse and rat.



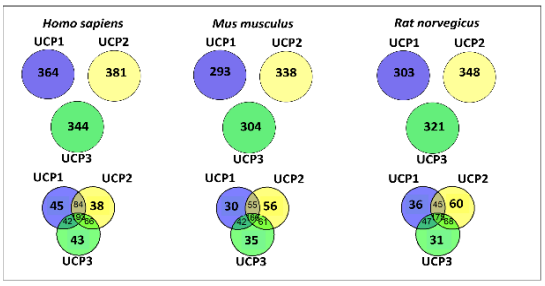
2B

H-UCP1 aligned with H-ANT1



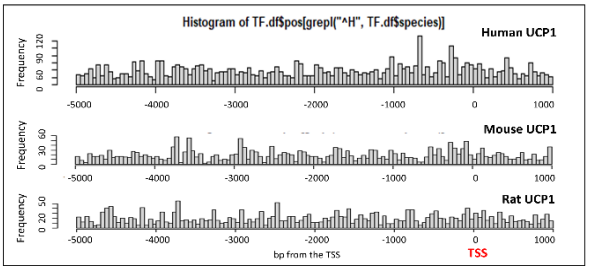
2C

Number of TFs with RE in human, mouse and rat UCP1, UCP2 and UCP3 Promoter



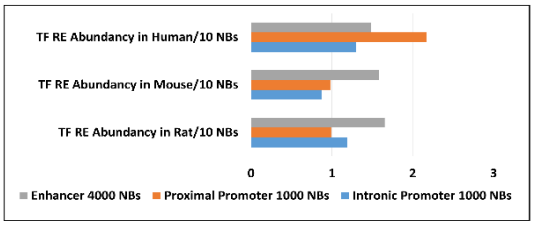
2D

TF RE frequency in human, mouse and rat UCP1 gene promoter



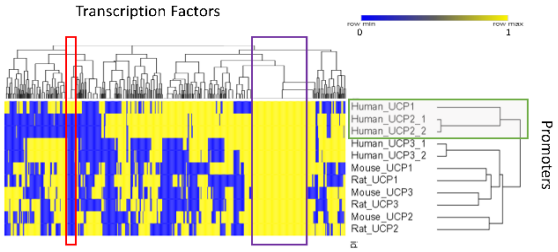
2E

TF RE abundancy in UCP1 gene promoter regions of human, mouse and rat



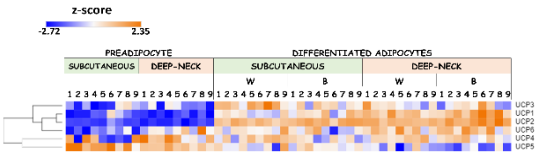
2F

Hierarchical cluster analysis of the human, mouse and rat UCP1, UCP2 and UCP3 genes based on the TF REs detected in their promoters.



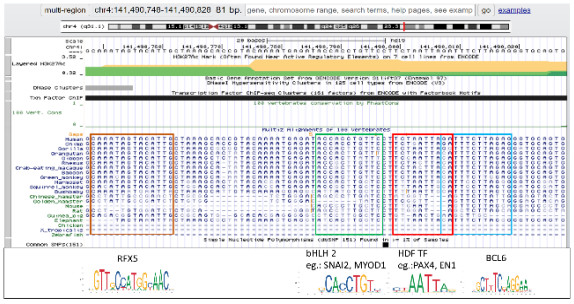
2G

Heatmap shows the gene expression profile of the six UCP genes in adipocyte samples of cervical origin.



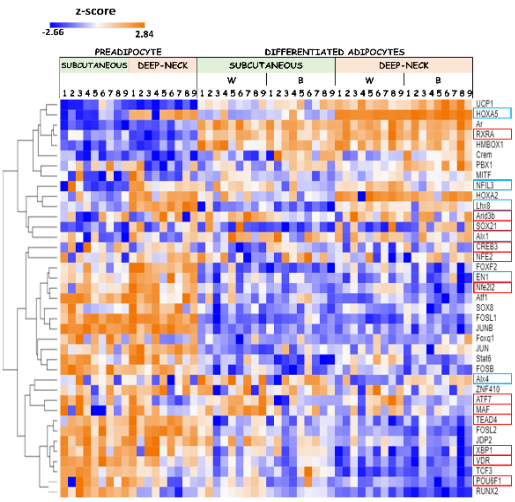
2I

Comparison of the sequence of the HDF1-class TF RE at the human UCP1 promoter -690 BP with sequences from other species.



2H

Heatmap shows the gene expression profile of 45 TFs with RE in the human UCP1 promoter in adipocyte samples of cervical origin.



2J

Heatmap shows the gene expression profile of 12 TFs with RE in the human, mouse and rat UCP1 promoter in adipocyte samples of cervical origin.

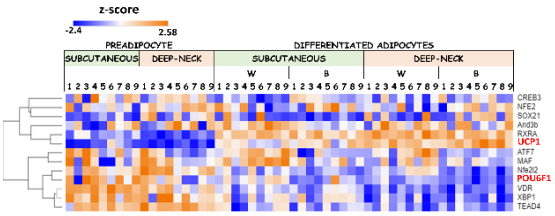


Figure 2. 1A. RCSB PBD aligned, AlphaFold 2 predicted protein structures of human, mouse and rat UCP1, UCP2 and UCP3 from three different perspectives. **1B.** AlphaFold 2 predicted H-UCP1 aligned with H-ANT1. **1C.** Number of TFs with RE in human, mouse and rat *UCP1*, *UCP2* and *UCP3* Promoter: -5000 bp – +1000 bp (p<1e-4) (Venn program in R environment). **1D.** TF RE frequency in human, mouse and rat *UCP1* gene promoters -5000 bp – +1000 bp (p<1e-4) (Excel). **1F.** Bars show the frequency of TF RE per 10 bp in different parts of the UCP1 promoter in humans, mouse and rat -5000 bp – +1000 bp (p<1e-4). Distal enhancer -5000 bp – -1001 bp; Proximal and Core promoter-1000 bp – TSS; Intronic promoter TSS – +1000 bp (Excel). **2F**. Hierarchical cluster analysis of the human, mouse and rat *UCP1*, *UCP2* and *UCP3* genes based on the TFs detected in their promoters by RE. The red box highlights the 13 TFs that are unique to the human *UCP1* promoter, the purple box shows the 93 TFs that are common to the examined UCPs, while the green box shows the clusters with the most similar TF RE profile to *UCP1* (Morpheus Interface). **2G** Heatmap shows the gene expression profile of the six *UCP* genes in adipocyte samples of cervical origin. Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies *UCP*s with similar gene expression profiles to *UCP1* (Morpheus Interface). **2H** Heatmap shows the gene expression profile of 45 TFs with RE in the human *UCP1* promoter in adipocyte samples of cervical origin. For comparison, *UCP1* is also shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies TFs with similar gene expression profiles to *UCP1* (Morpheus Interface). **2I** Comparison of the nucleotide sequence of the HDF1-class TF RE at the human *UCP1* promoter -690 bp with sequences from other species using the Multiz Alignment of 100 Vertebrates database in the UCSC genome browser (red box). In the vicinity of HDF1 RE the TF RE sequences of RFX5, MYOD1 and BCL6 identified by EPD are highlighted with brown, green and blue boxes, respectively. **2J** Heatmap shows the gene expression profile of 12 TFs with RE common in the human, mouse and rat *UCP1* promoter in adipocyte samples of cervical origin. For comparison, *UCP1* is also shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies TFs with similar gene expression profiles to *UCP1* (Morpheus Interface). Adipocytes were differentiated using a white (W) or brown (B) differentiation protocol (Tóth et al. 2020).

2. Comparative analysis of transcription factor binding sites of *UCP* promoters in humans, mice and rats

2.1 Cross-species comparative analysis of putative TF REs in the UCP1/2/3 promoters reveals greater similarity of human UCP1 to human UCP2 than to rodent UCP1.

After a thorough investigation of the promoter sequences of the *UCP1*, *UCP2* and *UCP3* genes in humans, mice and rats, we focussed on determining the specific transcriptional regulation of the human *UCP1* gene. Using the Eukaryotic Promoter Database (EPD) and a large number of TF motifs from the JASPAR databases, we performed an in-depth analysis with 579 TF motifs

291 (Dreos et al., 2017). Using stringent significance thresholds ($p < 1e-4$), we identified a variable
292 number of TFs with putative REs in the nine genes analysed, ranging from 293 to 381 (Fig. 2C,
293 Suppl Fig. 2B, Suppl Table 1). Applying a more stringent threshold of $p < 1e-5$ resulted in the
294 detection of 63 to 123 TFs with REs in the promoter of the *UCP* genes in the three species
295 (Suppl Fig. 2C). The distribution of TF REs within the *UCP1* promoters showed considerable
296 differences between regions and species. While in humans the proximal promoter has the
297 highest TF binding sites, in mice and rats it is the enhancer region (Fig. 2D, 2E; Suppl Table 2).
298 Visualisation on a heatmap and hierarchical clustering showed that the regulatory TF-RE
299 profile of the human *UCP1* gene is very similar to that of its paralogue, the human *UCP2* gene
300 (Fig. 2F). This is in contrast to the observations based on protein structure, amino acid
301 composition, and gene and promoter sequences (Fig. 1A, 2A), where the orthologues are more
302 similar. Comparing the relative gene expression profiles of the human *UCP* genes in the
303 cervical-derived adipocyte samples, the expression of *UCP1* was most closely related to that
304 of *UCP2*, as shown by the correlation using Spearman's Rank Similarity Matrix (Fig. 2E).

305 *2.2 Comparative analysis of transcription factor response elements of 9 UCPs reveals*
306 *unique regulatory sequences for the human UCP1 gene associated with TFs of the*
307 *homeodomain family*

308 Our comprehensive study identified a total of 364 TFs (including dimers, trimers and variants)
309 with putative REs in the analysed promoter region of the human *UCP1* gene (Figure 2C; Suppl
310 Table 1). A comparison with the results of the analyses of the regulatory regions of the human
311 *UCP2* and *UCP3* genes shows that 319 of these TF REs were also present in one of the analysed
312 paralogous human *UCP1* promoters. However, 45 TFs (including dimers and variants) were
313 present exclusively on the *UCP1* promoter. We examined these 45 *UCP1*-specific TFs in more

314 detail by first comparing their expression patterns with *UCP1* in cervical-derived subcutaneous
315 and deep neck preadipocytes and mature adipocytes, differentiated ex vivo (Figure 2H). Gene
316 expression of 5 TFs, including *AR, RXRA, HMBOX1, CREM* and *HOXA5* showed a significant
317 positive correlation with *UCP1* (Suppl Table 3). Conversely, the gene expression of 17 TFs,
318 including *JUNB, FOSL1* and *2, ATF1, VDR* and *NFE2L2* as the top six, showed a negative
319 correlation. However, the RE of RXRA shows only a specific occurrence in the human *UCP1*
320 promoter when dimerised with VDR. Similarly, only a single variant of the JUNB TF RE,
321 JUNB_var2 RE, shows a specific occurrence in the human *UCP1* promoter and only when
322 dimerised with FOSL1 or FOSL2. When considered alone or dimerised with other TFs, their REs
323 are identified in all *UCP*s tested, so their *UCP1* specificity in humans is context-dependent.

324 Next, in comparison to the TFs with REs detected in the regulatory regions of mouse and rat
325 *UCP1, UCP2*, and *UCP3* genes, we identified 13 of the 45 human *UCP*1-specific TFs that do not
326 have REs in the promoters of these rodent *UCP* genes, namely: ALX1, ALX4, DLX2, EN1, GSX1,
327 HOXA2, LHX8, MIXL1, NFIL3, PAX4, VSX1, VSX2 and HOXA5 (Figure 2F, red box). Except for
328 NFIL3, these TFs belong to the class of homeodomain family TFs (HDF; JASPAR database). The
329 Eukaryotic Promoter Database can also detect most of these REs with the highest probability
330 ($p < 1e-5$) (Table 2A, Suppl Table 2), supporting that these TFs may regulate thermogenic
331 processes in a species-specific manner. In addition, we examined the RE nucleotide sequences
332 identified by the Eukaryotic Promoter Database for these 13 TFs and detected a palindromic
333 DNA nucleotide sequence of 10 bp, 5'-TCTAATTAGA-3', positioned at -690 bp in the human
334 *UCP1* promoter (EPD database; UCSC Genome Browser GRCh 37/hg19: 141.490.801-
335 141,490,810), and serves as RE for 9 of the 13 human *UCP1*-specific TFs (Table 2A). While this
336 short RE sequence was located outside the 1419 bp and 658 bp long sequence region found
337 only in humans, it was indeed absent from the mouse and rat *UCP1, UCP2*, and *UCP3*

338 promoters, as well as the human *UCP2* and *UCP3* promoters, according to Eucaryotic Promoter
339 Database and NCBI BLAST searches, suggesting that it potentially play a unique regulatory role
340 in human adipocyte thermogenesis. Further data indicate that this sequence is conserved to
341 some degree in primates (one base difference), such as chimpanzees, gorillas, orangutans,
342 baboons and green monkeys (Figure 1B; Multiz Vertebrate Sequence Alignment in UCSC),
343 expanding the range of possible species in which this regulatory process may be implemented.
344 In addition to the homeodomain family TFs, the nuclear factor interleukin 3 (NFIL3), which
345 belongs to the basic leucine zipper TF family and binds to the activating transcription factor
346 (ATF) site, was found to be another potential human-specific *UCP1* regulatory factor with a
347 binding site at -881 bp in the human *UCP1* promoter, although not detected at a stringent
348 probability level (p<1e-5).

349 On the other hand, 93 TF REs were consistently present in all *UCP* promoters of the three
350 species analysed, indicating some degree of shared regulation (Figure 2F Purple box, Suppl
351 Table 1). These include TFs for which phylogenetically well-conserved binding sites were found
352 in the *UCP1* promoters in previous studies, such as PPARA, PPARG, RARE, TFAP2, SP, THR and
353 SREBF1 (Gaudry and Campbell 2017).

354 **Table 2.**

355 2A Human-specific *UCP1* putative TFs

TF	RE position in human UCP1	Nucleotide sequence in human UCP1	GE. Corr. TF vs.UCP1	RNA Expression in Adipocytes	ChipSeq data	PubMED abstract with UCP1	PWM in JASPAR	TF Class/Family	AdipoNET	EPD p<1e-5
Alx1	-690	TCTAATTAGA	NO	VERY LOW	NO	NO		HDF1/Paired-related	NO	NO
Alx4	-690	TCTAATTAGA	NO	VERY LOW	NO	NO		HDF1/Paired-related	NO	NO
Dlx2	-2478	CATAATTGCA	NO	VERY LOW	NO	NO		HDF2/NK family	NO	YES
EN1	-690 -689	TCTAATTAGA	NEG. ***	HIGH	NO	YES		HDF1/NK family	NO	YES
GSX1	-690 -689	TCTAATTAGA	NO	VERY LOW	NO	NO		HDF1/HOX family	NO	NO
HOXA2	-3959	TCTCATTAGC	NO	LOW	NO	NO		HDF1/HOX family	NO	NO
HOXA5	-4785	GCAAATTAGT	POS.*	LOW	NO	YES		HDF1/HOX family	NO	YES
Lhx8	-690 -689	TCTAATTAGA	NO	LOW	NO	YES		HDF1/LIM family	NO	YES
MIXL1	-690 -689	TCTAATTAGA	NO	VERY LOW	NO	NO		HDF1/Paired-related	NO	NO
NFIL3	-881	ATATTACATCATTATA	NO	HIGH	NO	NO		bZIP/CEBP-related	NO	NO
PAX4	-690 -689	TCTAATTAGA	NA	NA	NO	NO		HDF1/ Paired Plus Homeo Domain	NO	YES
VSX1	-690 -689	TCTAATTAGA	NO	LOW	NO	NO		HDF1/Paired-related	NO	NO
VSX2	-690 -689	TCTAATTAGA	NA	NA	NO	NO		HDF1/Paired-related	NO	YES

2B Homeodomain family TFs with putative RE in human *UCP1*

TF	RE position in human UCP1 by EPD	ConTra	GE Corrl. TF vs UCP1	RNA Expression in Adipocytes	ChipSeq data	PubMED abstract with UCP1	PWM in JASPAR	TF Class/Family	AdipoNET	EPD p<1e-5
HOXA5	-4785	YES	POS.*	MEDIUM	NO	YES		HDF1/HOX family	NO	YES
HOXB6	NA	NO	POS*	MEDIUM	NO	NO		HDF1/HOX family	NO	NO
VAX2	NO	YES	POS.*	LOW	NO	NO		HDF1/NK family	NO	NO
Dlx4	NO	YES	POS.****	LOW	NO	NO		HDF1/NK family	NO	NO
NOTO	NO	YES	POS.*	VERY LOW	NO	NO		HDF1/NK family	NO	NO
EN1	-690 -689	YES	NEG.***	HIGH	NO	YES		HDF1/NK family	NO	YES
PRRX1	NO	YES	NEG.****	HIGH	NO	NO		HDF1/Paired-related	NO	NO
PRRX2	-690 -689	YES	NEG.****	HIGH	NO	NO		HDF1/Paired-related	NO	NO
PROP1	-692	YES	NA	NA	NO	NO		HDF2/Paired-related	NO	NO
HOXD8	-4195 -4116 -2914	YES	NEG.*	MEDIUM	NO	NO		HDF1/HOX family	NO	NO
HOXC4	NA	YES	NEG.***	MEDIUM	NO	NO		HDF1/HOX family	NO	NO
HOXC8	NA	YES	NEG.***	MEDIUM	NO	NO		HDF1/HOX family	NO	NO
POU6F1	-4716 -4715 -2913 -2912	YES	NEG.**	MEDIUM	NO	NO		HDF1/POU domain	NO	NO
PHOX2A	-692	YES	NA	NO	NO	NO		HDF2/HOX family	NO	NO
LBX2	-690 -689	YES	NEG.*	LOW	NO	NO		HDF1/NK family	NO	NO
PAX4	-690 -689	YES	NA	NA	NO	NO		HDF1/ Paired Plus Homeo Domain	NO	YES
VSX2	-690 -689	YES	NA	NA	NO	NO		HDF1/Paired-related	NO	YES

Table 2. Tables 2A and 2B summarize our data on TFs with an RE in the human *UCP1* promoter identified by EPD (p<1e-4) and filtered according to specific criteria. These include the position of the RE relative to the TSS, their gene expression pattern relative to *UCP1*, their gene expression levels measured in neck-derived preadipocytes and *in vitro* differentiated adipocytes, their ChipSeq data from the ChiPBase and TFlink databases, their classification into TF class and family based on JASPAR databases, their PWM, their involvement in the adipocyte phenotype formation pathway according to the AdipoNET interface, EPD TF RE search results at p<1e-5 significance level in the human *UCP1* promoter. Bold highlights TF or value outstanding in the related category. **Table 2A** highlights the 13 putative regulatory TFs that possess REs only in the human *UCP1* promoter and are

366 absent from the other *UCP* genes examined. **Table 2B** summarizes our data for 17 TFs belonging to the
367 homeodomain Family 1 and 2 TF classes whose RE matches the human-specific binding sequence identified in
368 the *UCP1* promoter ("TCTAATTAGA") the expression of which is correlated with *UCP1* or for which we do not
369 have gene expression data and therefore cannot exclude them. GE: Gene Expression, Corr.: Correlation.

370 *2.3 Narrowing transcription factors that bind to the unique palindromic nucleotide*
371 *sequence of the human UCP1 regulatory region*

372 Next, we wanted to predict the TFs that could bind to the unique nucleotide sequence "5'-
373 TCTAATTAGA-3'" located at -690 bp in the human *UCP1* regulatory region. In addition to the
374 9 human-specific TFs from the homeodomain family, the Eukaryotic Promoter Database also
375 revealed several TFs from the homeodomain family ($p < 1e-4$) that potentially recognise this
376 RE but lack specificity for human *UCP1* (Suppl Table 2). Therefore we performed a
377 comprehensive analysis to determine which of the proposed TFs are likely to play a functional
378 role and to assess their importance in the adipocyte browning process (Table 2B). First, we
379 verified the presence of the predicted TF REs of homeodomain family 1 and 2 (HDF1-2) using
380 the ConTra v3 platform (stringency: core=0.95; similarity matrix=0.85;
381 http://bioit2.irc.ugent.be/contra/v3/#/step/1) (Suppl Fig. 3).

382 In addition to confirming the results of the Eukaryotic Promoter Database, the ConTra v3
383 interface can be used to gain insight into the evolutionary conservation of REs based on
384 aligned sequences from many species. Contra v3 analysis verified the set of TFs potentially
385 associated with the *UCP1* promoter at -690 bp predicted by the Eucaryota Promoter Database,
386 as the HDF TFs were also identified by ConTra v3 (Suppl. Fig 3.). Therefore, we extended our
387 investigation to all TFs of the HDF1 class based on JASPAR, as they could potentially bind to
388 the *UCP1* promoter at this position. To narrow down the group of potential homeodomain
389 family TFs, we examined their relative gene expression levels in our neck-derived adipocyte
390 samples and focused on those that showed a significant positive or negative correlation with

391 *UCP1* gene expression (Table 2B; Suppl Fig. 4, Suppl Table 3). For some TFs, we did not have
392 gene expression data, so we could not exclude them from the group of TFs that might bind to
393 this specific RE. We collected additional data on this narrowed group of TFs, including ChipSeq
394 data based on ChiPBase and TFlink databases, search results from the Eukaryotic Promoter
395 Database using more stringent $p < 1e-5$ criteria, and position weighting matrices (PWMs) from
396 the JASPAR database to see the sequences of the REs (Table 2A, 2B). Moreover, we used the
397 AdipoNET platform (Tóth et al. 2021) to categorise these TFs based on their potential
398 involvement in the formation of brown, white or both adipocyte phenotypes as "linker"
399 pathway genes. This categorization provided insights into the possible roles of these proteins
400 in specific adipocyte phenotypes. In addition, we performed an automated search of the
401 PubMed database to identify the co-occurrence of these TFs and *UCP1* in article abstracts to
402 determine what knowledge we have about these TFs concerning *UCP1* (Suppl Table 4, Suppl
403 Fig. 5.). The data collected indicate that Engrailed Homeobox 1 (EN1), Paired Related
404 Homeobox 2 (PRRX2), Homeobox A5 (HOXA5), Paired Box 4 (PAX4) and Visual System
405 Homeobox 2 (VSX2) are the TFs most likely to bind to the given RE sequence. While EN1 and
406 PRRX2 showed high expression in adipocytes and a strong negative correlation with *UCP1*,
407 HOXA5 showed a positive correlation, but a low expression. Although the expression profiles
408 of PAX4 and VSX2 were not investigated in our previous transcriptomic analyses (Tóth et al.
409 (2020), a more stringent search in the Eukaryotic Promoter Database ($p > 1e - 5$) also identified
410 REs for these TFs at -690 bp, supporting their potential for functional binding at this position.
411 In addition, the ISMARA platform (Balwierz et al. 2014) shows high activity of the PAX4 motif
412 in adipose tissue (Suppl Fig. 6). Of note, the AdipoNET database and publication data mining
413 so far have not previously linked these homeodomain family TFs with pathways belongs to

414 adipocyte phenotype formation or *UCP1* expression, so we present here a new area for further
415 investigation of the role of these TFs in browning processes.

416 *2.4 Investigation of the vicinity of the human UCP1-specific homeodomain family REs*
417 *identifies potential regulatory hubs*

418 TFs of the homeodomain family are known to be involved in embryonic developmental
419 processes, suggesting a plausible role in the regulation of tissue-specific *UCP1* expression
420 during adipocyte maturation (Gehring 1994; Banerjee-Basu and Baxevanis 2001). Their
421 recognised function involves the modulation of gene expression through the influence of
422 epigenetic markers that affect the accessibility of chromatin and thereby control the binding
423 of TFs to nearby REs. In addition, homo- or heterodimers are often formed as cooperating
424 complexes to improve their DNA binding specificity (Slattery et al. 2011; Cain et al. 2023).
425 Therefore, our further investigations focused on the regulatory elements of the TFs near the
426 identified human-specific genomic positions (in a range of +40 bp and -40 bp) (Table 2A) to
427 determine which TFs can be allowed to operate by these homeodomain family TFs.

428 First, we investigated TF REs near position -690 of the human *UCP1* promoter. Using a
429 stringent search ($p < 1e-4$), the Eukaryotic Promoter Database identified the presence of REs
430 for B-cell lymphoma 6 (BCL6) at -702, regulatory factor X5 (RFX5) at -651/650, and REs for
431 basic helix-loop-helix (bHLH2) class TFs around -679 bp (Fig. 2I). Remarkably, BCL6 and RFX5
432 showed high expression and a strong positive correlation with *UCP1* in our cervical adipocyte
433 samples (Suppl Fig. 7; Suppl Table 3). Of the bHLH2 TFs with putative RE at -679 bp, ID4, SNAI2
434 and TCF3/4/12 also showed high gene expression levels but a negative correlation with *UCP1*.
435 Furthermore, the primary interaction network of PAX4 described using the ISMARA platform
436 (Balwierz et al. 2014) also includes TFs identified near its RE at -690 bp, such as BCL6, TCF4,

437 ID4, SNAI2 and RFX5, indicating the potential formation of a regulatory complex
438 (enhanceosome) at this position of the human *UCP1* promoter (Suppl Fig. 8).

439 Next, we investigated the proximity of other identified human *UCP1*-specific putative REs of
440 the homeodomain family, the REs of HOXA5, HOXA2 and DLX2, located at -4785, -3959 and -
441 2478 bp, respectively (Table 2A). These REs were present in the primate genomes but absent
442 in the mouse and rat *UCP1* promoters (Table 2A, Suppl. Fig. 3 Contra v3). The HOXA5 and DLX2
443 REs were also identified by the Eukaryotic Promoter Database with a more conservative
444 probability of p<1e-5. All three TFs are expressed at low levels in adipocytes of cervical origin
445 and only *HOXA5* gene expression showed a positive correlation with *UCP1*. Remarkably, we
446 identified the presence of the HIC1 TF RE near the HOXA5 RE (Suppl Fig 9A), which showed
447 high expression levels in pre-adipocyte samples but a significant reduction in mature
448 adipocytes and a negative correlation with *UCP1* expression (Suppl Table 3). A context-
449 dependent regulatory aspect is suggested by the gene expression profile of TEAD RE, which
450 was detected at -2496 bp near the DLX2 RE (Suppl Figure 9B). TEAD1, TEAD3 and TEAD4, which
451 are highly expressed in neck-derived adipocytes, showed a negative correlation with *UCP1*
452 gene expression, while TEAD2 showed a positive correlation. At the same time, NFKB2, which
453 has a putative RE at -3976 bp, and ZNF410 and ZNF283 with RE at -3947 bp and -3948 bp,
454 respectively, possibly form a regulatory complex with HOXA2 (Suppl Fig 9C). While NFKB2
455 showed a negative correlation trend with *UCP1* gene expression, ZNF410 and ZNF283 showed
456 no obvious correlation (Suppl Table 3). Importantly, these TF REs detected near the
457 homeodomain family TF REs were also identified at a more stringent search threshold (p < 1e-
458 5), emphasising the likelihood of their involvement in regulatory complexes.

459 Additionally, our investigation extended to the vicinity of the unique NFIL3 RE identified at -
460 881 bp within the human *UCP1* promoter (Table 2A). Here, the presence of the REL RE was
461 detected at -902 bp (Suppl. Figure 9D, 9E) and also confirmed the conservation of these RE in
462 the *UCP1* promoter in primates, whereas their absence in the mouse and rat genomes (Suppl.
463 Figure 9F).

464 *2.5 Common RE of the orthologous UCP1 promoters occur at different genomic locations*
465 *in humans, mice and rats and show no sequence conservation at a particular genomic*
466 *position*

467 Examination and comparison of the regulatory components in the human, mouse and rat
468 *UCP1* promoters revealed a number of TF REs present in the human *UCP1* promoter, including
469 ARID3B, ATF7, CREB3, MAF::NFE2, NFE2L2, POU6F1, RXRA::VDR, SOX21, TEAD4 and XBP1,
470 which are absent in the human *UCP2* and *UCP3* promoters, but also have REs in the mouse
471 and rat *UCP1* promoters. This group consists mainly of TFs known to be involved in *UCP1* gene
472 expression, including well-characterised TFs such as CREB/ATF (Cao et al 2001; Liu et al 2019),
473 RXR (Alvarez et al 1995), NFE2 and NFE2L2 (Rim and Kozak 2002). A closer look at their gene
474 expression profiles (Fig. 2J) and the prediction of signalling pathways with AdipoNET suggests
475 that their main function is to promote the white adipocyte phenotype, as shown by the strong
476 negative correlation of POU6F1, TEAD4, XBP1 and NFE2L2 gene expression with *UCP1* (Table
477 3A). Only RXRA gene expression shows a positive correlation with *UCP1* (Suppl Table 3),
478 suggesting a potential enhancer role in browning processes. Consistent with this, AdipoNET
479 predictions suggest that RXRA, CREB3 and XBP1 may contribute to the development of both
480 white and brown adipocyte phenotypes. Interestingly, only POU6F1 from the family of
481 homeodomain TFs was exclusively associated with *UCP1* (it is absent in the promoter of *UCP2*

and *UCP3* in mice and rats), indicating its possible role in the specific regulation of the *UCP1* gene and thermogenesis in these species. This observation suggests the presence of evolutionarily conserved TF REs in the *UCP1* promoter in different species. However, a comparative analysis of the regulatory sequences of UCP1 in different species (https://genome.ucsc.edu/) revealed that most of these REs were detectable in the consensus sequences of primates, but not in mice and rats. This may indicate that these REs are not conserved in the consensus sequence of these species, as they occur at different positions in the UCP1 promoter of mice and rats (Suppl. Figure 10A, 10B; for the positions of the TF REs in the different species, see Suppl. Table 2).

Table 3.

3A Human, mouse and rat *UCP1* specific TFs

TF	RE position in human UCP1	GE Corr. TF vs.UCP1	RNA Expression in Adipocytes	ChipSeq data	PubMED abstract with UCP1	PWM in JASPAR	TF Class/Family	AdipoNET	EPD p<1e-5
POU6F1	-4716 -4715 -2913 -2912	NEG. **	MEDIUM	NO	NO		HDF1/POU domain factors	NO	NO
RXRA::VDR	-769	POS.****/ NEG.****	HIGH/HIGH	ChiPBase/ TFlink	YES		NR with C4 zinc fingers, (NR2):: (NR1)	B and W LINKER	NO
SOX21	-2243	NO	LOW	NO	NO		HMG domain/Sox-related	NO	YES
TEAD4	-2496 419 878	NEG. ****	MEDIUM	ChiPBase/ TFlink	NO		TEA domain factors, TEF-1-related factors	NO	YES
XBP1	-3762	NEG. ****	HIGH	TFlink	YES		bZIP/ XBP-related	B and W LINKER	NO
ARID3B	-4116 -3113	NO	LOW	NO	NO		ARID/ARID related	NO	NO
ATF7	-880 -879	NO	MEDIUM	NO	NO		bZIP/JUN-related	NO	NO
CREB3	-3761	NO	MEDIUM	NO	NO		bZIP/CREB-releted	B and W LINKER	NO
Maf::NFE2	-4388 -3441	NO/NO	HIGH/LOW	ChiPBase	YES		bZIP/Maf-related::Jun-related	WHITE	NO
NFE2L2	-4392 -3760 -3437	NEG. ****	HIGH	NO	YES		bZIP/JUN-related	WHITE	YES

3B. TFs with RE in the human *UCP1* promoter based on EPD (p<1e-4) and Chip-seq data (ChipBase and Tflink databases).

TF	RE position in human UCP1	GE. Corr. TF vs.UCP1	RNA Expression in Adipocytes	ChipSeq data	PubMED abstract with UCP1	PWM in JASPAR	TF Class/Family	AdipoNET	EPD p<1e-5
GATA2	-2789	Neg. **	Low	ChiPBase/ TFlink	Yes		Other C4 ZF/C4 GATA-related	NO	NO
HIF1A	-264 -261 -214	Neg. ****	High	ChiPBase/ TFlink	Yes		BHLH class/ PAS domain	WHITE	YES
JUN	-2832 -2586 -878 -585	Neg *.	High	ChiPBase/ TFlink	Yes		bZIP/JUN-related	B and W LINKER	YES
KLF4	-4653	Neg. ***	High	ChiPBase/ TFlink	Yes		C2H2 ZF/3 ZF Kruppel-related	WHITE	NO
MXI1	-263 -262	Neg. ****	Medium	ChiPBase/ TFlink	No		BHLH class; BHLH-ZIP	NO	NO
MYC	-263 -262	Neg. ***	Medium	ChiPBase/ TFlink	Yes		BHLH class; BHLH-ZIP	WHITE	NO
MYCN	-2593 -2592 -263 -262	NA	No	ChiPBase/ TFlink	No		BHLH class; BHLH-ZIP	WHITE	NO
NFKB1	-3976 -2405	Neg. ****	Medium	ChiPBase/ TFlink	Yes		RHR/NFKB-related	B and W LINKER	NO
RBPJ	-3973 -3944 -3610 -1505 -780 519	Neg. ****	High	ChiPBase/ TFlink	Yes		RHR/CSL-related	WHITE	YES
TEAD4	-2496 419 878	Neg. ****	Medium	ChiPBase/ TFlink	No		TEA domain factors/ TEF-1-related factors	NO	YES
USF1	-2593	Pos. ****	Medium	ChiPBase/ TFlink	No		BHLH class; BHLH-ZIP	NO	NO

Table 3. Tables 3A and 3B summarize our data on TFs with RE in the human *UCP1* promoter identified using EPD (p<1e-4) and filtered according to specific criteria. These include the position of their RE relative to the TSS, their gene expression pattern relative to *UCP1*, their gene expression levels measured in neck-derived preadipocytes and in vitro differentiated adipocytes, and their ChipSeq. data from the ChiPBase and TFlink databases, their classification into TF class and family based on JASPAR databases, their PWM, their involvement in the adipocyte phenotype formation pathway according to the AdipoNET interface, EPD TF RE search results at p<1e-5 significance level in the human *UCP1* promoter. Bold highlights TF or value outstanding in the related category. **3A** The table summarizes the 12 TFs with RE found to be common in human, mouse and rat *UCP1.* **3B.** The table summarizes the data of TFs with RE (EPD p<1e-4) in the human *UCP1* promoter that also has ChipSeq data from both ChiPBase and TFlink databases. GE: Gene Expression, Corr.: Correlation.

2.6 Common human-specific bHLH transcription factor response elements in the proximal promoter of UCP1-2-3 genes

Finally, our comparative analysis revealed four regulatory TF elements (HES1, MAX, MNT, TCFL5) that occur exclusively in human *UCP* genes (Suppl Table 1). These TFs have binding sequences within the regulatory domains of the human *UCP1, UCP2* and *UCP3* genes, but are not present in the promoters of the analysed mouse and rat *UCP* genes (Suppl Fig. 11A).

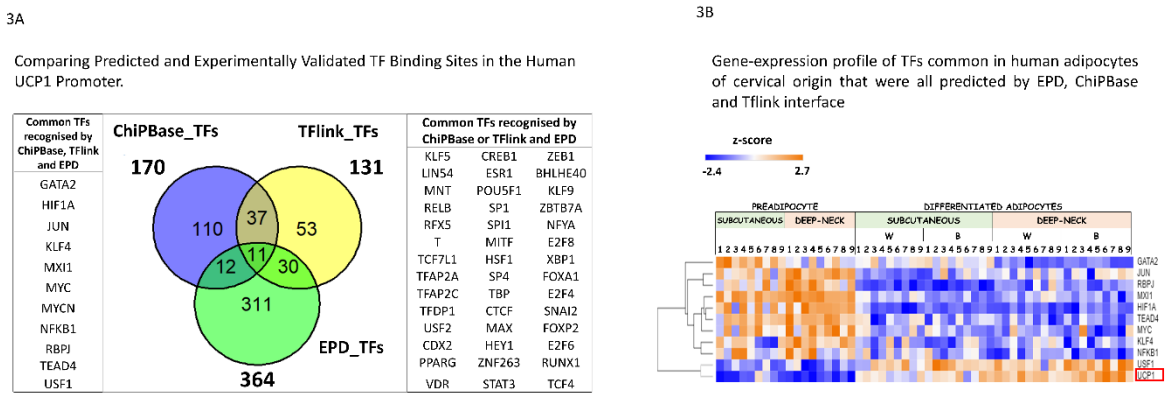
513 Interestingly, all of these TFs have RE in the proximal promoter region of human *UCP1* at -263
514 bp. This RE is recognised by the TFs of the basic helix-loop-helix (bHLH) family, and the
515 Eukaryotic Promoter Database assigns additional TFs to this nucleotide sequence beyond the
516 four mentioned above (Suppl Table 2). Therefore, further studies are needed to determine
517 which context-dependent regulatory mechanisms might enable the TFs to actually be users of
518 this RE.

519 2.7 *Predicted and experimentally validated TF binding sites in the human UCP1 promoter*
520 *show little overlap*

521 In our study, we performed a detailed comparative analysis of putative TF REs identified using
522 the Eukaryotic Promoter Database (https://epd.epfl.ch) in the regulatory region of the human
523 UCP1 gene extending from -5000 to +1000 bases from the TSS, with a significance level of $p <$
524 1e-4. We than compared these results with experimentally validated binding proteins/TFs for
525 the same DNA sequence using the ChiPBase (https://rnasysu.com) and TFlink
526 (https://tflink.net/) databases. While we identified 364 putative TF REs using the Eukaryotic
527 Promoter Database, ChiPBase recognised 170 and TFlink 131 experimentally validated TFs
528 (Fig. 3A). Significantly, only a small proportion of the experimentally identified binding TFs had
529 a matching RE recognised by the Eukaryotic Promoter Database (ChiPBase: 23 TFs, 14%; TFlink:
530 41 TFs, 31%). Conversely, we found that only 53 TFs (15%) of those with REs according to the
531 Eukaryotic Promoter Database were associated with laboratory validation based on these two
532 databases. Note that the two Chip-Seq databases also have only a small overlap. Notably, only
533 11 TFs were identified in all three databases (GATA2, HIF1A, JUN, KLF4, MXI1, MYC, MYCN,
534 NFKB1, RBPJ, TEAD4, and USF1), highlighting their potential role in regulating human *UCP1*
535 gene expression. The gene expression patterns of these TFs in neck-derived adipocyte samples

536 (Tóth et al. 2020) further support this, with most showing a strong correlation with *UCP1*,
537 mainly negative, while Upstream Transcription Factor 1 (USF1) showed a positive correlation
538 (Fig. 3B). USF1, whose REs were identified in all *UCP*s of the three species studied, appears to
539 act as a general regulator of gene expression of *UCP*s and not as a specific regulator of *UCP1*
540 and browning processes (Suppl Table 2) (Laurila et al. 2016). While the predictions of
541 AdipoNET also suggest that most of these TFs are involved in the development of the white
542 adipocyte phenotype, JUN and NFKB1 may play a role in both white and brown adipocyte
543 phenotypes (Table 3B). Response elements of HIF1A, JUN, RBPJ and TEAD4 were identified at
544 a more stringent significance level ($p < 1e-5$), supporting their possible involvement in the
545 regulation of *UCP1* gene expression. Interestingly, four of these 11 TFs (HIF1A, MYC, MYCN
546 and MXI1) share a common binding site around -263 bp in the proximal promoter region,
547 which was previously identified in our study as a potential site for fine-tuned, state-dependent
548 regulation of human *UCP* gene expression (Suppl Fig 11A). This RE is not evolutionarily
549 conserved in the mouse and rat *UCP* genes, highlighting the unique control aspects of *UCP1* in
550 humans and the potential for the formation of a regulatory module with TFs SP2 and EGR1,
551 which have REs nearby (Suppl Table 2), based on data revealed by ConTra v3 (Suppl Figs 11B,
552 11C, 11E).

553 **Figure 3.**



554

555 **Figure 3. 3A** The Venn diagrams show the number of TFs with putative RE in the human *UCP1* promoter (-5000
556 – +1000 BP) based on EPD, ChiPBase and TFlink databases. TFs commonly found in the databases are named in
557 the side panels (Venn program in R environment). **3B** The heatmap shows the gene expression profiles of 11 TFs
558 in human adipocytes of cervical origin that have been predicted or experimentally shown to have RE in the human
559 *UCP1* promoter based on EPD, ChiPBase and TFlink interfaces. For comparison, *UCP1* is also shown. Hierarchical
560 cluster analysis based on the Spearman Rank similarity matrix identifies TFs with similar gene expression profiles
561 to *UCP1* (Morpheus interface).

562 **DISCUSSION**

563 **1. Evolutionary dynamics of *UCP1* genes and their promoters differ in humans and**
564 **rodents**

565 1.1 *Research trends in the phylogenetic analysis of genes.*

566 *In silico* studies based on DNA sequences developed into a fundamental approach for
567 understanding the evolutionary dynamics of gene-specific functions. In general, phylogenetic
568 analyses have primarily focused on coding sequences, and there are now more than a hundred
569 species in comparative analyses of orthologous genes for *UCP1*, providing insights into
570 complex phylogenetic relationships (Cassard et al. 1990; Klingenspor et al. 2008; McGaugh
571 and Schwartz 2017, Hughes et al. 2009; Jastroch et al. 2005, 2008, 2018; Berg et al. 2006; Hou
572 et al. 2017; Gaudry et al. 2017). In the meantime, evolutionary studies on transcriptional
573 regulation have remained strikingly limited. In particular, Gaudry and Campbell (2017)
574 performed a comparative analysis of the transcriptional regulatory elements of *UCP1* in
575 mammals, although they focused only on the REs already described, particularly in the core
576 promoter and distal enhancer. Their work has shed light on evolutionarily conserved key sites
577 such as the TATA box, RARE-3 and PPRE motifs which are expected to be functionally relevant
578 in the phylogeny of eutherians (Gaurdy and Campbell, 2017; McGaugh and Schwartz, 2017).
579 Our studies demonstrate the ancient origin of these regulatory elements, indicating a
580 remarkable level of conservation. According to searches in the Eukaryotic Promoter Database,

581 these elements are widespread in the promoter regions of the *UCP* genes in all three species
582 studied. This abundance may indicates a fundamental role in the regulation of gene expression
583 within this gene family. However, due to their widespread distribution in the *UCP* family, these
584 elements may be less suitable as specific targets for precise modification of the transcription
585 of the *UCP1* gene, which is the ultimate goal of metabolic disorder control. Nonetheless, the
586 results of Gaudry and Campbell (2017) suggest that transcriptional control of gene expression
587 is only conserved to a limited extent within the mammalian clade. This observation suggests
588 the possible existence of novel regulatory and functional elements within the human *UCP1*
589 promoters that may adapt and coordinate to species-specific conditions and enable the
590 development of more targeted interventions.

591 *1.2 Differences between the regulatory region of UCP1 in humans and rodents*

592 Currently, several interfaces are developed for the prediction of DNA binding of TFs with
593 known PWM based on the promoter nucleotide sequence, such as the Eukaryotic Promoter
594 Database we used, TRANSFAC2.0 (https://genexplain.com/transfac-2-0/), GTRD (Yevshin et al.
595 2017) or ConTra v3. Using these interfaces, the set of potential regulatory sequences identified
596 in the 9 UCP promoters and their putative associated TFs supports the possibility of distinct
597 regulation of gene expression in these species, as similarity analysis of RE profiles reveals
598 differential clustering between human and rodent UCPs. In particular, the regulatory profile
599 of *UCP1* in humans shows the closest proximity to *UCP2* in humans, whereas *UCP1* in mice
600 shows a closer relationship to *UCP1* in rats. Our comparative analysis of upstream regulatory
601 segments in human, mice and rats *UCP1* genes reveals the intricate interplay between
602 sequence conservation and divergence that characterises the evolutionary trajectory of these
603 genes. While a substantial proportion of the identified cis-acting RE in these segments of
604 rodent and human *UCP1* genes are common (174 of 364; Suppl Figure 2B), broadly consistent

605 with previous findings (Collins et al., 2004), our analysis of sequence similarity reveals major
606 differences (Figure 1B), suggesting the possibility of the emergence of distinct regulatory
607 modules (Saito et al. 2008). Remarkably, a significant portion of the distal enhancer region of
608 the human *UCP1* promoter is missing in the mouse and rat genomes. This may be due to
609 secondary losses in the ancestors of mice and rats in the DNA segment that arose after the
610 opossum lineage split. Remarkably, these DNA sequence losses did not lead to extinction of
611 the rodent ancestors, suggesting that their existence was not essential for life or that possible
612 compensatory mechanisms allow sufficient TF binding to other parts of the regulatory region.
613 Despite this sequence deletion, human-specific *UCP1*-REs were identified in the DNA regions
614 whose consensus sequences are found in rodents, which could potentially form the basis for
615 human-specific UCP1 regulation. Recent studies also suggest interspecies variability in the
616 regulation of *UCP1* expression. They show that the 3'-untranslated region (3'-UTR) of *UCP1*
617 mRNA is processed differently in mice and humans, which quantitatively affects *UCP1*
618 synthesis and thermogenic activity (Lu et al., 2021). These differences urge caution when
619 interpreting research results obtained in rodent models to humans.

620 1.3 *Repeat elements in the human UCP1 promoter*

621 We identified numerous repeat elements in the human *UCP1* promoter region, including those
622 embedded in evolutionarily novel sequences that are absent in the corresponding rodent
623 regulatory regions (Fig. 1B). BLAST searches revealed the presence of these repeat sequences
624 at several locations in the human genome (Fig. 1H). However, the mechanisms underlying
625 their integration into a genome remain unclear, in particular whether these mobile elements
626 are randomly integrated or controlled by specific sequences. We observed the presence of
627 short palindromic sequences at the ends of the novel DNA segments, some of which are

628 annotated as TF REs in the Eukaryotic Promoter Database. Palindromic sequences in DNA are
629 traditionally recognised as specific binding sites for nucleotides or proteins, leading to
630 functional consequences (Pingoud and Jeltsch 2001; Linheiro and Bergman 2008). This may
631 suggest that these binding motifs could be ancient genetic elements used for integration into
632 the genome and repurposed by eukaryotic cells as regulatory modules to control gene
633 expression (Thompson et al. 2016). For example, the multiple integration of the MLT1G repeat
634 sequence into the human genome may organise the expression of nearby genes into a
635 regulatory circuit, leading to the simultaneous expression of the genes involved. This could
636 facilitate the development of complex processes such as adipocyte thermogenesis.

637 An interesting observation related to the 5'-TCTAATTAGA-'3 palindromic homeodomain RE
638 sequence identified at -690 bp in the human *UCP1* promoter may also support the repurposing
639 hypothesis. The same nucleotide sequence was previously recognised by Li and co-workers
640 (2005) as a characteristic DNA fragment located at one endpoint of the 1165 bp long sequence
641 that distinguishes two genotypic variants of Mamestra configurata nucleopolyhedrovirus-A
642 (MacoNPV-A). These variants include the lower virulence variant 90/4 and the archetypal
643 variant 90/2. They concluded that the recognised palindrome sequence raises the possibility
644 that the difference between these viral variants is due to the insertion or deletion of a
645 transposable element. RepeatMasker does not recognise this particular sequence region as a
646 repetitive element in humans, but mutational changes since the ancient insertion may have
647 masked the repetitive element.

648 Other studies also suggest that transposable elements and short tandem repeats may
649 influence the regulation of gene expression in the genome (Horton et al., 2023; Chuong et al.,
650 2016; Kunarso et al., 2010). For example, Scarpato and colleagues (2015) have suggested that

651 some repetitive elements, such as SINEs, are highly prone to hybridisation with microRNAs,
652 thereby modifying the transcriptional activity of nearby genes. Overall, our results suggest
653 that species-specific differences in transposable and repetitive elements in the *UCP1*
654 promoters may underlie gene-specific regulatory mechanisms.

655 **2. Comparative analyses of putative RE in *UCP1, UCP2* and *UCP3* promoters in humans,**
656 **mice and rats underline the possible regulatory role of TFs of the homeodomain**
657 **family.**

658 *2.1 The human UCP1-specific 5'-TCTAATTAGA-3' binding sequence*

659 One of the main goals of our study was to identify unique REs within the human *UCP1*
660 promoter which may have remained hidden in "exploratory" studies commonly performed in
661 rodents. Results indicate that most of the known putative regulatory factors of human *UCP1*
662 are general regulators of the studied *UCP*s, which classifies them as less specific targets.
663 Uncovering the specific regulatory elements allows for targeting metabolic processes by safely
664 promoting fat cell browning. Our study focused on the identified unique TF REs of the
665 homeodomain family of human *UCP1* and their vicinity, in particular a palindromic nucleotide
666 sequence, 5'-TCTAATTAGA-3', located at -690/689 bp in the human *UCP1* promoter (Fig. 2I,
667 Fig. 4). It confers the ability of homeodomain family TFs to bind to both DNA strands, allowing
668 dimerisation and stronger tethering (Jolma et al. 2013; Cain et al. 2023; Dror et al. 2014). It is
669 also possible that multiple TFs of the homeodomain family alternate the use of this binding
670 site depending on the circumstances, and these TFs could differ in whether they promote or
671 block gene expression (Foulkes et al. 1991). The homeodomain family is one of the largest
672 groups of TFs in metazoans and consists of almost 200 family members in humans (Lambert
673 et al. 2018). Their recognised functions include the control of developmental and homeostatic
674 processes through the modulation of DNA availability (Waldén et al. 2012; Hilton et al. 2015;

675 Cantile et al. 2003), but their potential for species-specific regulation of adipocyte phenotype
676 remodelling has been underestimated. These TFs can bind very similar AT-rich monomer
677 sequences, with small differences in nucleotides being sufficient for a specific interaction and
678 modulating the openness of DNA to allow or deny TFs access to nearby REs (Park et al. 2009).
679 In our analysis, regulatory elements such as the bHLH type 2 TFs, BCL6 and RFX5, identified
680 near this palindromic sequence unique to human *UCP1*, may further increase the likelihood of
681 binding by forming a complex/enhanceosome with homeodomain family TFs via a process
682 termed "epistatic capture" (Emera and Wagner 2012). The combined effect of these factors
683 can lead to a regulatory output that is not simply additive but is influenced by the specific
684 combination of transcription factor binding. Furthermore, in a previous study analysing the
685 signalling pathways enriched by genes with differential expression in subcutaneous and deep
686 neck samples, both in preadipocytes and differentiated mature adipocytes, the homeodomain
687 family TF pathway was found to significantly dominate in subcutaneous adipocytes, indicating
688 their possible active involvement in the regulation of browning processes (Tóth et al. 2020).

689 *2.2 Prediction of transcription factors targeting the human-specific sequence '5'-*
690 *TCTAATTAGA-3'"*

691 Among the TFs of the homeodomain family, we propose several potential binding candidates
692 for the human UCP1-specific RE at -690 bp. EN1, which is abundant in human neck adipocytes
693 and whose gene expression negatively correlates with *UCP1*, has previously been shown to be
694 a positive regulator of brown adipogenesis in mice (Zhang et al. 2016; Singh et al. 2015; Atit
695 et al. 2006), but was downregulated after animals were exposed to cold. These studies show
696 that EN1 is a marker for the brown adipocyte lineage, suggesting a role in brown adipocyte
697 development, but not necessarily a direct regulator of *UCP1* expression during thermogenic

activation in rodents. Our previous human study has shown that it is upregulated in the obesity phenotype in adipocytes carrying the single nucleotide polymorphism (SNP) FTO rs1421085 T-to-C (Tóth et al. 2020, unpublished data), suggesting a diverse, species-specific role in the regulation of *UCP1* expression. In contrast, another candidate, HOXA5 has been reported to be increasingly expressed in the thermogenic fat cells of rodents and to play a role in promoting adipose tissue browning through anti-inflammatory processes (Cao et al. 2018; Kato et al. 2021; Singh et al. 2016). Consistent with this study, HOXA5, showed a gene expression profile which may suggest its beneficial effect on browning in cervical-derived human adipocytes (Fig. 2H, Suppl Table 3)., At the same time, other human study showed increased expression in white adipocytes derived from induced pluripotent cells (Mohsen-Kanson et al. 2014). This discrepancy also suggests that HOXA5 may play a context-dependent role in the expression of the adipocyte phenotype in different species and fat depots.

The known role of PAX4 in pancreatic beta cell development and function is associated with the maintenance of glucose homeostasis (Brun et al. 2008; Xu et al. 2017), and abnormalities in its expression have been linked to metabolic diseases such as diabetes (Lau et al. 2023). However, its role in adipocyte function is less understood. PAX4 is expressed in mesenchymal progenitor cells of adipocytes (Xu et al. 2017), and our study suggests that it may be involved in the regulation of adipogenesis, including the balance between white and brown adipocyte formation (Suppl Fig. 6, Fig. 8). Role of the other candidate TFs of the homeodomain family, PRRX2 and VSX1/2, in the context of the thermogenic response has not yet been characterised. Nonetheless, single nuclear RNA sequencing studies have identified PRRX2 with differential expression activity in brown versus white adipogenesis in human cells (Gupta et al. 2023), indicating a possible role in thermogenic adipocyte formation.

721 The TF REs identified in the human *UCP1* promoter are often evolutionarily conserved in
722 primates, but not in mice and rats. Even the identified common TF REs in orthologous *UCP1*
723 promoters, such as POU6F1 TF from the homeodomain family, are located in the promoter
724 segment in humans that is absent in mice and rats and are not conserved in the genomic
725 consensus sequences. This indicates a kind of "mobility" of the regulatory elements and
726 possible convergent evolutionary processes. Nonetheless, according to our data mining
727 efforts, POU6F1 has not been previously associated with adipocyte function. Recent research
728 has identified its role in lung biology and cancer-related processes, revealing its interaction
729 with RORA in the lung leading to downregulation of *HIF1A* gene expression and potential
730 inhibition of adenocarcinoma cell proliferation (Xiao et al., 2022). However, as shown in
731 previous studies (Takahashi et al., 2007), certain POU factors serve as effective mediators of
732 cell reprogramming. Given its strong expression in mature adipocytes and a remarkable
733 negative correlation with *UCP1* gene expression (Table 3A, Suppl Table 3), it is plausible that
734 POU6F1 may play a role in adipocyte phenotype switching and that this function is conserved
735 in mammalian evolution.

736 *2.3 Limitations in the identification of functional regulatory elements in the human UCP1*
737 *promoter*

738 Our comparative analysis revealed little overlap between the TF REs predicted in our study
739 based on DNA sequence analyses and those identified experimentally with ChipSeq in the
740 human *UCP1* promoter, highlighting the challenges in identifying functional regulatory
741 elements. TFs, like other proteins, have the capacity for non-specific DNA binding, which is
742 essential for efficient target site localisation by facilitating linear diffusion (Pingoud and Jeltsch
743 2001). This non-specific binding usually involves interactions with the DNA backbone and not
744 with the bases. In contrast, specific binding is characterised by a close interplay between direct

745 (interaction with the bases) and indirect (interaction with the backbone) selection. The
746 frequent occurrence of non-functional TF binding can lead to considerable ambiguity about
747 functional DNA-TF binding in ChipSeq laboratory experiments. In addition, the infinite
748 combinations of biological tissues, developmental stages and specific conditions significantly
749 affect *ad hoc* TF binding to specific DNA segments, making generalisation of results difficult.
750 While PWMs are useful for predicting functional TF binding within promoter sequences, they
751 also have numerous limitations. One of these is the random occurrence of short binding
752 sequences in the genome, which emphasises the need for rigorous statistical analysis to
753 distinguish between biologically meaningful TF binding events and random matches. In
754 addition, the use of species- and gene-specific screening approaches can further increase the
755 reliability of predictions. The used Eukaryotic Promoter Database employs robust statistical
756 tests to predict PWM in a DNA sequence and thus provides the best possible prediction. Other
757 limitations of the *in silico* approach are our incomplete knowledge of the binding sequences
758 of TFs and the inability of PWMs to fully capture the sequence flexibility of TF binding sites.
759 TFs can tolerate variations in DNA sequence at their binding sites, leading to a degree of
760 "sloppiness" or degeneracy. This sequence flexibility allows TFs to recognise and bind a range
761 of related sequences which may differ from the consensus motif defined by the PWM.
762 Therefore, PWMs may miss functional TF binding sites that differ from the consensus
763 sequence or erroneously predict non-functional sites that are very similar to it. In addition,
764 the PWMs do not take into account the context-dependent nature of TF binding, where
765 neighbouring DNA sequences or chromatin structures may influence binding specificity. As a
766 result, both predictions may miss biologically relevant binding events or lead to false positives.
767 This highlights the importance of integrating experimental and complementary computational
768 approaches to improve the accuracy of TF binding site identification.

769 **3. Hotspots in human UCP1 promoter**

770 The prediction of human *UCP1*-specific TFs identified using the Eukaryotic Promoter Database
771 indicates specific DNA regions with higher TF-RE density within the human *UCP1* promoter
772 (Fig. 4). Such an accumulation of regulatory binding sites may lead to the formation of
773 enhanceosomes, in which molecular elements such as TFs and cofactors form complex
774 multilevel regulatory networks that finely orchestrate gene expression. In particular, in the
775 enhancer region of the human *UCP1* promoter, novel regulatory cluster-forming nucleotide
776 sequences are observed at various homeodomain family binding sites, often located at the
777 beginning or end of DNA regions identified as repetitive elements. In addition, several REs
778 appear in the proximal promoter near the RE of the bHLH TF family identified at -262/-263 bp,
779 suggesting the possibility of regulatory complex formation (see Suppl. Figs 11B, 11C, 11D). The
780 proximity of EGR1 and SP2 binding sites provides a scenario for dynamic regulation that could
781 occur through interaction with HIF1A, modulating *UCP1* expression according to cellular
782 conditions. This proposed mechanism shares similarities with mechanisms described in non-
783 small cell lung cancer, where interactions between HIF1A and EGR1 as well as HIF1A and SP1
784 play a central role in regulating *EPOR* gene expression in response to changing hypoxic
785 conditions (Su et al. 2019). Given the known role of hypoxia and HIF1A in adipose tissue
786 function, the possible involvement of HIF1A in *UCP1* transcription through promoter binding
787 and interaction with other TFs could be an interesting area of research (Tóth et al. 2021; Basse
788 et al. 2017). In addition, EGR1 has already been shown to be involved in metabolic diseases
789 and can influence the insulin sensitivity of fat cells and decide on PI3K/Akt and MAPK signalling
790 pathways (Zhang et al. 2013). It is conceivable that slight variations in the concentrations of
791 different components within the *UCP1* enhanceosomes may interact synergistically and lead
792 to significant differences in *UCP1* expression, as has previously been hypothesised in

comparative studies in mice (Xue et al. 2005). Among the potentially related cofactors, the p300/CBP-associated factor with well-described histone acetyltransferase activity stands out as a frequently described partner of TFs with a putative RE in the human *UCP1* promoter, indicating an essential role for them in shaping the fat cell phenotype (Fig. 4). Nevertheless, PGC1A, another described coactivator of TFs that regulate *UCP1*, such as PPARG, appears to interact with TFs that lack *UCP* gene or species specificity (Figure 4). Determining the key components of the regulatory complexes, describing their role and identifying the components that can be replaced or omitted in the initiation of human UCP1 transcription under certain conditions is a promising future research direction.

Figure 4.

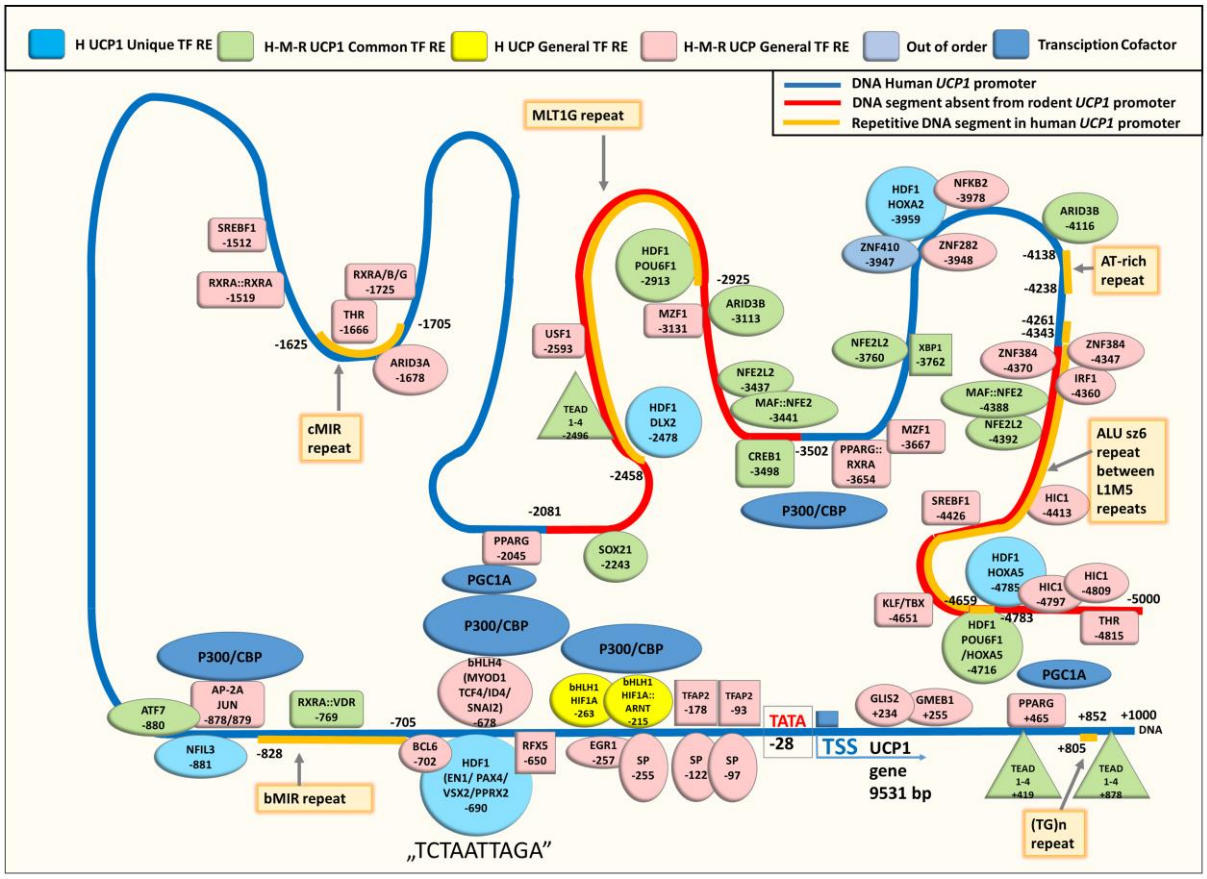


Figure 4. This figure provides a comprehensive overview of the transcription factor response elements identified and filtered by our study within the human *UCP1* promoter. These REs exhibit gene or species specificity. Additionally, several previously reported TF REs, which did not display such specificity, are included for reference.

807 The figure highlights specific promoter segments in the human *UCP1* promoter that are absent in the mouse and
808 rat *UCP1* promoters (indicated by the red line). The yellow line represents repeat elements identified by
809 RepeatMasker. Numerical annotations within the promoter denote base pair positions relative to the
810 transcription start site (TSS). For clarity, a list of TF abbreviations is provided in the "List of abbreviations" section
811 at the end of the article. H: human, M: mouse, R: rat. HDF: homeodomain family, bHLH: basic helix-loop-helix
812 family (PowerPoint).

813 **METHODOLOGY**

814 **1. Reconstruction of the Phylogenetic Trees**

815 To generate a combined human, mouse and rat UCP1, UCP2, and UCP3 amino acid, coding
816 sequence and promoter sequence phylogenetic tree, the amino acid and the mRNA data set
817 was downloaded from the National Library of Medicine (https://www.ncbi.nlm.nih.gov),
818 whereas the sequences of gene promoters were downloaded from the EPD database (link:
819 https://epd.epfl.ch/). Accession numbers of the promoters' FASTA sequences: human *UCP1*
820 FP007435, *UCP2_1* FP016503, *UCP2_2* FP016504, *UCP3_1* FP016505, *UCP3_2* FP016506;
821 mouse *UCP1* FP011827, *UCP2* FP010492, *UCP3* FP010490; rat *UCP1* FP011518, *UCP2*
822 FP000767, *UCP3* FP000766; Accession numbers of the mRNAs' FASTA sequences: human UCP1
823 NM_021833.5, UCP2 NM_001381943.1, UCP3 NM_003356.4; mouse UCP1 NM_009463.3,
824 UCP2 NM_011671.6, UCP3 NM_009464.3; rat UCP1 NM_012682.2, UCP2 NM_019354.3,
825 UCP3 NM_013167.2; Accession numbers of the amino-acids' FASTA sequences: human UCP1
826 NP_068605, UCP2 AAC51336, UCP3 AAC51367.1; mouse UCP1 NP_033489, UCP2 AAB17666,
827 UCP3 AAB87084; rat UCP1 NP_036814, UCP2 BAA25698, UCP3 AAB71523. The resulting
828 amino acid, exon and promoter sequences were aligned using MAFFT 7.505 (Katoh &
829 Standley, 2013), and a maximum likelihood tree was constructed using IQtree2 2.0.7 (Minh et
830 al., 2020) with automatic model selection ('-m MFP') turned on and using 1 000 aLRT and 1
831 000 rapid bootstrap replications to test the robustness of the analysis. The alignment of the
832 mRNA sequences had a length of 3,398 bp, of which 1536 were variable and 878 informative,
833 whereas the promoters had 7,956 variables and 5,412 informative sites on an alignment
834 length of 12585 bp. Of the 313 aligned amino acid positions, we identified 172 polymorphic
835 and 150 informative characters. Phylogenetic trees were visualized with packages ape 5.7.1
836 (Paradis & Schliep, 2019) and phangorn 2.11.1 (Schliep, 2011) using R 4.2.2 (R Foundation for
837 Statistical Computing, n.d.). Pairwise genetic distances of *UCP* promoters and exons were
838 calculated with the R package pegas 1.2 (Paradis, 2010) using the "K80" model of nucleotide

839 substitution (Kimura's 2-parameters distance metric; Kimura, 1980). Pairwise distances were
840 visualized as a neighbour-joining phylogram with pegas, as a UPGMA dendrogram with
841 phangorn, and as a heatmap with pheatmap 1.0.12 (https://cran.r-
842 project.org/web/packages/pheatmap/index.html).

843 **2. 3D Protein Structure Prediction, Alignment and Pairwise Structural Comparison**

844 The 3D structures of human, mouse and rat UCP1, UCP2 and UCP3 proteins were downloaded
845 using the AlphaFold 2 protein structure prediction interface (https://alphafold.ebi.ac.uk/)
846 developed by Deep Mind (Jumper et al. 2021). Colors indicate model confidence, AlphaFold
847 produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50
848 pLDDT may be unstructured in isolation. Dark blue: Very high (pLDDT > 90); Light blue:
849 Confident (90 > pLDDT > 70); Yellow: Low (70 > pLDDT > 50); Red: Very low (pLDDT < 50).

850 For alignment of the predicted protein structures we used the RCSB PBD
851 (https://www.rcsb.org; Berman et al. 2000) 3D protein structure comparison interface. For
852 protein comparisons, pairwise comparisons were performed with human UCP1. The metrics
853 used to describe the degree of pairwise structure alignment and structural similarity are as
854 follows:

855 **RMSD** (root mean square deviation) is computed between aligned pairs of the backbone C-
856 alpha atoms in superposed structures in Å resolution. The lower the RMSD, the better the
857 structure alignment between the pair of structures. This is the most commonly reported
858 metric when comparing two structures, but it is sensitive to local structure deviation. If a few
859 residues in a loop are not aligned, the RMSD value is large, even though the rest of the
860 structure is well aligned.

861 **TM-score** (template modeling score) is a measure of topological similarity between the
862 template and model structures. The TM-score ranges between 0 and 1, where 1 indicates a
863 perfect match and 0 is no match between the two structures. Scores < 0.2 usually indicate that
864 the proteins are unrelated while those > 0.5 generally have the same protein fold (e.g.,
865 classified by SCOP/CATH).

866 **Identity** (sequence identity percentage) is the percentage of paired residues in the alignment
867 that are identical in sequence.

868 **Equivalent Residues** is the number of residue pairs that are structurally equivalent in the
869 alignment.

870 **Sequence Length** is the total number of polymeric residues in the deposited sequence for a
871 given chain.

872 **Modeled Residues** is the number of residues with coordinates that were used for structure
873 alignment.

874 **3. BLAST Search to Identify Specific Nucleotide Sequences in the Human, Mouse and**
875 **Rat Genome**

876 The BLAST interface (https://blast.ncbi.nlm.nih.gov/) was used to identify sequences similar
877 to a given nucleotide sequence in the human, mouse and rat genomes (Altschul et al. 1997).
878 The human RefSeqGene, rat mRatBN7.2 and mouse GRCm39 reference genome sequences
879 were used in the searches. The graphical summary of the results in Suppl. Fig. 1H shows the
880 percentage of homology in different colors, while the position indicates the sequence region
881 of the analysed DNA segment in which the homologous segment is located, showing the hits
882 in separate rows.

883 **4. Sequence Presence Assessment for Estimating Evolutionary Divergence**

884 We used the phylogenetic tree based on the genome of the species provided by the Multiz
885 Alignment of 100 Species interface integrated into the UCSC genome browser. We then
886 manually checked the presence of the 1419 bp sequence missing in the mouse and rat
887 genomes in the corresponding sequences of the other species. In this way, we estimated the
888 evolutionary divergence based on the available genomic data and the presence of the
889 sequence in the different species.

890 **5. Computational Identification of Putative Transcription Factor Response Elements in**
891 **Gene Regulatory Sequences Using the EPD Interface**

892 The identification of putative TF REs within the regulatory sequences of the genes under study
893 was performed using the Eukaryotic Promoter Database (EPD) interface (https://epd.epfl.ch)
894 (Périer et al. 2000). EPD's approach to identify TF REs relies on the JASPAR Transcription Motifs
895 Library (specifically JASPAR CORE 2018 vertebrates) (Sandelin et al. 2004), which includes a
896 diverse set of TF motifs, including generic promoter motifs such as the TATA box, Initiator, GC

box and CCAAT box. This library contains a large body of information represented as a position weight matrix (PWM) for 579 TF-binding sequences. Many approaches have been developed to model and learn a protein's DNA-binding specificity, however, comparative analyses indicate that simple models based on mononucleotide PWMs trained by the best methods perform similarly to more complex models (Weirauch et al. 2013).

To determine whether a given binding sequence is present within the regulatory sequence of the target gene, the EPD interface uses a probability-based algorithm with four different p-value thresholds (p-value < 0.01, 0.001, 0.0001 and 0.00001). In this study, we focused on the two most stringent p-value thresholds ($p < 0.0001$ and $p < 0.00001$), as empirical evidence from the literature suggests that more permissive probabilities can lead to high false positive rates (Benita et al., 2009). However, our exploratory studies showed that for some genes with experimentally proven regulatory TFs, the software accurately identified the binding sequence only at $p < 0.0001$, as exemplified by the occurrence of the HIF1A TF RE for genes such as *VEGFA* and *GLUT1*. In addition, considering that every 4000 nucleotides, a given RE of average length 12 bp can be expected to be found randomly (Wunderlich and Mirny, 2009), the identified occurrence itself has limited meaning. It rather suggests that TFs for which we did not identify the RE are less likely to be involved in the regulation of *UCP* gene expression. Consequently, we presented our findings primarily utilizing the less stringent $p < 1e\text{-}4$ criteria. To mitigate the problem of false positives, we applied additional screening methods to filter out less relevant REs and emphasise potential species- and gene-specific TF-REs. Although a stricter threshold ($p < 1e\text{-}5$) would reduce the number of false positives, it would also increase the number of false negatives, potentially omitting significant findings. False-positive results can be managed through further filtering and by collecting more detailed information. In choosing the significance level ($p < 1e\text{-}4$), we were guided by the possibility of comparing the results of similar tests from paralogous and orthologous UCP promoters and thus filtering out false positives in the analysis. However, for reasons of comparison and transparency, we also present results with a more stringent threshold ($p < 1e\text{-}5$) in the tables.

In this study, we analysed regulatory sequences 5 kb upstream and 1 kb downstream of the TSS for the genes under study (eg.: human *UCP1*: GRCh37/hg19; Chr. 4: 141,489,115-141,495,114; TSS: Chr. 4:141,489,115). The relevant base-pair boundaries of the upstream

927 regulatory DNA segment may vary among genes, therefore the widest DNA region allowed by
928 the EPD interface was analysed. Data retrieval was automated using a custom script, which
929 we have made publicly available on GitHub, increasing transparency and reproducibility.

930 **6. Comparative Analysis of the Putative TFs in the Regulatory Sequences of Human,**
931 **Mouse and Rat *UCP1*, *UCP2* and *UCP3* Genes**

932 Data analysis and screening were conducted in an R interactive environment. Initially, data
933 were extracted from the EPD (https://epd.epfl.ch) interface as individual data files (text files),
934 corresponding to each target gene. EPD utilizes the JASPAR database, which contains PWMs
935 of TF REs for examining their presence within specific DNA sequences
936 (https://jaspar.genereg.net). The EPD database contains two promoter sequences for human
937 UCP2 and UCP3, as these genes have alternative transcription start sites, with the TSSs of the
938 two variants differing by a few 100 bp. Since we did not want to favour either variant, both
939 variants were included in the analysis. Subsequently, the data extracted for each gene were
940 consolidated into a unified database that compiled occurrence information for the 579 TF
941 binding sequences within the promoter regions of three genes: *UCP1*, *UCP2*, and *UCP3* in
942 humans, mice, and rats. This resulting matrix was then utilized to identify distinct clusters of
943 TF REs based on specific criteria. These criteria included identifying TF REs that were specific
944 to human *UCP1* (not present in human *UCP2* or *UCP3* but potentially present in mouse and rat
945 *UCP1*, *UCP2*, or *UCP3*), unique to human *UCP1* (exclusive to human *UCP1*, absent in human
946 *UCP2*, *UCP3* as well as in mouse and rat *UCP1*, *UCP2*, and *UCP3*), specific to *UCP1* (present in
947 both human and mouse and rat *UCP1*, absent in human *UCP2* and *UCP3* but potentially
948 present in mouse and rat *UCP2* and *UCP3*), common to human, mouse, and rat *UCP1* (absent
949 in human, mouse, and rat *UCP2* and *UCP3*), and general human *UCP* TF REs (occurring in both
950 human *UCP1* and *UCP2* and *UCP3* but not in the promoter regions of mouse and rat *UCP1*,
951 *UCP2*, and *UCP3* genes).

952 **7. Chromosome position of the investigated *UCP1*, *UCP2* and *UCP3* promoters and**
953 **genes in the human genom**

954 *UCP1* promoter in human: GRCh 37/hg19 Chr 4:141,489,115 – 141,495,114 size: 6000 bases;
955 *UCP1* gene in human: GRCh 37/hg19 Chr 4:141,480,585 – 141,490,115 size: 9531 bases, minus
956 strand; *UCP2* promoter in human: GRCh 37/hg19 Chr11:73,693,248 - 73,699,247 size: 6000

957 bases; *UCP2* gene in human: GRCh 37/hg19 Chr11:73,685,717 - 73,694,247 size: 8531 bases;
958 *UCP3* promoter in human: GRCh 37/hg19 Chr11:73,719,130 - 73,725,129 size: 6000 bases;
959 *UCP3* gene in human: GRCh 37/hg19 chr11:73,711,322 - 73,720,130; size: 8809 bases.

960 **8. Mapping The Regulatory DNA Regions of the Identified Putative TF REs in Human**
961 **UCP1 and Comparing them with Aligned DNA Sequences of Other Species**

962 During the course of the study, a group of TF REs with a distance between the filtered out TF
963 REs below +/- 40 bp were identified, potentially indicating a complex regulatory mechanism.
964 This clustering of TF binding sites aligns with a common phenomenon observed in numerous
965 regulatory regions of multicellular eukaryotes, as described in previous studies (Wunderlich
966 and Mirny 2009). Using the Contra v3 interface (http://bioit2.irc.ugent.be/contra/v3/; Kreft et
967 al. 2017) and the UCSC browser (http://genome.ucsc.edu/), putative human *UCP1* TF REs
968 identified by EPD and filtered by comparative analysis of orthologs and paralogs were
969 examined for nucleotide sequence at the defined positions. The results are fully linked to NCBI
970 (http://www.ncbi.nlm.nih.gov/) and UCSC (https://genome.ucsc.edu/). In addition, using
971 these interfaces, the human DNA sequence of the region of interest was compared with the
972 aligned DNA region of 100 vertebrate species to examine the degree of conservation of the TF
973 RE (hg19 Multiz Alignments of 100 Vertebrates).

974 **9. Heat Map Visualization and Hierarchical Cluster Analyses of the TF RE Profile of the**
975 **UCP genes and the Transcriptomic Data of Human Adipocytes**

976 Hierarchical cluster analyses and heat map visualization were performed with the
977 Morpheus web tool (https://software.broadinstitute.org/morpheus/) using Spearman rank
978 correlation of rows and columns and complete linkage based on the presence or absence of
979 the TF RE in the given gene upstream regulatory region. The same tools were used to visualise
980 relative gene-expression profile of human neck-derived adipocytes (Tóth et al. 2020). RNAseq
981 data is deposited to Sequence Read Archive (SRA) database
982 (**https://www.ncbi.nlm.nih.gov/sra**) under accession number PRJNA607438. Heatmaps
983 depict the z-scores calculated from the filtered and normalized transcriptomic data of isolated
984 and maintained preadipocytes from subcutaneous and deep neck adipose stromal cells and

985 lipid-laden mature adipocytes differentiated *in vitro* with white and brown protocols
986 described in Tóth et al. (2020).

987 **10. Comparison of the Identified Putative TF REs in the Regulatory Sequence of UCP1**
988 **with the Corresponding Data Found in the Chip Based Databases**

989 The list of putative TF REs identified by the EPD interface (https://epd.epfl.ch) at p<1e-4
990 confidence level in human *UCP1* was compared with similar data from public ChipSeq
991 databases. ChipSeq databases collect experimentally validated DNA binding data for TFs to
992 the promoter of individual genes. Two ChipSeq databases, ChiPBase (https://rnasysu.com)
993 and TFlink (https://tflink.net/), were used to identify proteins/TFs binding in the regulatory
994 DNA sequence of the human *UCP1* gene range between -5000 bp and +1000 bp.

995 **11. Identification of Repeat Elements in the Genome**

996 The RepeatMasker interface (https://www.repeatmasker.org/) integrated into the UCSC
997 genome browser (https://genome.ucsc.edu/) was used to identify, categorize and determine
998 the exact position of repeat elements in the human *UCP1* promoter (GRCh 37/hg19 by NCBI
999 Gene, Chr4: 141.489.115-141.495.114).

1000 **12. Literature Mining Using a Custom Script**

1001 We programmed a simple Python script which uses the Entrez E-utilities (Sayers 2009) ESearch
1002 and EFetch to query the PubMed database and retrieve relevant abstracts. The script
1003 summarizes the search results in a convenient, human-readable tabular format which is
1004 manually examined by the researcher.

1005 We compiled a list of abbreviations of TFs of interest (e.g. HOXA5, JUN, TEAD1, etc.) based on
1006 the JASPAR database; Altogether the list of TFs we examined comprised 579 items.

1007 In addition, we added the UCP keyword to the search.

1008 For each TF in the list, the script iteratively executes searches for each pair consisting of that
1009 TF and a keyword, e.g. HOXA5 and UCP1, HIF1A and UCP1 etc. These pairs are concatenated
1010 with a '+' character and passed to ESearch, e.g. *'HOXA5+UCP1'*. When searching for a complex
1011 of several TFs rather than a single TF, e.g. EWSR1 and FLI1, the constituent TFs in the complex
1012 were also combined with a '+' character in the search, e.g. *'EWSR1+FLI1+UCP1'*.

1013 When searching for TF abbreviations that are identical or similar to a regular English word, in
1014 particular Ar (for which the database returns hits containing the verb form 'are'), CLOCK
1015 (English noun), REST (English noun and verb), MAX (short for maximum and a name) and JUN
1016 (very frequent abbreviation of the month June), we found that simply searching for these
1017 abbreviations returns a very large number of abstracts which contain the English word rather
1018 than the TF in question. Thus we had to add a further search term to disambiguate the
1019 abbreviation and narrow down the search to abstracts which contain a reference to the TF
1020 that we were interested in. We found that searching for *"Ar androgen receptor", "CLOCK*
1021 *circadian", "REST RE1 silencing", "MAX MYC associated x",* and *"JUN proto-oncogene"*
1022 respectively constrained the search effectively so that the results contained few irrelevant
1023 hits.

1024 All combinations of TFs and the keyword were searched, i.e. in total 579 searches were
1025 executed. For each such combined TF + keyword search expression, the ESearch tool returns
1026 a list of PubMed ids of articles the abstract of which contains both the TF and the keyword.
1027 For each TF the number of abstracts containing the keyword is summarized in a table by the
1028 script. In addition, the full text of all abstracts is downloaded and saved along with the
1029 bibliographical data of the corresponding articles, as well as the URL of the full text of the
1030 article where available up to a pre-set maximum number, e.g. 50 abstracts.

1031 Finally, the downloaded abstracts were manually reviewed. We considered the use of
1032 advanced natural language processing techniques, in particular large language models, to
1033 partially automate the review of the abstracts, but this was judged unnecessary for this
1034 specific set of results.

1035 **13. Identification of TF Motif Activity in Different Tissues and Exploration of TF**
1036 **Interaction Partners Using the ISMARA Online Platform.**

1037 We obtained RNA-seq expression profiles from 16 human cell types using the Illumina Body
1038 Map 2 dataset (GEO accession GSE30611), accessible within the Integrated System for Motif
1039 Activity Response Analysis (ISMARA) online platform
1040 (https://ismara.unibas.ch/supp/dataset1_IBM_v2/ismara_report/) (Balwierz et al. 2014). We
1041 employed ISMARA to analyze the motif activity of putative species-specific and species-
1042 common homeodomain family TFs across different human tissues for identification of TF motif

activity. By leveraging the ISMARA platform, we investigated the motif activity patterns of these TFs, providing insights into their potential roles in tissue-specific gene regulation. Furthermore, we explored potential TFs and their firs-level TF interaction partners using ISMARA to uncover regulatory networks and complex formations, such as enhanceosomes.

14. Statistical Analyses

Statistical analysis was carried out using the stats package in R. Correlation analysis of *UCP1* gene expression with TFs showing gene expression in adipocytes from the cervical region (Tóth et al. 2020) was performed using Sperman's rank correlation. For multiple comparisons, an adjusted p-value was calculated using Holm's test, with $p < 0.05$ considered statistically significant (Suppl. Table 3).

Supplementary Table Legends

Supplementary Table 1 The table lists TFs identified through RE analysis in the promoters of human, mouse, and rat UCP1, UCP2, and UCP3 genes, categorized by species and genes, employing the EPD interface (p<1e-4) for identification. Subgroup categorization is based on the Venn diagram representation presented in Figure 1D.

Supplementary Table 2 The table presents the list of TFs identified through RE analysis in the promoters of human, mouse, and rat UCP1, UCP2, and UCP3 genes. The listing includes the TFs' relative positions to the TSS, categorized by species and genes, utilizing the EPD interface (p<1e-4, p<1e-5).

Supplementary Table 3 The table presents correlation coefficients depicting the associations between the gene expression profile of human UCP1 and the TFs identified within human UCP1 promoters using putative REs. The data utilized for this analysis were derived from RNA sequencing of human adipocytes obtained from the cervical region.

Supplementary Table 4. The table shows the TF abbreviations and the number of abstracts where they occur together with the keyword UCP1 in publications in the PubMed database.

DECLARATIONS

Ethics approval and consent to participate

Not Applicable

Consent for Publication

Not Applicable

Availability of data and material

RNAseq data have been deposited to Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA607438. We automated the retrieval of EPD data and the screening of abstracts from the PubMed database using custom scripts that were made publicly available on GitHub (https://github.com/DEpt-metagenom/TFRE-in-UCP1-promoter). The main and supplementary figures and

tables are available in the Mendeley database: https://data.mendeley.com/preview/z5b6stz3zx?a=828a4650-5239-487a-b17f-43c667b02454

Competing interests

The authors declare no conflict of interest.

Funding

During this research BBT, LL, EV and GP was funded by the European Union and the European Regional Development Fund (GINOP-2.3.4-15-2016-00002 „A felsőoktatás és az ipar együttműködése az egészségiparban”), GH was supported by the European Union and the European Regional Development Fund (GINOP-2.3.4-15-2020-00008 „Komplex Egészségipari Multidiszciplináris Kompetencia Központ kialakítása a Debreceni Egyetemen új innovatív termékek és technológiák fejlesztése érdekében”), and we are thankful for this. BBT was also supported by the University of Debrecen Scientific Research Bridging Fund (DETKA24). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

Conceptualization, BBT., methodology, BBT., GH, EV, GP, and LL; software, BBT., GH, EV, GP and LL; formal analysis, BBT and GP; visualization, BBT and LL; investigation, BBT.; data curation, BBT; writing—original draft preparation, BBT.; writing—review and editing, BBT., PG, LL, LF; project administration, LL;

Acknowledgements

We extend our appreciation to Professors Jan Nedergaard and Barbara Canon for their insightful feedback and valuable suggestions to enhance the clarity of the manuscript. On behalf of the project "Bioinformatic analysis of third generation sequencing data" ("Harmadik generációs szekvenálási adatok bioinformatikai elemzése") we are grateful for the possibility to use HUN-REN Cloud (see Héder et al. 2022; https://science-cloud.hu/) which helped us achieve the results published in this paper.

REFERENCES

1. Alvarez R, de Andrés J, Yubero P, Viñas O, Mampel T, Iglesias R, Giralt M, Villarroya F. A novel regulatory pathway of brown fat thermogenesis. Retinoic acid is a transcriptional activator of the mitochondrial uncoupling protein gene. J Biol Chem. 1995 Mar 10;270(10):5666-73. doi: 10.1074/jbc.270.10.5666. PMID: 7890689.
2. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402. doi: 10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.
3. Ambele MA, Pepper MS. Identification of transcription factors potentially involved in human adipogenesis in vitro. Mol Genet Genomic Med. 2017 Mar 3;5(3):210-222. doi: 10.1002/mgg3.269. PMID: 28546992; PMCID: PMC5441431.

4. Atit R, Sgaier SK, Mohamed OA, Taketo MM, Dufort D, Joyner AL, Niswander L, Conlon RA. Beta-catenin activation is necessary and sufficient to specify the dorsal dermal fate in the mouse. Dev Biol. 2006 Aug 1;296(1):164-76. doi: 10.1016/j.ydbio.2006.04.449. Epub 2006 Apr 21. PMID: 16730693.
5. Balwierz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, van Nimwegen E. ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. Genome Res. 2014 May;24(5):869-84. doi: 10.1101/gr.169508.113. Epub 2014 Feb 10. PMID: 24515121; PMCID: PMC4009616.
6. Banerjee-Basu S, Baxevanis AD. Molecular evolution of the homeodomain family of transcription factors. Nucleic Acids Res. 2001 Aug 1;29(15):3258-69. doi: 10.1093/nar/29.15.3258. PMID: 11470884; PMCID: PMC55828.
7. Basse AL, Isidor MS, Winther S, Skjoldborg NB, Murholm M, Andersen ES, Pedersen SB, Wolfrum C, Quistorff B, Hansen JB. Regulation of glycolysis in brown adipocytes by HIF-1α. Sci Rep. 2017 Jun 22;7(1):4052. doi: 10.1038/s41598-017-04246-y. PMID: 28642579; PMCID: PMC5481455.
8. Becher T, Palanisamy S, Kramer DJ, Eljalby M, Marx SJ, Wibmer AG, Butler SD, Jiang CS, Vaughan R, Schöder H, Mark A, Cohen P. Brown adipose tissue is associated with cardiometabolic health. Nat Med. 2021 Jan;27(1):58-65. doi: 10.1038/s41591-020-1126-7. Epub 2021 Jan 4. PMID: 33398160; PMCID: PMC8461455.
9. Benita Y, Kikuchi H, Smith AD, Zhang MQ, Chung DC, Xavier RJ. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. Nucleic Acids Res. 2009 Aug;37(14):4587-602. doi: 10.1093/nar/gkp425. Epub 2009 Jun 2. PMID: 19491311; PMCID: PMC2724271.
10. Berg F, Gustafson U, Andersson L. The uncoupling protein 1 gene (UCP1) is disrupted in the pig lineage: a genetic explanation for poor thermoregulation in piglets. PLoS Genet. 2006 Aug 18;2(8):e129. doi: 10.1371/journal.pgen.0020129. PMID: 16933999; PMCID: PMC1550502.
11. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235-42. doi: 10.1093/nar/28.1.235. PMID: 10592235; PMCID: PMC102472.
12. Bonet ML, Oliver P, Palou A. Pharmacological and nutritional agents promoting browning of white adipose tissue. Biochim Biophys Acta. 2013 May;1831(5):969-85. doi: 10.1016/j.bbalip.2012.12.002. Epub 2012 Dec 12. PMID: 23246573.
13. Brun T, Hu He KH, Lupi R, Boehm B, Wojtusciszyn A, Sauter N, Donath M, Marchetti P, Maedler K, Gauthier BR. The diabetes-linked transcription factor Pax4 is expressed in human pancreatic islets and is activated by mitogens and GLP-1. Hum Mol Genet. 2008 Feb 15;17(4):478-89. doi: 10.1093/hmg/ddm325. Epub 2007 Nov 7. PMID: 17989064.
14. Cadenas S. Mitochondrial uncoupling, ROS generation and cardioprotection. Biochim Biophys Acta Bioenerg. 2018 Sep;1859(9):940-950. doi: 10.1016/j.bbabio.2018.05.019. Epub 2018 May 31. PMID: 29859845.
15. Cain B, Webb J, Yuan Z, Cheung D, Lim HW, Kovall RA, Weirauch MT, Gebelein B. Prediction of cooperative homeodomain DNA binding sites from high-throughput-SELEX data. Nucleic Acids Res. 2023 Jul 7;51(12):6055-6072. doi: 10.1093/nar/gkad318. PMID: 37114997; PMCID: PMC10325903.

16. Cantile M, Procino A, D'Armiento M, Cindolo L, Cillo C. HOX gene network is involved in the transcriptional regulation of in vivo human adipogenesis. J Cell Physiol. 2003 Feb;194(2):225-36. doi: 10.1002/jcp.10210. PMID: 12494461.
17. Cao W, Huang H, Xia T, Liu C, Muhammad S, Sun C. Homeobox a5 Promotes White Adipose Tissue Browning Through Inhibition of the Tenascin C/Toll-Like Receptor 4/Nuclear Factor Kappa B Inflammatory Signaling in Mice. Front Immunol. 2018 Mar 29;9:647. doi: 10.3389/fimmu.2018.00647. PMID: 29651293; PMCID: PMC5884924.
18. Cao W, Medvedev AV, Daniel KW, Collins S. beta-Adrenergic activation of p38 MAP kinase in adipocytes: cAMP induction of the uncoupling protein 1 (UCP1) gene requires p38 MAP kinase. J Biol Chem. 2001 Jul 20;276(29):27077-82. doi: 10.1074/jbc.M101049200. Epub 2001 May 21. PMID: 11369767.
19. Cassard AM, Bouillaud F, Mattei MG, Hentz E, Raimbault S, Thomas M, Ricquier D. Human uncoupling protein gene: structure, comparison with rat gene, and assignment to the long arm of chromosome 4. J Cell Biochem. 1990 Jul;43(3):255-64. doi: 10.1002/jcb.240430306. PMID: 2380264.
20. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016 Mar 4;351(6277):1083-7. doi: 10.1126/science.aad5497. PMID: 26941318; PMCID: PMC4887275.
21. Collins S, Cao W, Robidoux J. Learning new tricks from old dogs: beta-adrenergic receptors teach new lessons on firing up adipose tissue metabolism. Mol Endocrinol. 2004 Sep;18(9):2123-31. doi: 10.1210/me.2004-0193. Epub 2004 Jul 8. PMID: 15243132.
22. Crichton PG, Lee Y, Kunji ER. The molecular features of uncoupling protein 1 support a conventional mitochondrial carrier-like mechanism. Biochimie. 2017 Mar;134:35-50. doi: 10.1016/j.biochi.2016.12.016. Epub 2017 Jan 3. PMID: 28057583; PMCID: PMC5395090.
23. del Mar Gonzalez-Barroso M, Pecqueur C, Gelly C, Sanchis D, Alves-Guerra MC, Bouillaud F, Ricquier D, Cassard-Doulcier AM. Transcriptional activation of the human ucp1 gene in a rodent cell line. Synergism of retinoids, isoproterenol, and thiazolidinedione is mediated by a multipartite response element. J Biol Chem. 2000 Oct 13;275(41):31722-32. doi: 10.1074/jbc.M001678200. PMID: 10921912.
24. Dreos R, Ambrosini G, Groux R, Cavin Périer R, Bucher P. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. Nucleic Acids Res. 2017;45(D1):D51–5. https://doi.org/10.1093/nar/gkw1069 Epub 2016 Nov 28. PMID: 27899657; PMCID: PMC5210552.
25. Dror I, Zhou T, Mandel-Gutfreund Y, Rohs R. Covariation between homeodomain transcription factors and the shape of their DNA binding sites. Nucleic Acids Res. 2014 Jan;42(1):430-41. doi: 10.1093/nar/gkt862. Epub 2013 Sep 27. PMID: 24078250; PMCID: PMC3874178.
26. Emera D, Wagner GP. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc Natl Acad Sci U S A. 2012 Jul 10;109(28):11246-51. doi: 10.1073/pnas.1118566109. Epub 2012 Jun 25. PMID: 22733751; PMCID: PMC3396485.
27. Enerbäck S, Jacobsson A, Simpson EM, Guerra C, Yamashita H, Harper ME, Kozak LP. Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. Nature. 1997 May 1;387(6628):90-4. doi: 10.1038/387090a0. PMID: 9139827.

1186 28. Feldmann HM, Golozoubova V, Cannon B, Nedergaard J. UCP1 ablation induces obesity and abolishes diet-
1187 induced thermogenesis in mice exempt from thermal stress by living at thermoneutrality. Cell Metab. 2009
1188 Feb;9(2):203-9. doi: 10.1016/j.cmet.2008.12.014. PMID: 19187776.
1189 29. Foulkes NS, Borrelli E, Sassone-Corsi P. CREM gene: use of alternative DNA-binding domains generates
1190 multiple antagonists of cAMP-induced transcription. Cell. 1991 Feb 22;64(4):739-49. doi: 10.1016/0092-
1191 8674(91)90503-q. PMID: 1847666.
1192 30. Gaudry MJ, Campbell KL. Evolution of *UCP1* Transcriptional Regulatory Elements Across the Mammalian
1193 Phylogeny. Front Physiol. 2017 Sep 20;8:670. doi: 10.3389/fphys.2017.00670. PMID: 28979209; PMCID:
1194 PMC5611445.
1195 31. Gaudry MJ, Jastroch M, Treberg JR, Hofreiter M, Paijmans JLA, Starrett J, et al. Inactivation of thermogenic
1196 UCP1 as a historical contingency in multiple placental mammal clades. *Sci Adv.* (2017) 3:e1602878
1197 10.1126/sciadv.1602878
1198 32. Gehring WJ, Affolter M, Bürglin T. Homeodomain proteins. Annu Rev Biochem. 1994;63:487-526. doi:
1199 10.1146/annurev.bi.63.070194.002415. PMID: 7979246.
1200 33. Gupta A, Efthymiou V, Kodani SD, Shamsi F, Patti ME, Tseng YH, Streets A. Mapping the transcriptional
1201 landscape of human white and brown adipogenesis using single-nuclei RNA-seq. Mol Metab. 2023
1202 Aug;74:101746. doi: 10.1016/j.molmet.2023.101746. Epub 2023 Jun 5. PMID: 37286033; PMCID:
1203 PMC10338377.
1204 34. Harms MJ, Lim HW, Ho Y, Shapira SN, Ishibashi J, Rajakumari S, Steger DJ, Lazar MA, Won KJ, Seale P. PRDM16
1205 binds MED1 and controls chromatin architecture to determine a brown fat transcriptional program. Genes
1206 Dev. 2015 Feb 1;29(3):298-307. doi: 10.1101/gad.252734.114. PMID: 25644604; PMCID: PMC4318146.
1207 35. Héder, M., Rigó, E., Medgyesi, D., Lovas, R., Tenczer, S., Török, F., Farkas, A., Emődi, M., Kadlecsik, J., Mező,
1208 G., Pintér, Á., & Kacsuk, P. (2022). The Past, Present and Future of the ELKH Cloud. Információs Társadalom,
1209 22(2), 128.
1210 36. Hilton C, Karpe F, Pinnick KE. Role of developmental transcription factors in white, brown and beige adipose
1211 tissues. Biochim Biophys Acta. 2015 May;1851(5):686-96. doi: 10.1016/j.bbalip.2015.02.003. Epub 2015 Feb
1212 8. PMID: 25668679.
1213 37. Himms-Hagen, J. and Harper, M. (2001). Physiological role of ucp3 may be export of fatty acids from
1214 mitochondria when fatty acid oxidation predominates: an hypothesis. Experimental Biology and Medicine,
1215 226(2), 78-84. https://doi.org/10.1177/153537020122600204
1216 38. Hirschenson J, Melgar-Bermudez E, Mailloux RJ. The Uncoupling Proteins: A Systematic Review on the
1217 Mechanism Used in the Prevention of Oxidative Stress. Antioxidants (Basel). 2022 Feb 6;11(2):322. doi:
1218 10.3390/antiox11020322. PMID: 35204205; PMCID: PMC8868465.
1219 39. Hoang T, Smith MD, Jelokhani-Niaraki M. Expression, folding, and proton transport activity of human
1220 uncoupling protein-1 (UCP1) in lipid membranes: evidence for associated functional forms. J Biol Chem. 2013
1221 Dec 20;288(51):36244-58. doi: 10.1074/jbc.M113.509935. Epub 2013 Nov 6. PMID: 24196960; PMCID:
1222 PMC3868739.

1223 40. Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, Shah N, Suzuki PH, Shrikumar
1224 A, Afek A, Greenleaf WJ, Gordân R, Zeitlinger J, Kundaje A, Fordyce PM. Short tandem repeats bind
1225 transcription factors to tune eukaryotic gene expression. Science. 2023 Sep 22;381(6664):eadd1250. doi:
1226 10.1126/science.add1250. Epub 2023 Sep 22. PMID: 37733848.
1227 41. Hou L, Shi J, Cao L, Xu G, Hu C, Wang C. Pig has no uncoupling protein 1. Biochem Biophys Res Commun.
1228 2017 Jun 10;487(4):795-800. doi: 10.1016/j.bbrc.2017.04.118. Epub 2017 Apr 23. PMID: 28442343.
1229 42. Hughes J, Criscuolo F. Evolutionary history of the UCP gene family: gene duplication and selection. BMC Evol
1230 Biol. 2008 Nov 3;8:306. doi: 10.1186/1471-2148-8-306. PMID: 18980678; PMCID: PMC2584656.
1231 43. Hughes DA, Jastroch M, Stoneking M, Klingenspor M. Molecular evolution of UCP1 and the evolutionary
1232 history of mammalian non-shivering thermogenesis. BMC Evol Biol. 2009 Jan 7;9:4. doi: 10.1186/1471-2148-
1233 9-4. PMID: 19128480; PMCID: PMC2627829.S
1234 44. Ikeda K, Kang Q, Yoneshiro T, Camporez JP, Maki H, Homma M, Shinoda K, Chen Y, Lu X, Maretich P, Tajima
1235 K, Ajuwon KM, Soga T, Kajimura S. UCP1-independent signaling involving SERCA2b-mediated calcium cycling
1236 regulates beige fat thermogenesis and systemic glucose homeostasis. Nat Med. 2017 Dec;23(12):1454-1465.
1237 doi: 10.1038/nm.4429. Epub 2017 Nov 13. PMID: 29131158; PMCID: PMC5727902.
1238 45. Inagaki T, Sakai J, Kajimura S. Transcriptional and epigenetic control of brown and beige adipose cell fate and
1239 function. Nat Rev Mol Cell Biol. 2016 Aug;17(8):480-95. doi: 10.1038/nrm.2016.62. Epub 2016 Jun 2. Erratum
1240 in: Nat Rev Mol Cell Biol. 2017 Aug;18(8):527. PMID: 27251423; PMCID: PMC4956538.
1241 46. Jastroch M, Oelkrug R, Keipert S. Insights into brown adipose tissue evolution and function from non-model
1242 organisms. *J Exp Biol.* (2018) 221(Pt Suppl 1):jeb169425. 10.1242/jeb.169425.
1243 47. Jastroch M, Wuertz S, Kloas W, Klingenspor M. Uncoupling protein 1 in fish uncovers an ancient evolutionary
1244 history of mammalian nonshivering thermogenesis. Physiol Genomics. 2005 Jul 14;22(2):150-6. doi:
1245 10.1152/physiolgenomics.00070.2005. Epub 2005 May 10. PMID: 15886331.
1246 48. Jastroch M, Withers KW, Taudien S, Frappell PB, Helwig M, Fromme T, Hirschberg V, Heldmaier G, McAllan
1247 BM, Firth BT, Burmester T, Platzer M, Klingenspor M. Marsupial uncoupling protein 1 sheds light on the
1248 evolution of mammalian nonshivering thermogenesis. Physiol Genomics. 2008 Jan 17;32(2):161-9. doi:
1249 10.1152/physiolgenomics.00183.2007. Epub 2007 Oct 30. PMID: 17971503.
1250 49. Jastroch M, Wuertz S, Kloas W, Klingenspor M. Uncoupling protein 1 in fish uncovers an ancient evolutionary
1251 history of mammalian nonshivering thermogenesis. Physiol Genomics. 2005 Jul 14;22(2):150-6. doi:
1252 10.1152/physiolgenomics.00070.2005. Epub 2005 May 10. PMID: 15886331.
1253 50. Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin
1254 K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J. DNA-
1255 binding specificities of human transcription factors. Cell. 2013 Jan 17;152(1-2):327-39. doi:
1256 10.1016/j.cell.2012.12.009. PMID: 23332764.
1257 51. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with
1258 AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
1259 52. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000
1260 Sep;16(9):418-20. doi: 10.1016/s0168-9525(00)02093-x. PMID: 10973072.

1261 53. Kazak L, Spiegelman BM. Mechanism of futile creatine cycling in thermogenesis. Am J Physiol Endocrinol
1262 Metab. 2020 Nov 1;319(5):E947-E949. doi: 10.1152/ajpendo.00444.2020. Epub 2020 Sep 21. PMID:
1263 32954828.
1264 54. Keipert S, Gaudry MJ, Kutschke M, Keuper M, Dela Rosa MAS, Cheng Y, Monroy Kuhn JM, Laterveer R, Cotrim
1265 CA, Giere P, Perocchi F, Feederle R, Crichton PG, Lutter D, Jastroch M. Two-stage evolution of mammalian
1266 adipose tissue thermogenesis. Science. 2024 Jun 7;384(6700):1111-1117. doi: 10.1126/science.adg1947.
1267 Epub 2024 Jun 6. PMID: 38843333.
1268 55. Kato H, Ario T, Kishida T, Tadano M, Osawa S, Maeda Y, Takakura H, Izawa T. Homeobox A5 and C10 genes
1269 modulate adaptation of brown adipose tissue during exercise training in juvenile rats. Exp Physiol. 2021
1270 Feb;106(2):463-474. doi: 10.1113/EP089114. Epub 2021 Jan 5. PMID: 33369800.
1271 56. Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements
1272 in Performance and Usability. Molecular Biology and Evolution, 30(4), 772–780.
1273 https://doi.org/10.1093/molbev/mst010
1274 57. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative
1275 studies of nucleotide sequences. J Mol Evol. 1980 Dec;16(2):111-20. doi: 10.1007/BF01731581. PMID:
1276 7463489.
1277 58. Klingenspor M, Fromme T, Hughes DA Jr, Manzke L, Polymeropoulos E, Riemann T, Trzcionka M, Hirschberg
1278 V, Jastroch M. An ancient look at UCP1. Biochim Biophys Acta. 2008 Jul-Aug;1777(7-8):637-41. doi:
1279 10.1016/j.bbabio.2008.03.006. Epub 2008 Mar 18. PMID: 18396149.
1280 59. Kottke FJ, Phalen JS, et al. Effect of hypoxia upon temperature regulation of mice, dogs, and man. Am J
1281 Physiol. 1948 Apr;153(1):10-5. doi: 10.1152/ajplegacy.1948.153.1.10. PMID: 18867027.
1282 60. Kreft L, Soete A, Hulpiau P, Botzki A, Saeys Y, De Bleser P. ConTra v3: a tool to identify transcription factor
1283 binding sites across species, update 2017. Nucleic Acids Res. 2017 Jul 3;45(W1):W490-W494. doi:
1284 10.1093/nar/gkx376. PMID: 28472390; PMCID: PMC5570180.
1285 61. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have
1286 rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010 Jul;42(7):631-4. doi:
1287 10.1038/ng.600. Epub 2010 Jun 6. PMID: 20526341.
1288 62. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The
1289 Human Transcription Factors. Cell. 2018 Feb 8;172(4):650-665. doi: 10.1016/j.cell.2018.01.029. Erratum in:
1290 Cell. 2018 Oct 4;175(2):598-599. PMID: 29425488.
1291 63. Lau HH, Krentz NAJ, Abaitua F, Perez-Alcantara M, Chan JW, Ajeian J, Ghosh S, Lee Y, Yang J, Thaman S,
1292 Champon B, Sun H, Jha A, Hoon S, Tan NS, Gardner DS, Kao SL, Tai ES, Gloyn AL, Teo AKK. PAX4 loss of
1293 function increases diabetes risk by altering human pancreatic endocrine cell development. Nat Commun.
1294 2023 Sep 30;14(1):6119. doi: 10.1038/s41467-023-41860-z. PMID: 37777536; PMCID: PMC10542369.
1295 64. Laurila PP, Soronen J, Kooijman S, Forsström S, Boon MR, Surakka I, Kaiharju E, Coomans CP, Van Den Berg
1296 SA, Autio A, Sarin AP, Kettunen J, Tikkanen E, Manninen T, Metso J, Silvennoinen R, Merikanto K, Ruuth M,
1297 Perttilä J, Mäkelä A, Isomi A, Tuomainen AM, Tikka A, Ramadan UA, Seppälä I, Lehtimäki T, Eriksson J,
1298 Havulinna A, Jula A, Karhunen PJ, Salomaa V, Perola M, Ehnholm C, Lee-Rueckert M, Van Eck M, Roivainen

1299 A, Taskinen MR, Peltonen L, Mervaala E, Jalanko A, Hohtola E, Olkkonen VM, Ripatti S, Kovanen PT, Rensen
1300 PC, Suomalainen A, Jauhiainen M. USF1 deficiency activates brown adipose tissue and improves
1301 cardiometabolic health. Sci Transl Med. 2016 Jan 27;8(323):323ra13. doi: 10.1126/scitranslmed.aad0015.
1302 PMID: 26819196.
1303 65. Li L, Li Q, Willis LG, Erlandson M, Theilmann DA, Donly C. Complete comparative genomic analysis of two
1304 field isolates of Mamestra configurata nucleopolyhedrovirus-A. J Gen Virol. 2005 Jan;86(Pt 1):91-105. doi:
1305 10.1099/vir.0.80488-0. PMID: 15604435.
1306 66. Li Y, Schwalie PC, Bast-Habersbrunner A, Mocek S, Russeil J, Fromme T, Deplancke B, Klingenspor M.
1307 Systems-Genetics-Based Inference of a Core Regulatory Network Underlying White Fat Browning. Cell Rep.
1308 2019 Dec 17;29(12):4099-4113.e5. doi: 10.1016/j.celrep.2019.11.053. PMID: 31851936.
1309 67. Linheiro RS, Bergman CM. Testing the palindromic target site model for DNA transposon insertion using the
1310 Drosophila melanogaster P-element. Nucleic Acids Res. 2008 Nov;36(19):6199-208. doi:
1311 10.1093/nar/gkn563. Epub 2008 Oct 1. PMID: 18829720; PMCID: PMC2577343.
1312 68. Liu Y, Maekawa T, Yoshida K, Muratani M, Chatton B, Ishii S. The Transcription Factor ATF7 Controls
1313 Adipocyte Differentiation and Thermogenic Gene Programming. iScience. 2019 Mar 29;13:98-112. doi:
1314 10.1016/j.isci.2019.02.013. Epub 2019 Feb 18. PMID: 30826729; PMCID: PMC6402263.
1315 69. Lu WH, Chang YM, Huang YS. Alternative Polyadenylation and Differential Regulation of *Ucp1*: Implications
1316 for Brown Adipose Tissue Thermogenesis Across Species. Front Pediatr. 2021 Feb 9;8:612279. doi:
1317 10.3389/fped.2020.612279. PMID: 33634052; PMCID: PMC7899972.
1318 70. McGaugh S, Schwartz TS. Here and there, but not everywhere: repeated loss of uncoupling protein 1 in
1319 amniotes. Biol Lett. 2017 Jan;13(1):20160749. doi: 10.1098/rsbl.2016.0749. PMID: 28052937; PMCID:
1320 PMC5310580.
1321 71. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R.
1322 (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.
1323 Molecular Biology and Evolution, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015
1324 72. Mohsen-Kanson T, Hafner AL, Wdziekonski B, Takashima Y, Villageois P, Carrière A, Svensson M, Bagnis C,
1325 Chignon-Sicard B, Svensson PA, Casteilla L, Smith A, Dani C. Differentiation of human induced pluripotent
1326 stem cells into brown and white adipocytes: role of Pax3. Stem Cells. 2014 Jun;32(6):1459-67. doi:
1327 10.1002/stem.1607. PMID: 24302443.
1328 73. Nedergaard J, Bengtsson T, Cannon B. Unexpected evidence for active brown adipose tissue in adult humans.
1329 Am J Physiol Endocrinol Metab. 2007 Aug;293(2):E444-52. doi: 10.1152/ajpendo.00691.2006. Epub 2007
1330 May 1. PMID: 17473055.
1331 74. Nedergaard J, Cannon B. The 'novel' 'uncoupling' proteins UCP2 and UCP3: what do they really do? Pros and
1332 cons for suggested functions. Exp Physiol. 2003 Jan;88(1):65-84. doi: 10.1113/eph8802502. PMID:
1333 12525856.
1334 75. Nedergaard J, Cannon B. The browning of white adipose tissue: some burning issues. Cell Metab. 2014 Sep
1335 2;20(3):396-407. doi: 10.1016/j.cmet.2014.07.005. Epub 2014 Aug 7. PMID: 25127354.

76. Nedergaard J, Golozoubova V, Matthias A, Asadi A, Jacobsson A, Cannon B. UCP1: the only protein able to mediate adaptive non-shivering thermogenesis and metabolic inefficiency. Biochim Biophys Acta. 2001 Mar 1;1504(1):82-106. doi: 10.1016/s0005-2728(00)00247-4. PMID: 11239487.
77. Nicholls DG. Mitochondrial proton leaks and uncoupling proteins. Biochim Biophys Acta Bioenerg. 2021 Jul 1;1862(7):148428. doi: 10.1016/j.bbabio.2021.148428. Epub 2021 Mar 31. PMID: 33798544.
78. Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. Bioinformatics, 26(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696
79. Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633
80. Park MJ, Kim HY, Kim K, Cheong J. Homeodomain transcription factor CDX1 is required for the transcriptional induction of PPARgamma in intestinal cell differentiation. FEBS Lett. 2009 Jan 5;583(1):29-35. doi: 10.1016/j.febslet.2008.11.030. Epub 2008 Dec 4. PMID: 19059241.
81. Périer RC, Praz V, Junier T, Bonnard C, Bucher P. The eukaryotic promoter database (EPD). Nucleic Acids Res. 2000 Jan 1;28(1):302-3. doi: 10.1093/nar/28.1.302. PMID: 10592254; PMCID: PMC102462.
82. Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. Nucleic Acids Res. 2001 Sep 15;29(18):3705-27. doi: 10.1093/nar/29.18.3705. PMID: 11557805; PMCID: PMC55916.
83. R Foundation for Statistical Computing. (n.d.). R: A language and environment for statistical computing. Vienna, Austria.
84. Razzoli M, Bartolomucci A. The Dichotomous Effect of Chronic Stress on Obesity. Trends Endocrinol Metab. 2016 Jul;27(7):504-515. doi: 10.1016/j.tem.2016.04.007. Epub 2016 May 5. PMID: 27162125; PMCID: PMC4912918.
85. Ricquier D. Uncoupling protein 1 of brown adipocytes, the only uncoupler: a historical perspective. Front Endocrinol (Lausanne). 2011 Dec 28;2:85. doi: 10.3389/fendo.2011.00085. PMID: 22649389; PMCID: PMC3355862.
86. Rim JS, Kozak LP. Regulatory motifs for CREB-binding protein and Nfe2l2 transcription factors in the upstream enhancer of the mitochondrial uncoupling protein 1 gene. J Biol Chem. 2002 Sep 13;277(37):34589-600. doi: 10.1074/jbc.M108866200. Epub 2002 Jun 25. PMID: 12084707.
87. Rosen ED, Walkey CJ, Puigserver P, Spiegelman BM. Transcriptional regulation of adipogenesis. Genes Dev. 2000 Jun 1;14(11):1293-307. PMID: 10837022.
88. Rousset S, Alves-Guerra MC, Mozo J, Miroux B, Cassard-Doulcier AM, Bouillaud F, Ricquier D. The biology of mitochondrial uncoupling proteins. Diabetes. 2004 Feb;53 Suppl 1:S130-5. doi: 10.2337/diabetes.53.2007.s130. PMID: 14749278.
89. Saito S, Saito CT, Shingai R. Adaptive evolution of the uncoupling protein 1 gene contributed to the acquisition of novel nonshivering thermogenesis in ancestral eutherian mammals. *Gene*. (2008) 408:37–44. 10.1016/j.gene.2007.10.018.
90. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D91-4. doi: 10.1093/nar/gkh012. PMID: 14681366; PMCID: PMC308747.

91. Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. Bioinformatics, 27(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706
92. Scarpato M, Angelini C, Cocca E, Pallotta MM, Morescalchi MA, Capriglione T. Short interspersed DNA elements and miRNAs: a novel hidden gene regulation layer in zebrafish? Chromosome Res. 2015 Sep;23(3):533-44. doi: 10.1007/s10577-015-9484-6. PMID: 26363800.
93. Seale P., Kajimura S., Yang W., et al. Transcriptional control of brown fat determination by PRDM16. *Cell Metabolism.* 2007;6(1):38–54. doi: 10.1016/j.cmet.2007.06.001.
94. Shapira SN, Seale P. Transcriptional Control of Brown and Beige Fat Development and Function. Obesity (Silver Spring). 2019 Jan;27(1):13-21. doi: 10.1002/oby.22334. PMID: 30569639; PMCID: PMC6309799.
95. Schrauwen P, Hinderling V, Hesselink MK, Schaart G, Kornips E, Saris WH, Westerterp-Plantenga M, Langhans W. Etomoxir-induced increase in UCP3 supports a role of uncoupling protein 3 as a mitochondrial fatty acid anion exporter. FASEB J. 2002 Oct;16(12):1688-90. doi: 10.1096/fj.02-0275fje. Epub 2002 Aug 21. PMID: 12206997.
96. Seki T, Yang Y, Sun X, Lim S, Xie S, Guo Z, Xiong W, Kuroda M, Sakaue H, Hosaka K, Jing X, Yoshihara M, Qu L, Li X, Chen Y, Cao Y. Brown-fat-mediated tumour suppression by cold-altered global metabolism. Nature. 2022 Aug;608(7922):421-428. doi: 10.1038/s41586-022-05030-3. Epub 2022 Aug 3. PMID: 35922508; PMCID: PMC9365697.
97. Singh S, Rajput YS, Barui AK, Sharma R, Datta TK. Fat accumulation in differentiated brown adipocytes is linked with expression of Hox genes. Gene Expr Patterns. 2016 Mar;20(2):99-105. doi: 10.1016/j.gep.2016.01.002. Epub 2016 Jan 26. PMID: 26820751.
98. Singh S, Rajput YS, Barui AK, Sharma R, Grover S. Expression of developmental genes in brown fat cells grown in vitro is linked with lipid accumulation. In Vitro Cell Dev Biol Anim. 2015 Nov;51(10):1003-11. doi: 10.1007/s11626-015-9930-y. Epub 2015 Jul 7. PMID: 26148884.
99. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell. 2011 Dec 9;147(6):1270-82. doi: 10.1016/j.cell.2011.10.053. PMID: 22153072; PMCID: PMC3319069.
100. Su T, Liu P, Ti X, Wu S, Xue X, Wang Z, Dioum E, Zhang Q. HIF1α, EGR1 and SP1 co-regulate the erythropoietin receptor expression under hypoxia: an essential role in the growth of non-small cell lung cancer cells. Cell Commun Signal. 2019 Nov 21;17(1):152. doi: 10.1186/s12964-019-0458-8. PMID: 31752873; PMCID: PMC6869211.
101. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007 Nov 30;131(5):861-72. doi: 10.1016/j.cell.2007.11.019. PMID: 18035408.
102. Tattersall GJ, Milsom WK. Hypoxia reduces the hypothalamic thermogenic threshold and thermosensitivity. J Physiol. 2009 Nov 1;587(Pt 21):5259-74. doi: 10.1113/jphysiol.2009.175828. Epub 2009 Sep 21. PMID: 19770191; PMCID: PMC2790263.

103. Tayanloo-Beik A, Nikkhah A, Alaei S, Goodarzi P, Rezaei-Tavirani M, Mafi AR, Larijani B, Shouroki FF, Arjmand B. Brown adipose tissue and alzheimer's disease. Metab Brain Dis. 2023 Jan;38(1):91-107. doi: 10.1007/s11011-022-01097-z. Epub 2022 Nov 2. PMID: 36322277.

104. Thompson PJ, Macfarlan TS, Lorincz MC. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell. 2016 Jun 2;62(5):766-76. doi: 10.1016/j.molcel.2016.03.029. PMID: 27259207; PMCID: PMC4910160.

105. Timmons J. A., Wennmalm K., Larsson O., et al. Myogenic gene expression signature establishes that brown and white adipocytes originate from distinct cell lineages. *Proceedings of the National Academy of Sciences of the United States of America.* 2007;104(11):4401–4406. doi: 10.1073/pnas.0610615104.

106. Tóth BB, Arianti R, Shaw A, Vámos A, Veréb Z, Póliska S, Győry F, Bacso Z, Fésüs L, Kristóf E. FTO Intronic SNP Strongly Influences Human Neck Adipocyte Browning Determined by Tissue and PPARγ Specific Regulation: A Transcriptome Analysis. Cells. 2020 Apr 16;9(4):987. doi: 10.3390/cells9040987. PMID: 32316277; PMCID: PMC7227023.

107. B Tóth B, Barta Z, Barta ÁB, Fésüs L. Regulatory modules of human thermogenic adipocytes: functional genomics of large cohort and Meta-analysis derived marker-genes. BMC Genomics. 2021 Dec 11;22(1):886. doi: 10.1186/s12864-021-08126-8. PMID: 34895148; PMCID: PMC8665548.

108. Vaishnav ED, de Boer CG, Molinet J, Yassour M, Fan L, Adiconis X, Thompson DA, Levin JZ, Cubillos FA, Regev A. The evolution, evolvability and engineering of gene regulatory DNA. Nature. 2022 Mar;603(7901):455-463. doi: 10.1038/s41586-022-04506-6. Epub 2022 Mar 9. PMID: 35264797; PMCID: PMC8934302.

109. Villarroya F, Peyrou M, Giralt M. Transcriptional regulation of the uncoupling protein-1 gene. Biochimie. 2017 Mar;134:86-92. doi: 10.1016/j.biochi.2016.09.017. Epub 2016 Oct 5. PMID: 27693079.

110. Waldén TB, Hansen IR, Timmons JA, Cannon B, Nedergaard J. Recruited vs. nonrecruited molecular signatures of brown, "brite," and white adipose tissues. Am J Physiol Endocrinol Metab. 2012 Jan 1;302(1):E19-31. doi: 10.1152/ajpendo.00249.2011. Epub 2011 Aug 9. PMID: 21828341.

111. Warfel JD, Elks CM, Bayless DS, Vandanmagsar B, Stone AC, Velasquez SE, Olivares-Nazar P, Noland RC, Ghosh S, Zhang J, Mynatt RL. Rats lacking Ucp1 present a novel translational tool for the investigation of thermogenic adaptation during cold challenge. Acta Physiol (Oxf). 2023 May;238(1):e13935. doi: 10.1111/apha.13935. Epub 2023 Feb 7. PMID: 36650072.

112. Weirauch MT, Cote A, Norel R, Annala M, Zhao Y, Riley TR, Saez-Rodriguez J, Cokelaer T, Vedenko A, Talukder S; DREAM5 Consortium; Bussemaker HJ, Morris QD, Bulyk ML, Stolovitzky G, Hughes TR. Evaluation of methods for modeling transcription factor sequence specificity. Nat Biotechnol. 2013 Feb;31(2):126-34. doi: 10.1038/nbt.2486. Epub 2013 Jan 27. PMID: 23354101; PMCID: PMC3687085.

113. Wibmer AG, Becher T, Eljalby M, Crane A, Andrieu PC, Jiang CS, Vaughan R, Schöder H, Cohen P. Brown adipose tissue is associated with healthier body fat distribution and metabolic benefits independent of regional adiposity. Cell Rep Med. 2021 Jul 7;2(7):100332. doi: 10.1016/j.xcrm.2021.100332. PMID: 34337558; PMCID: PMC8324464.

114. Wunderlich Z, Mirny LA. Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. 2009 Oct;25(10):434-40. doi: 10.1016/j.tig.2009.08.003. Epub 2009 Oct 6. PMID: 19815308; PMCID: PMC3697852.

115. Xiao W, Geng W, Zhou M, Xu J, Wang S, Huang Q, Sun Y, Li Y, Yang G, Jin Y. POU6F1 cooperates with RORA to suppress the proliferation of lung adenocarcinoma by downregulation HIF1A signaling pathway. Cell Death Dis. 2022 May 3;13(5):427. doi: 10.1038/s41419-022-04857-y. PMID: 35504868; PMCID: PMC9065044.

116. Xu L, Xu C, Zhou S, Liu X, Wang J, Liu X, Qian S, Xin Y, Gao Y, Zhu Y, Tang X. PAX4 promotes PDX1-induced differentiation of mesenchymal stem cells into insulin-secreting cells. Am J Transl Res. 2017 Mar 15;9(3):874-886. PMID: 28386318; PMCID: PMC5375983.

117. Xue B, Coulter A, Rim JS, Koza RA, Kozak LP. Transcriptional synergy and the regulation of Ucp1 during brown adipocyte induction in white fat depots. Mol Cell Biol. 2005 Sep;25(18):8311-22. doi: 10.1128/MCB.25.18.8311-8322.2005. PMID: 16135818; PMCID: PMC1234324.

118. Yevshin I, Sharipov R, Valeev T, Kel A, Kolpakov F. GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Res. 2017 Jan 4;45(D1):D61-D67. doi: 10.1093/nar/gkw951. Epub 2016 Oct 24. PMID: 27924024; PMCID: PMC5210645.

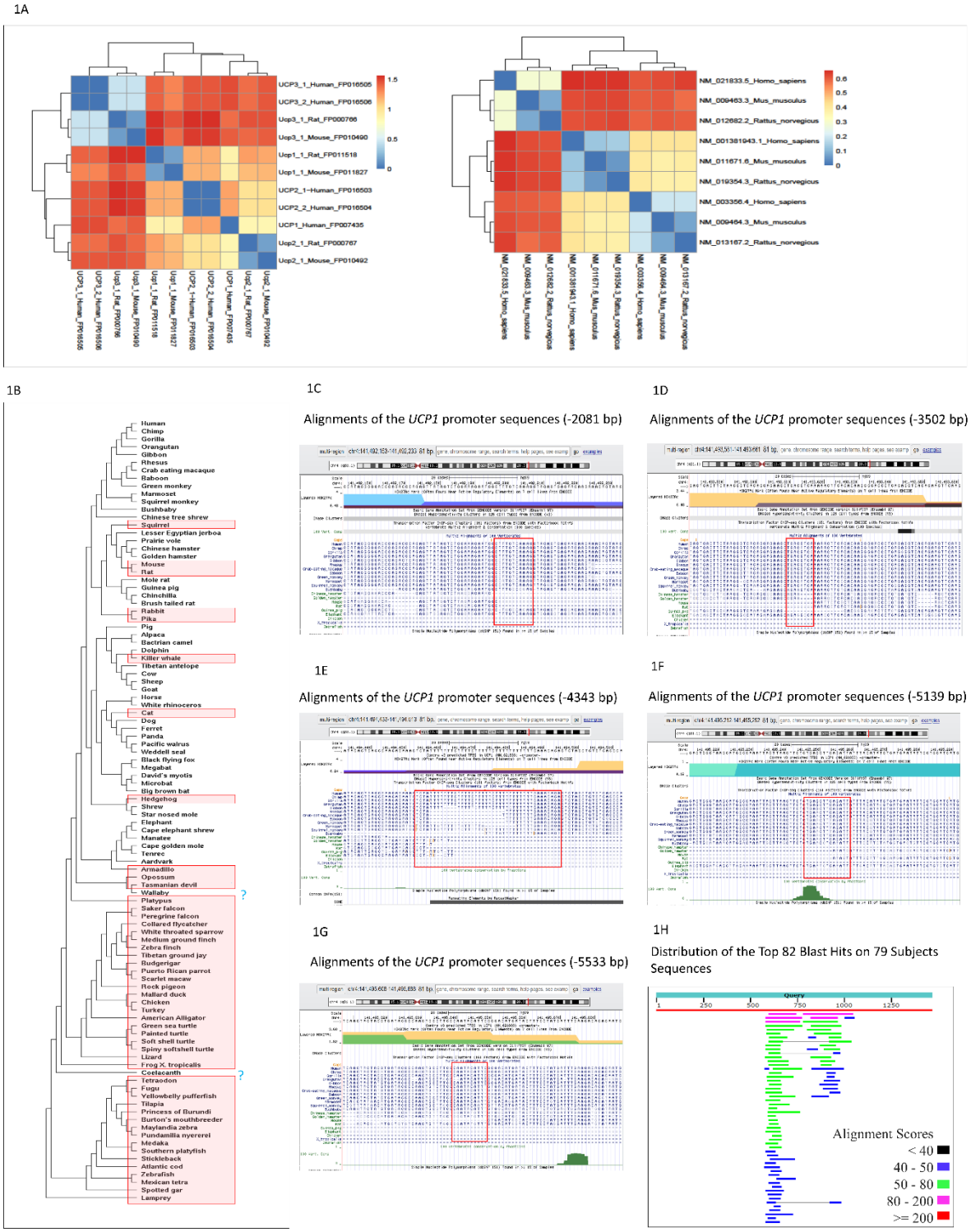
119. Zhang X, Wang C, Liu M; 77-OR: Targeting JunB Induces a Shift from Low- to High-Thermogenic Adipocytes and Ameliorates Diet-Induced Insulin Resistance. *Diabetes* 20 June 2023; 72 (Supplement_1): 77–OR. doi: 10.2337/db23-77-OR

120. Zhang C, Weng Y, Shi F, Jin W. The Engrailed-1 Gene Stimulates Brown Adipogenesis. Stem Cells Int. 2016;2016:7369491. doi: 10.1155/2016/7369491. Epub 2016 Apr 11. PMID: 27148369; PMCID: PMC4842372.

121. Zhang J, Zhang Y, Sun T, Guo F, Huang S, Chandalia M, Abate N, Fan D, Xin HB, Chen YE, Fu M. Dietary obesity-induced Egr-1 in adipocytes facilitates energy storage via suppression of FOXC2. Sci Rep. 2013;3:1476. doi: 10.1038/srep01476. PMID: 23502673; PMCID: PMC3600596.

SUPPLEMENTARY FIGURES

Supplementary figure 1.



Supplementary Figure 1. Heatmaps show the pairwise genetic distances of the nucleotid sequences of the *UCP1*,

1486 *UCP2* and *UCP3* genes' regulatory and coding regions in humans, mice and rats (pheatmap program in R
1487 environment). **1B** The phylogenetic tree shows the presence and the absence (red box) of the human UCP1
1488 promoter DNA fragment spans between -2083 bp and -3502 bp (GRCh37/hg19: Chr4 141,492,160-141,493,617)
1489 in *UCP1* promoter of different vertebrate species according to the Multiz Alignment of 100 Vertebrates database
1490 in the UCSC genome browser. No information is available for the three genera marked with blue question marks.
1491 **1C – 1G** Alignments of the *UCP1* promoter sequences of different species highlight the palindromic nucleotide
1492 sequences of the beginning and end of the DNA fragments (1419 bp and 658 bp) missing from the mouse and rat
1493 *UCP1* promoter according to the Multiz Alignment of 100 Vertebrates database in the UCSC genome browser.
1494 **1H** Graphical summary showing the human genomic occurrence of the *UCP1* promoter DNA fragment (1419 bp)
1495 absent from the mice and rat *UCP1* promoter (https://blast.ncbi.nlm.nih.gov/). The graphic is an overview of the
1496 database sequences aligned to the query sequence. These are represented by horizontal bars colored coded by
1497 score and showing the extent of the alignment on the query sequence. Separate aligned regions on the same
1498 database sequence are connected by a thin grey line.

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

1509

1510

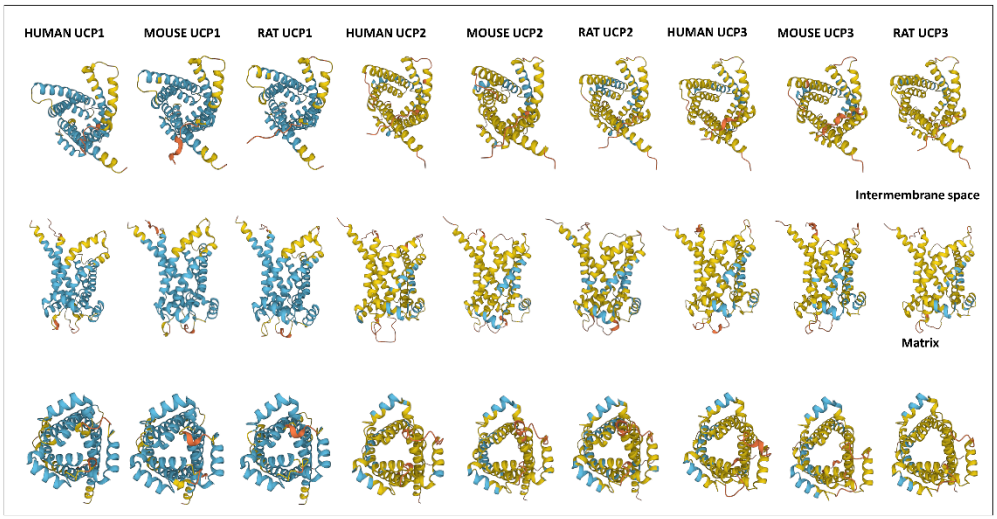
1511

1512

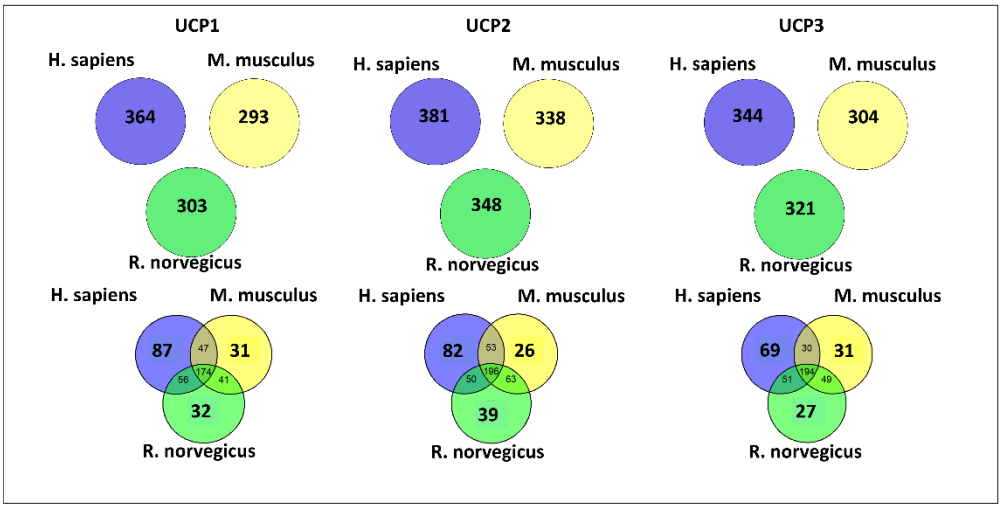
1513

1514 **Supplementary Figure 2.**

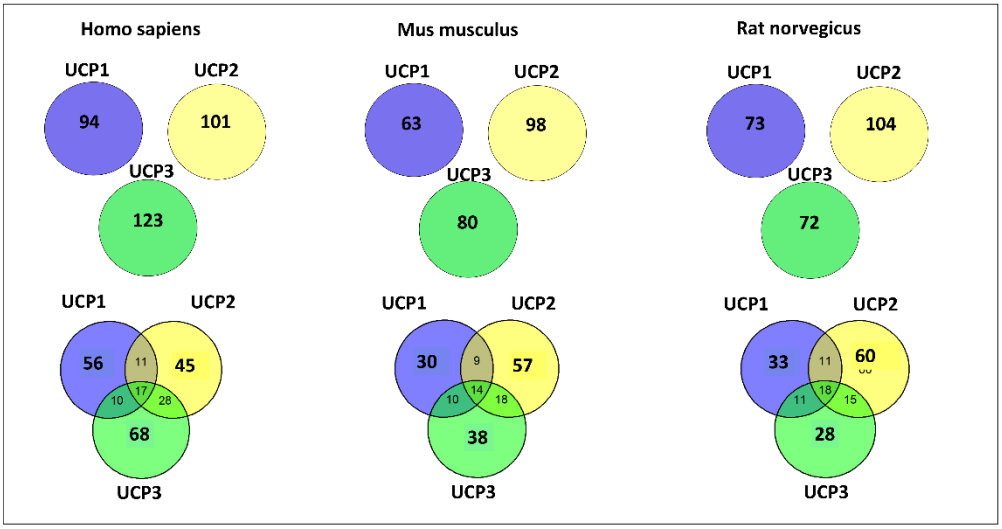
2A



2B



2C

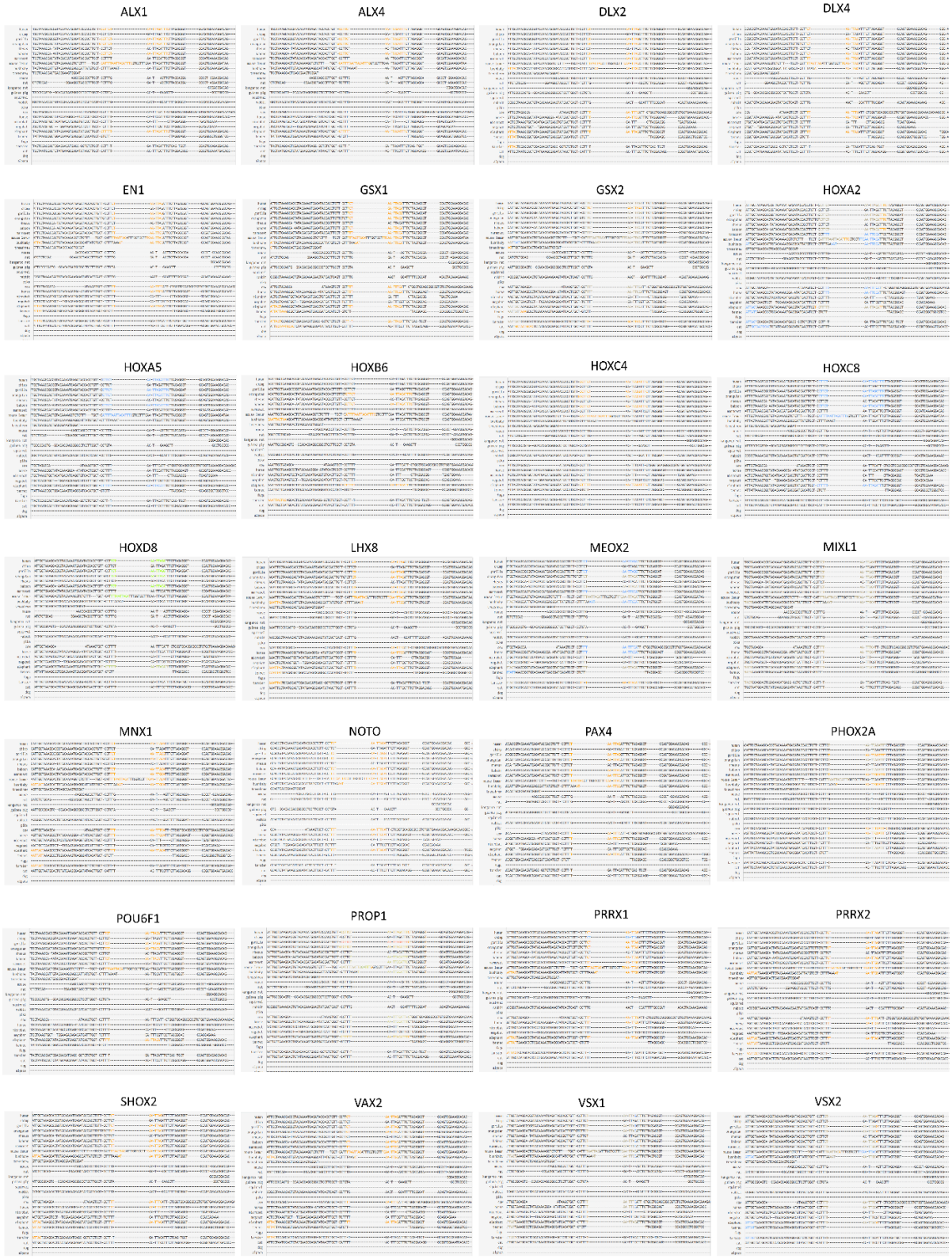


1515

1516 **Supplementary Figure 2. 2A** The spatial structure of human, mouse and rat *UCP1*, *UCP2* and *UCP3* proteins
1517 predicted by AlphaFold. The figures show the predicted spatial structure of the protein from the inter-membrane
1518 face, the membrane-embedded side and the matrix face. The blue color indicates that for the *UCP1* gene, the
1519 model predicts the spatial structure with high confidence for most of the part of the protein for all three species,
1520 while for the other UCPs, the model confidence value is generally low. Typically, a short highly disordered portion
1521 facing the matrix is well conserved in the UCP proteins of all three species. This potentially context-dependent
1522 variable spatial structure may affect the function of the protein at a given time. **2B** Venn diagram showing the
1523 number of TF REs identified in the promoter (-5000 bp –TSS – +1000 bp) of human, mouse and rat *UCP1*, *UCP2*
1524 and *UCP3* genes using the EPD interface, with probability level p<1e-4. Merged Venn diagrams among species
1525 show the number of unique and common TF REs in *UCP1*, *UCP2* and *UCP3* genes. **2C** Venn diagram showing the
1526 number of TF REs identified in the promoter (-5000 bp –TSS – +1000 bp) of human, mouse and rat *UCP1*, *UCP2*
1527 and *UCP3* genes using the EPD interface, with probability level p<1e-5. Merged Venn diagrams among *UCP* genes
1528 show the number of unique and common TF REs in human, mouse and rat *UCP* genes (Venn program in R
1529 environment).

1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549

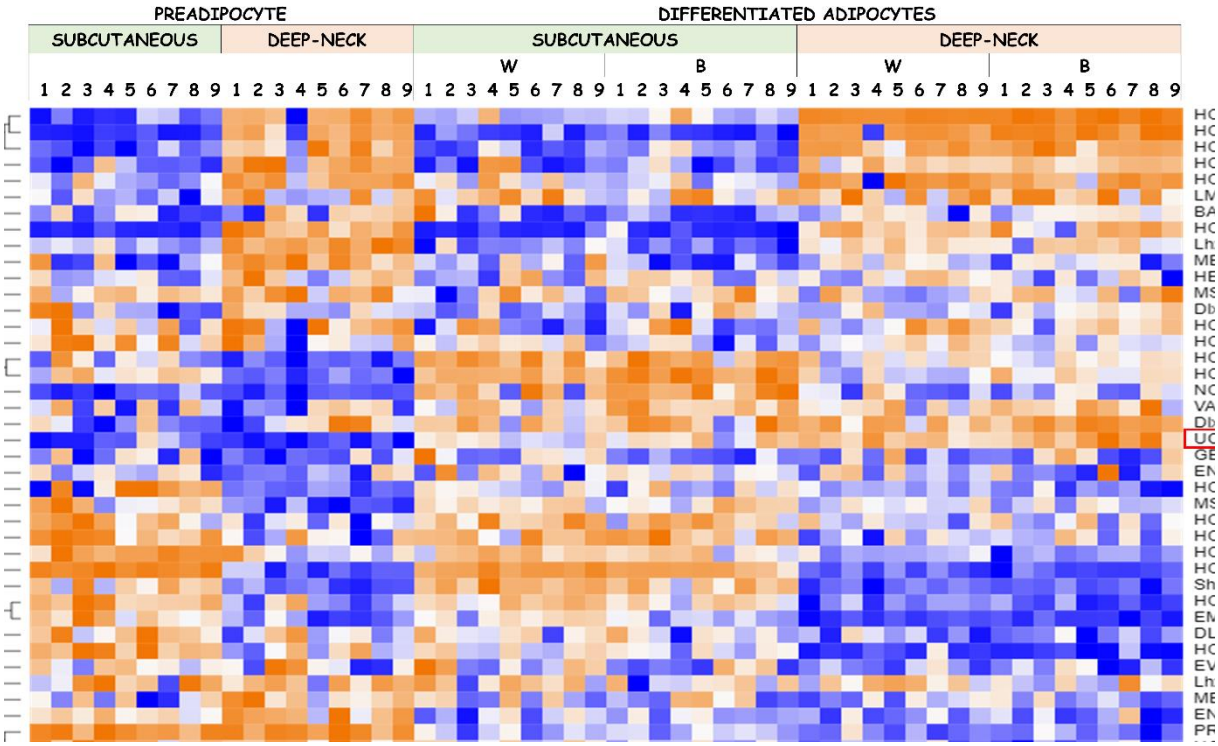
1550 **Supplementary Figure 3.**



1551

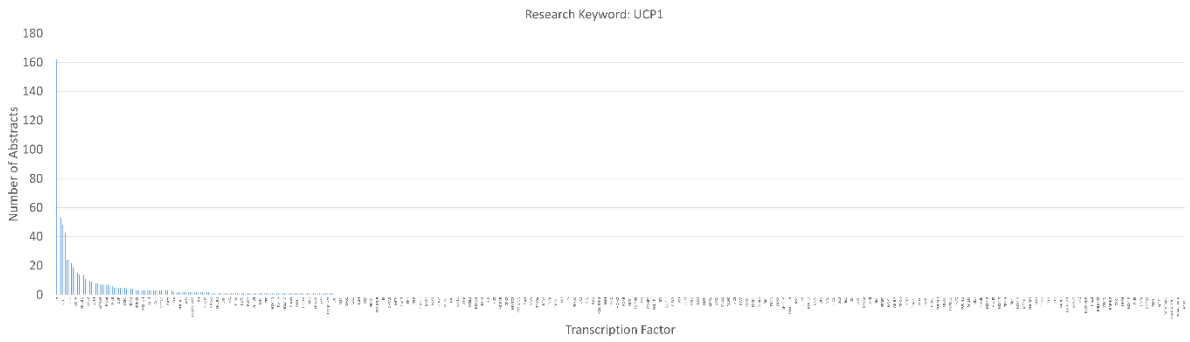
1552 **Supplementary Figure 3.** In the aligned *UCP1* promoter sequence of different species at position -690 bp, we
1553 investigated the presence of TF binding sequences on the ConTra v3 platform, which gave positive search results
1554 on the EPD interface, as well as those that were not tested using EPD but have a potential RE based on their TF
1555 class classification. Stringency: core = 0.95, similarity matrix = 0.85.

Supplementary Figure 4.



Supplementary Figure 4. The heatmap depicts the relative gene expression profile of homeodomain Factor 1 class TFs (based on the JASPAR database) in adipocytes of cervical origin. For comparison, *UCP1* is also shown (Morpheus interface). Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies TFs with gene expression profiles similar to *UCP1*. Adipocytes were differentiated using a white (W) or brown (B) differentiation protocol.

1576 **Supplementary Figure 5.**



1577

1578 **Supplementary Figure 5.** The bar graphs show the number of abstracts deposited in the PubMed database in
1579 which the TF under study appears together with the keyword UCP1, highlighting the "longtail" distribution of the
1580 data. On the x-axis only a few representative TFs are named, the y-axis shows the number of abstracts (Excel).

1581

1582

1583

1584

1585

1586

1587

1588

1589

1590

1591

1592

1593

1594

1595

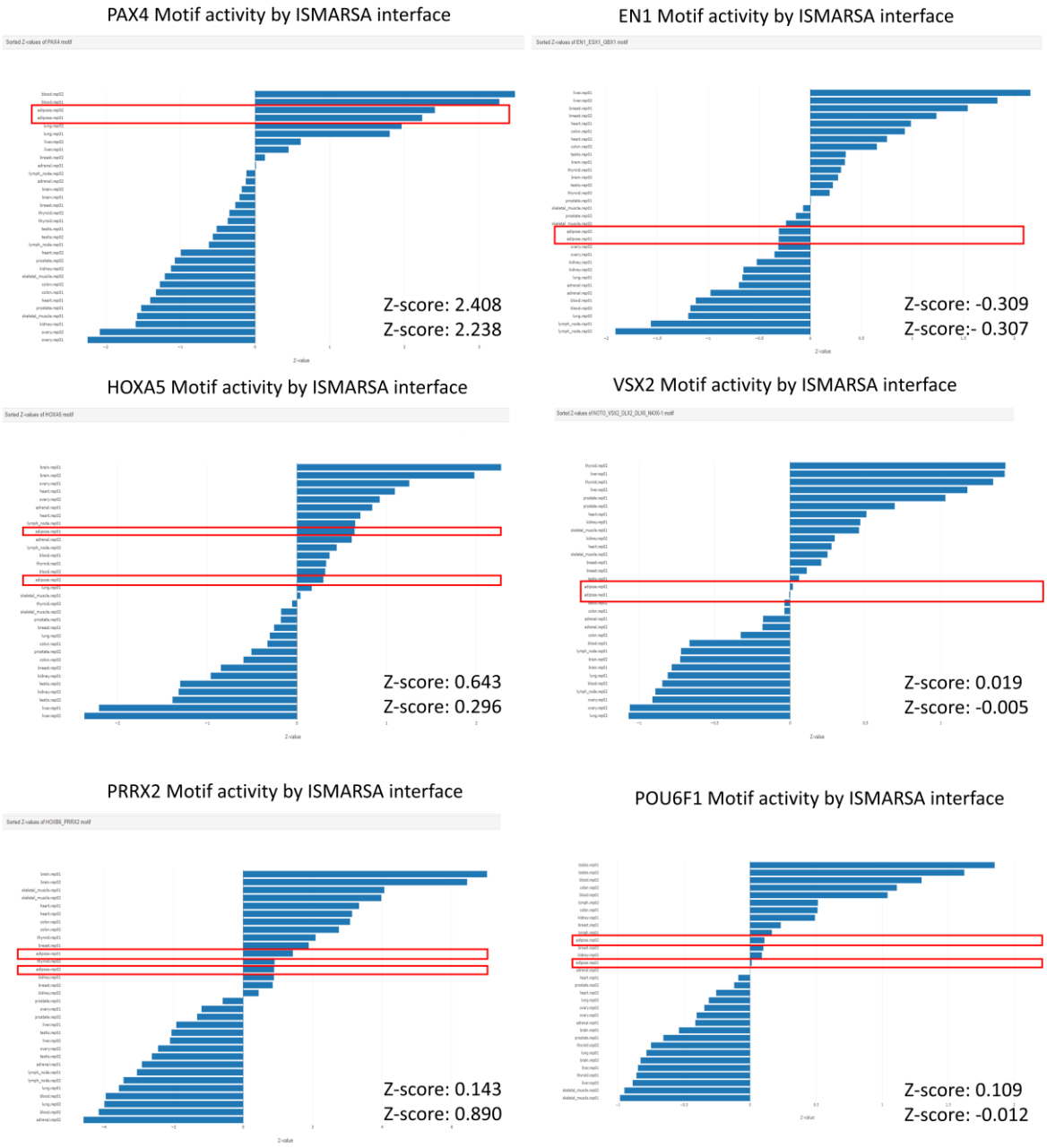
1596

1597

1598

1599

1600 **Supplementary Figure 6.**



1601

1602 **Supplementary Figure 6.** Bar graphs show the relative activity (z-score) of the putative species-specific and
1603 species-common homeodomain family TFs under study assessed per human tissue (according to ISMARA
1604 interface). The red box highlights the result recorded in adipose tissue samples.

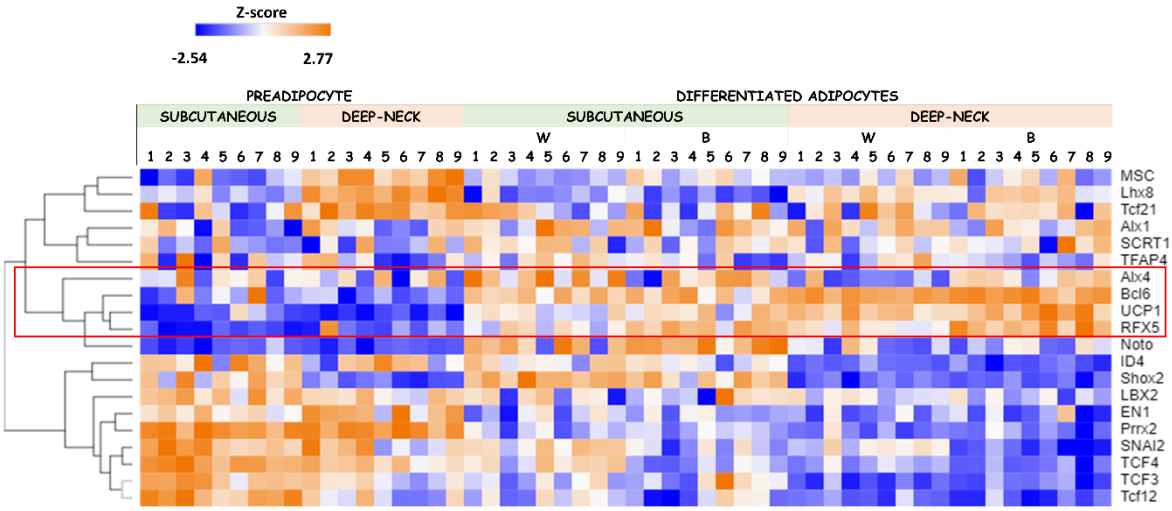
1605

1606

1607

1608

1609 **Supplementary Figure 7.**



1610

1611

1612 **Supplementary Figure 7.** The heatmap shows the relative gene expression profiles of TFs, for which a binding
1613 sequence was identified at -690 bp and near +/- 40 bp in the promoter of human *UCP1* using EPD interface (p<1e-
1614 4). For comparison, *UCP1* is also shown. Hierarchical cluster analysis based on Spearman Rank similarity matrix
1615 identifies TFs with gene expression profiles similar to *UCP1* (red box) (Morpheus interface). Adipocytes were
1616 differentiated using a white (W) or brown (B) differentiation protocol.

1617

1618

1619

1620

1621

1622

1623

1624

1625

1626

1627

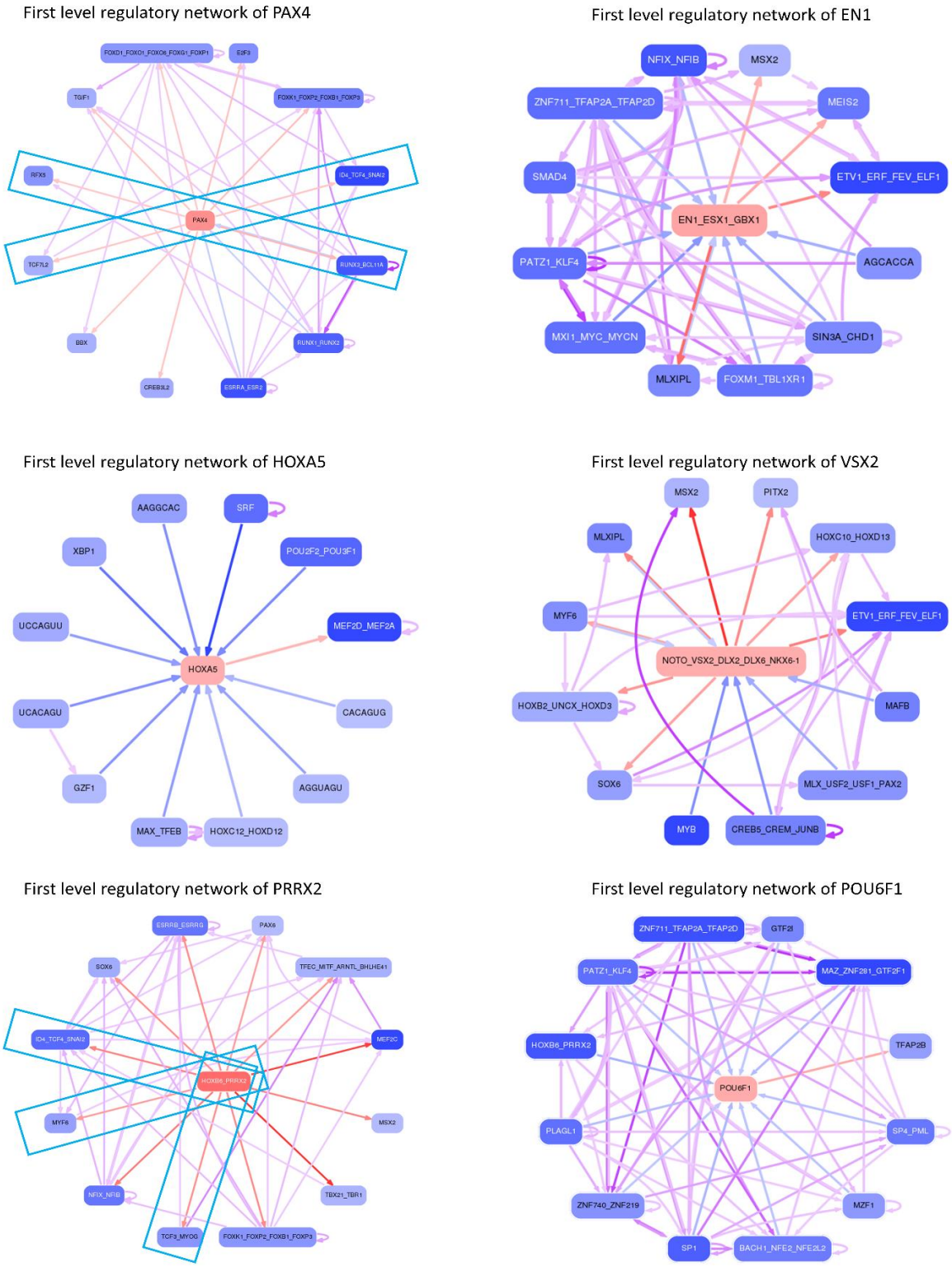
1628

1629

1630

1631

1632 **Supplementary Figure 8.**



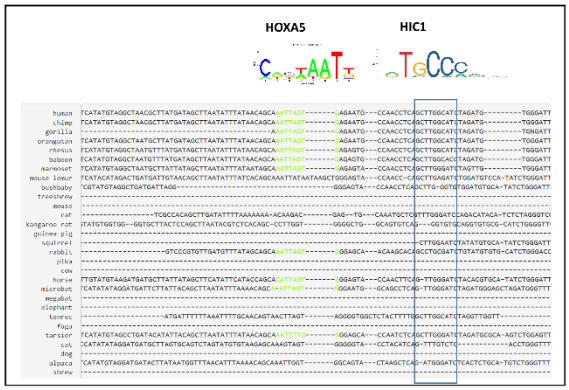
1633

1634 **Supplementary Figure 8.** Interactomes show the first level of the regulatory network of the putative species-
1635 specific and species-common homeodomain family TFs under study, such as PAX4, EN1, HOXA5, DLX2, PRRX2
1636 and POU6F1 according to the ISMARA interface. The blue box highlights the partners that were recognized as
1637 having a RE close to (within +/- 40 bp of) the RE of the TF of interest.

Supplementary Figure 9.

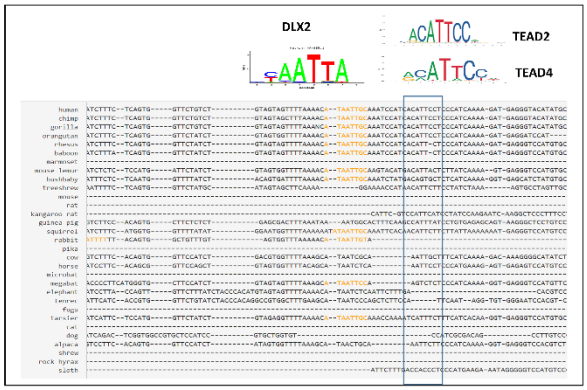
9A

HOXA5 RE at -4785 bp and HIC1 -4797 bp in human *UCP1* promoter



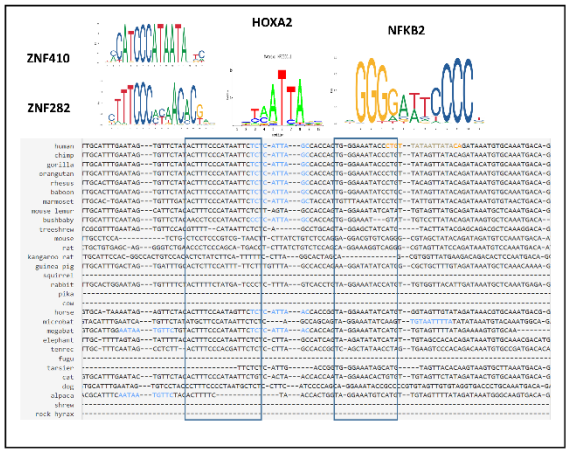
9B

DLX2 RE at -2478 bp and TEAD4 at -2496 bp in human *UCP1* promoter



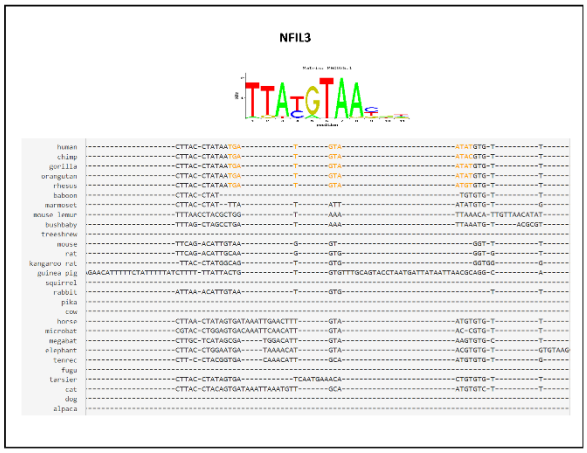
9C

HOXA2 RE at -3959 bp, NFKB2 RE at -3978 bp, ZNF410 and ZNF282 RE at -3948 bp in human *UCP1* promoter



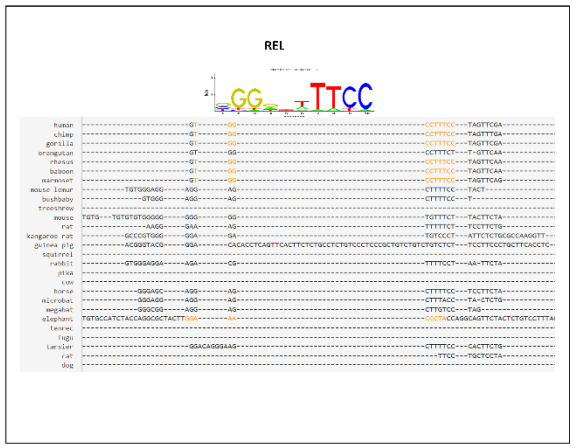
9D

NFIL3 RE at -881 bp in human *UCP1* promoter



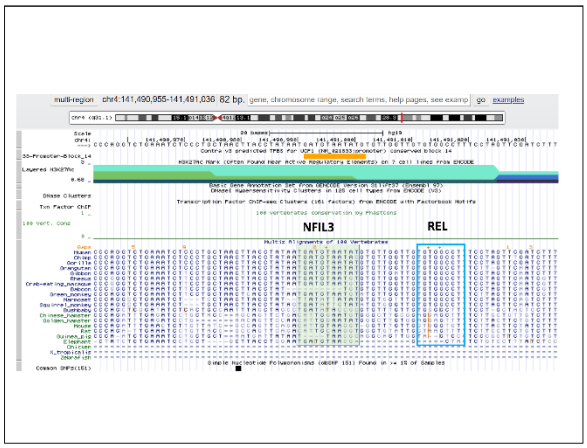
9E

REL RE at -902 bp in human *UCP1* promoter



9F

NFIL3 RE at -881 bp and REL RE at -904 bp in human *UCP1* promoter



Supplementary Figure 9. Alignment of *UCP1* promoter sequences from different species at different positions on the ConTra v3 platform. Human *UCP1*-specific TF REs are highlighted in different colors, and nearby TF REs are outlined in a blue box. **9A** shows the presence of the HOXA5 TF RE, unique to the human *UCP1* promoter at -4785 bp, and the nearby HIC1 RE at -4797 (EPD interface data). **9B** shows the presence of the DLX2 TF RE, unique

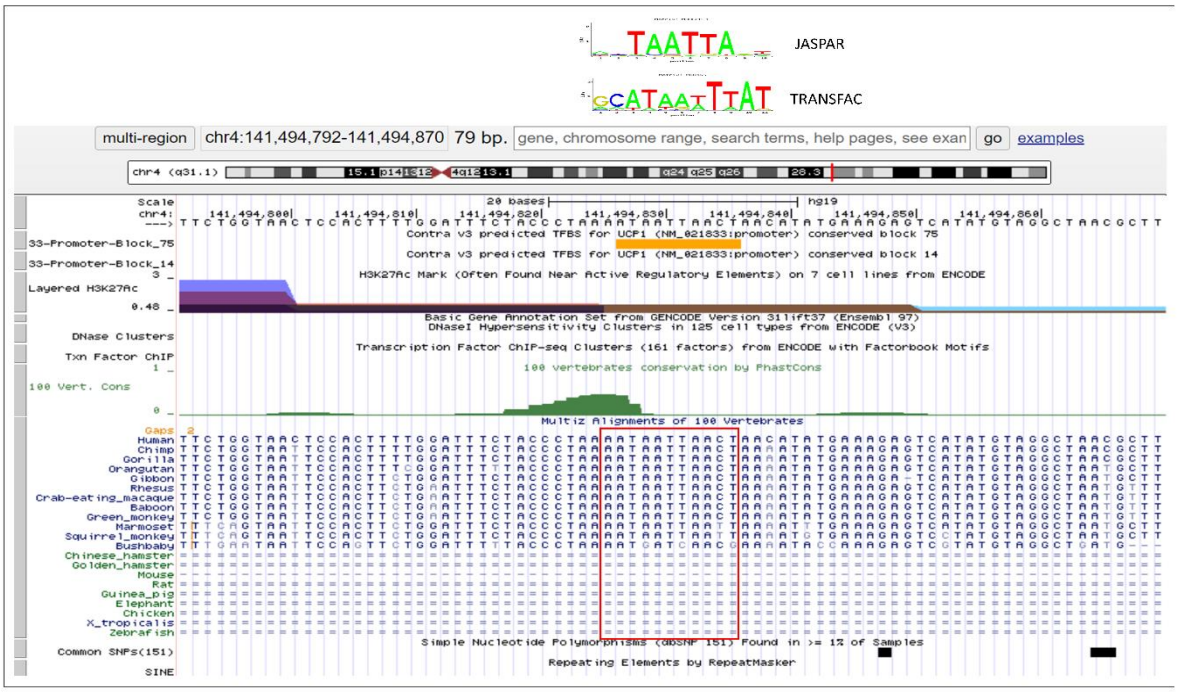
1644 to the human *UCP1* promoter at -2478 bp, and the nearby TEAD1-4 RE at -2496 (EPD interface data). **9C** shows
1645 the presence of the HOXA2 TF RE, unique to the human *UCP1* promoter at -3959 bp, and the nearby NFKB RE at
1646 -3978, and ZNF410 and ZNF282 at - 3948 bp (EPD interface data). **9D** shows the presence of the NFIL3 TF RE,
1647 unique to the human *UCP1* promoter at -881 bp, and **9E** the nearby REL RE at -902 bp (EPD interface data). **9F**
1648 shows the presence of the NFIL3 TF RE at -881 bp and the nearby REL RE at -902 bp across species in UCSC
1649 genome browser (ConTra v3 and EPD interface data).

1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675

Supplementary Figure 10.

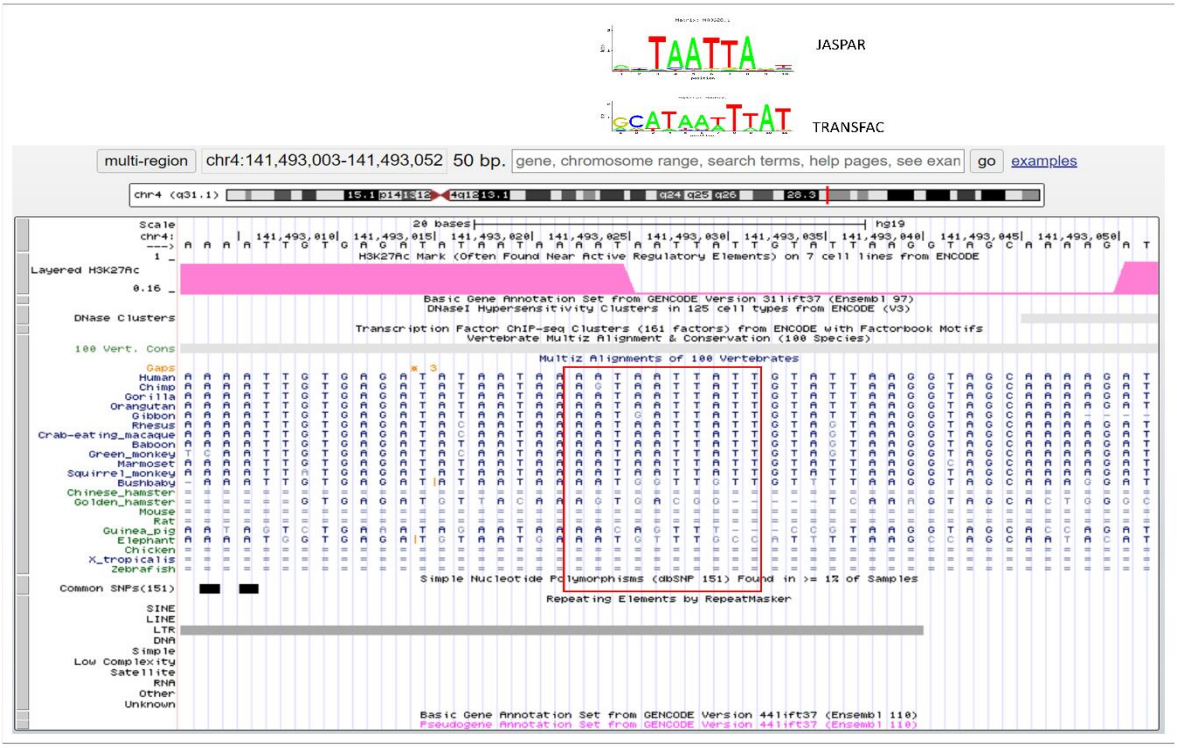
10A

POU6F1 RE at -4715 bp in human UCP1 promoter



10B

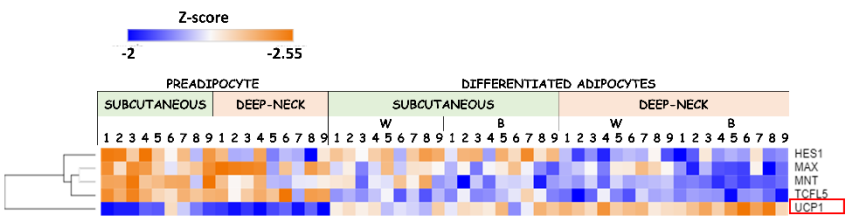
POU6F1 RE at -2913 bp in human UCP1 promoter



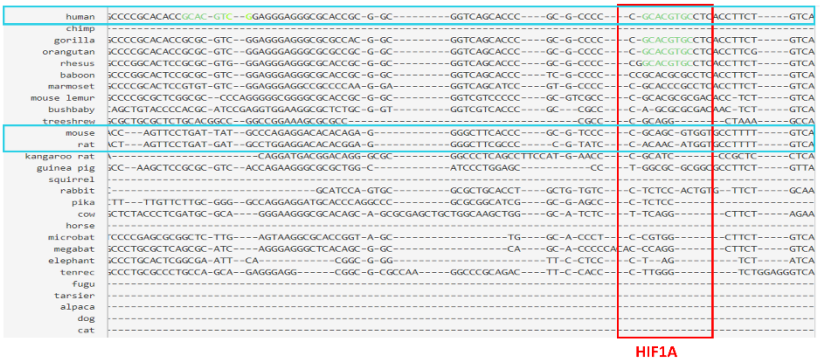
Supplementary Figure 10. Alignment of the *UCP1* promoter sequences of different species **10A** at -4715 bp position and **10B** at -2913 bp position to investigate the presence of POU6F1 TF-binding sequences in human, mouse and rat promoter according to the Multiz Alignment of 100 Vertebrates database in the UCSC genome browser.

Supplementary Figure 11.

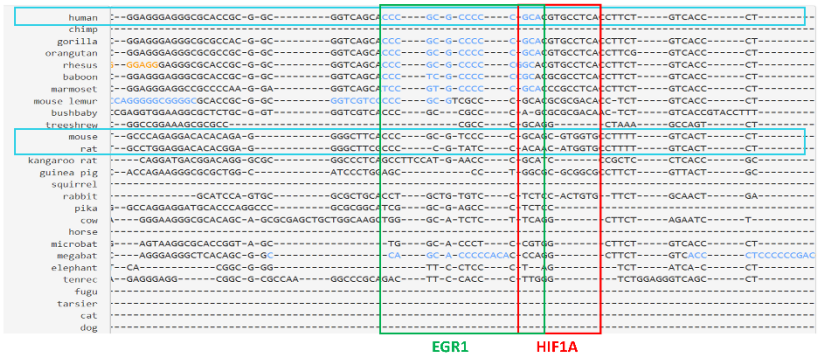
11A



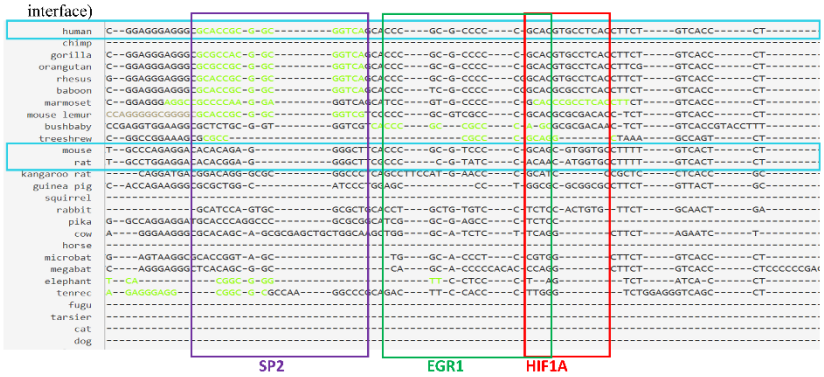
11B Phylogenetic footprinting: The HIF1A binding site in the human UCP1 promoter (ConTra 3v interface)



11C Phylogenetic footprinting: The EGR1 and HIF1A binding site in the human UCP1 promoter (ConTra 3v interface)



11D Phylogenetic footprinting: The SP2 EGR1 and HIF1A binding site in the human UCP1 promoter (ConTra 3v interface)



Supplementary Figure 11. **11A** Heatmap showing the relative gene expression profile of common TFs identified in the promoters of human *UCP1*, *UCP2* and *UCP3* around -264 bp missing the binding sequence in the promoters

1686 of mouse and rat orthologous genes using EPD interface (p<1e-4) in cervical derived adipocyte samples. For
1687 comparison, *UCP1* is also shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix
1688 identifies TFs with similar gene expression profiles to *UCP1*. (Morpheus Interface) **11B** Phylogenetic footprinting:
1689 the green color/red box highlights the HIF1A binding site in the *UCP1* promoter at -264 (ConTra v3 interface).
1690 **11C** Phylogenetic footprinting: Blue color/green box highlights the EGR1 RE at -257 and and red box the HIF1A
1691 RE at -264 bp in the *UCP1* promoter (ConTra v3 interface). **11D** Phylogenetic footprinting: the green color/purple
1692 box highlights the SP2 RE at -223 bp, the green box shows the EGR1 RE at -257 bp and the red box the HIF1A RE
1693 at -264 bp in the *UCP1* promoter (ConTra v3 interface). Adipocytes were differentiated using a white (W) or
1694 brown (B) differentiation protocol.