# Deciphering Species-Specific and Conserved Regulation of Human *UCP1* Gene Expression: Insights from Predicted Homeodomain Family Transcription Factor Binding Sites

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

## Highlights

- Evolutionary dynamics of UCP1 genes and their regulatory elements differ between humans and rodents.
- Majority of the putative regulatory factors of human *UCP1* are general regulators of the UCPs.
- Species-specific regulatory hotspots primarily recognized by TFs of the homeodomain family.
- Repetitive elements may shape specific regulatory mechanisms during evolution.

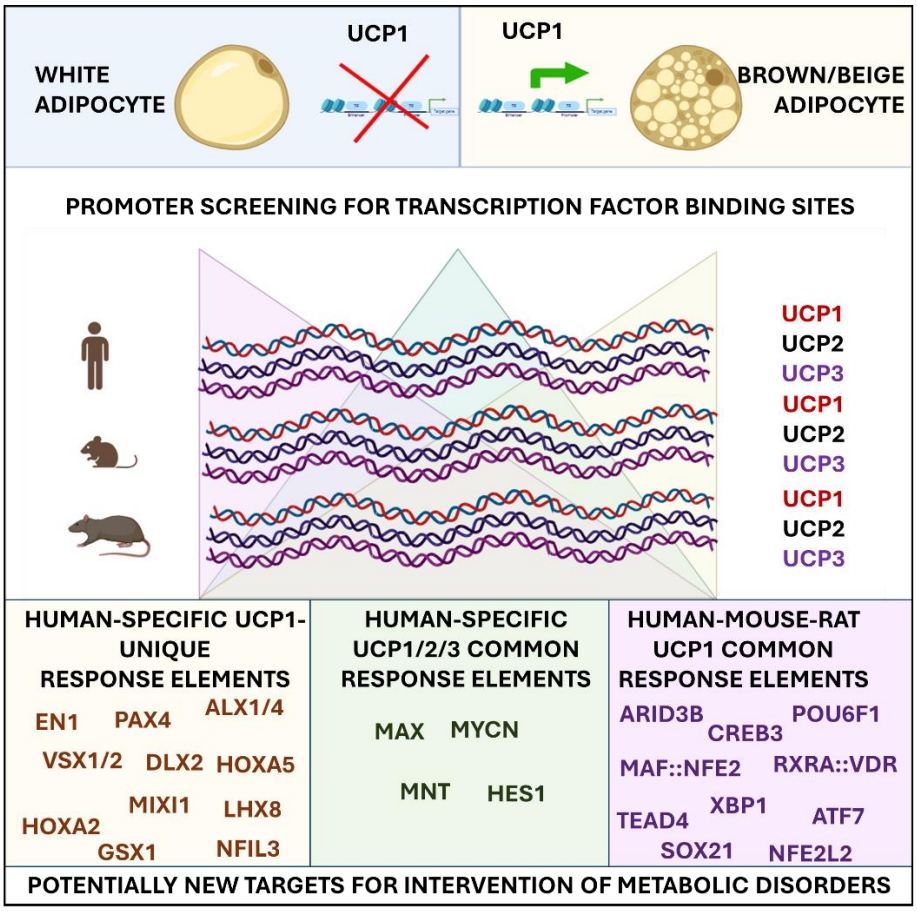
## Abstract

While the potential of adipocyte thermogenesis in influencing metabolic diseases is well-recognized, its precise regulatory mechanisms remain elusive. To address this, we explore the complex network of transcription factors (TFs) governing the human *UCP1* gene, a central player in thermogenesis, to shed light on regulatory processes driving human fat cell phenotypes. By employing computational analyses, we screen *UCP1* ortholog and paralog promoters across humans, mice and rats to uncover conserved and species-specific regulatory elements. Leveraging transcriptomic data, chipSeq results and AdioNET phenotyping we delineate TFs potentially bind to the human *UCP1* promoter. Significant sequence variation is revealed in the enhancer regions of *UCP1s*, characterised by the absence of large segments in the rodent genome, potentially influenced by the presence/absence of repetitive elements. Furthermore, our analyses identify species-specific regulatory hotspots mostly recognized by homeodomain family TFs providing the potential for the formation of enhanceosomes with

nearby TFs. Exploring phylogenetically conserved regulatory elements reveals TFs exhibiting binding sites across species and *UCP* paralogs assigning again homeodomain family TFs among others, and bHLH TFs, respectively. These results potentially advance our understanding of thermogenic fat cell development and offer new targets for intervention in metabolic disorders.

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

**Graphical Abstract**




## I. INTRODUCTION

Adipocytes are pivotal in maintaining energy balance and metabolic health within the human body. While white adipocytes primarily serve as energy reservoirs, brown and beige adipocytes specialize in energy expenditure through thermogenesis. This thermogenic

capacity holds significant promise for influencing overall energy homeostasis, especially in the context of metabolic disorders (Tayanloo-Beik et al., 2023; Seki et al., 2022; Wibmer et al., 2021; Becher et al., 2021; Razzoli et al., 2016, Nedergaard et al., 2007). Non-shivering thermogenesis, a crucial component of the body's heat-producing mechanisms, can be initiated by cold exposure and various environmental cues, yet the intricate mechanisms remain incompletely understood (Hoang et al., 2013). Cold-induced thermogenesis involves the upregulation of Uncoupling Protein 1 (UCP1) within the mitochondrial membrane, leading to controlled heat generation and presenting an attractive target for weight management strategies (Feldman et al., 2009; Ricquier 2011, Nedergaard et al. 2001, Enerbäck 1997). Despite extensive research, a comprehensive understanding of the regulatory mechanisms governing adipocyte browning remains elusive (Waldén et al. 2012; Nedergaard and Cannon 2014). Several TFs have been proposed as direct regulators of UCP1 gene expression including peroxisome proliferator-activated receptors (PPARs), retinoic acid receptors (RXRs, RARs), thyroid hormone receptor (THR), cAMP response element binding proteins (CREBPs), nuclear factor erythroid response element 2 like 2 (NFE2L2), Sterol Regulatory Element-Binding Protein 1 (SREBP1), early B-cell factor 1-2 (EBF1-2), the c-Jun::c-Fos complex, and the CCAAT/enhancer-binding proteins (CEBPA) and indirect regulators like PPARG Coactivator 1 alpha (PGC-1α) and PR domain containing 16 (PRDM16) (Rosen et al. 2000; del Mar Gonzalez-Barroso 2000; Rim and Kozak 2002; Seale et al. 2007; Timmons et al. 2007; Ricquier 2011, Villaroya et al., 2017; Zhang et al. 2023). Recent reviews offering 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), however, the intricate web of regulatory mechanisms is still to some extent obscured.

This investigation aims to utilize computational methodologies to identify transcription factors (TFs) potentially regulating human *UCP1* gene expression by discerning TFs possess putative functional DNA-binding motifs within the promoter regions of the human *UCP1* gene (GRCh37/hg19; Chr. 4: 141,489,115-141,495,114). To simplify nomenclature, the term "promoter" is used here to refer to the regulatory sequence segments under study, including intergenic regions (DNA segments in the positive direction from the transcription start site (TSS)) and distal enhancers. By comparing *UCP1* orthologs and paralogs across humans, mice and rats, we aimed to uncover conserved and species-specific *UCP1* regulatory elements.

While Mice and rats are valuable as model organisms for thermogenic studies, they present challenges in translating findings to human physiology due to inherent differences (Enerbäck et al. 1997). Particularly pertinent to metabolic and thermogenic processes, the evolutionary divergence in strategies for maintaining energy balance based on factors like body size and environmental temperature has been documented (Tattersall and Milsom, 2009; Kottke et al., 1948), indicating potential regulatory differences. Exploring regulatory variances in the expression of paralogous genes with shared origins but distinct tissue expressions and protein functions, such as *UCP2* and *UCP3* (Nedergaard and Canon 2003, Rousset et al. 2004, Nicholls 2021), opens up the possibility of unravelling *UCP1* gene-specific regulators and offering insight into potentially more specific intervention targets that may not be apparent in traditional model organisms (Hughes and Criscuolo 2008, Enerbäck et al. 1997, Warfel et al. 2023). The identification of these elements may elucidate specific regulatory processes driving human fat cell phenotypes and advance our understanding of thermogenic fat cell development, with significant therapeutic implications.

## II. RESULTS

### 1. Comparative Analysis of UCP1, UCP2, and UCP3 Promoter, Gene and Protein Sequences Across Species: Structural and Evolutionary Aspects.

To elucidate the structural properties underpinning the distinctive function of three proteins of the mitochondrial Solute Carrier Family 25 (SLC25), the UCP1 (SLC25A7), UCP2 (SLC25A8) and UCP3 (SLC25A9), we undertook a thorough comparative analysis including the nucleotide sequences of the genes and their promoters in the human, mouse and rat genomes. Additionally, we aligned the amino acid sequences of the encoded proteins and explored their complex three-dimensional (3D) spatial structures.

#### *1.1 Sequence Similarity Analysis Reveals Distinct Evolutionary Relationships Between UCP1, UCP2, and UCP3 Genes and Their Regulatory Regions*

A phylogenetic tree, reconstructed by merging data from human, mouse, and rat *UCP1*, *UCP2*, and *UCP3*, clearly illustrates that orthologous sequences grouped together indicating their closer relationship than to their paralogous counterparts, considering amino acid composition, coding sequences (mRNA), and regulatory DNA sequences with high support

113 values in all cases (Fig. 1A). As anticipated, an elevated similarity is also evident between
114 mouse and rat concerning orthologous genes, as they cluster more closely compare to human.
115 Conversely, for paralogous genes, the reconstruction of mRNA and amino acid sequences
116 clearly identified *UCP2* and *UCP3* as a separate cluster from *UCP1*. In contrast, *UCP1* and *UCP2*
117 were grouped together based on their promoters. These findings are further substantiated by
118 the pairwise distance analysis of coding and promoter sequences of the *UCP* genes (Suppl. Fig.
119 1A).

120 *1.2 Presence of Transposable and Repetitive Elements Result in Significant Sequence*
121 *Divergence within the Human and Rodent Distal Enhancer Regions.*

122 Cross-species alignment of the investigated 6000 bp long nucleotide sequences of the *UCP1*
123 promoter revealed that a 1419 bp sequence interval spanning from -2083 bp to -3502 bp
124 (GRCh37/hg19: Chr4 141,492,198-141,493,617) in the distal enhancer region of the human
125 *UCP1* promoter and the final 658 bp (Chr4: 141,494,458-141,495,115; at -4341 bp to -5000
126 bp) segment of the examined 6000 bp sequence were strikingly absent in the *UCP1* promoters
127 of mice and rats (Fig. 1B). This absence was corroborated through a BLAST search on the NCBI
128 database (https://blast.ncbi.nlm.nih.gov/). Moreover, examining the presence of the 1419 bp
129 sequence in other species using the Multiz Alignment of 100 Vertebrate tool in the UCSC
130 genome browser (https://genome.ucsc.edu/) suggests its emergence as an evolutionarily
131 novel sequence around the time of opossum divergence, albeit it is missing in certain clades,
132 including mice and rats (Suppl. Fig. 1B). It is noteworthy that short palindromic sequences or
133 their residues can be detected in the human *UCP1* promoter at the beginning and end of these
134 evolutionarily novel DNA fragments (Suppl. Fig. 1C, 1D,1E, 1F, 1G; red boxes). Eukaryotic
135 Promoter Database analysis ($p < 1e-5$) shows that the RE of CREB3 at -3498 bp corresponds to
136 the palindromic "TGACGTCA" sequence at the end of the 1419 bp long segment (Suppl. Fig.
137 1D). Notably, the 468 bp central region of this new sequence (Chr4: 141.492.573-141.493.040;
138 at -2458 bp to -2925 bp), identified as a long terminal repeat (LTR) retrotransposon repetitive
139 element by RepeatMasker (http://www.repeatmasker.org) (Jurka 2000) named MLT1G from
140 the endogenous retroviruses ERVL-MaLR family (Fig. 1B). Its seems to be multiply integrated
141 within the human genome, as indicated by BLAST search (Suppl. Fig. 1H). It is also interesting
142 to note that at the beginning and end of this central segment, homeodomain family TF RE are
143 identified, DLX2 RE at -2478 bp, while at the end for HOXD8, HOXD9, HOXB5, POU6F1 at -2913

144 bp. Remarkably, Eukaryotic Promoter Database does not identify TF RE at the intermediate
145 segment at the p < 1e-5 probability level. While, towards the 3' terminus of the 658 bp
146 sequence absent in the mouse and rat genomes, a 317 bp (Chr4: 141.494.458-141.495.774;
147 located at -4343 bp to -4659 bp) and a 124 bp segment (Chr4: 141.494.775-141.494.898; at -
148 4660 bp to -4783bp) were also recognized by RepeatMasker as a short interspersed nuclear
149 element (SINE) and a long interspersed nuclear element (LINE), respectively. Specifically, the
150 SINE was classified as an Alu sz6 repeat from the ALU family, and the subsequent LINE, was
151 identified as an L1M5 repeat belonging to the L1 family. The BLAST searches have found
152 similar repeat sequences in over 1000 locations in the human genome. Using RepeatMasker,
153 we found several additional repetitive elements in the human *UCP1* promoter, whose possible
154 role requires further investigation (Fig. 1B, grey lines). This observation may underscore the
155 dynamic nature of genetic elements and their potential roles in gene expression regulation
156 during evolution.

157 Nevertheless, the intronic regions of *UCP1* at around +250 bp and the distal enhancer region
158 at around -3600 bp exhibit above-average sequence similarity across species and may indicate
159 conserved DNA regions with REs that are important in the control of transcription, such as
160 Myeloid Zinc Finger 1 (MZF1) at -3667 bp and glucocorticoid modulatory element binding
161 protein 1 (GMEB1) and GLIS family zinc finger 2 (GLIS2) at around +250 bp (Fig. 1B blue boxes;
162 Suppl. Table 2).

163 *1.3 Subtle Structural Differences of UCP1 Orthologous and Paralogous Proteins May Have*
164 *Functional Implications*

165 The primary 3D structures of human UCP1, UCP2, and UCP3 proteins reveal a substantial
166 degree of homology with their mouse and rat counterparts, as established through predictive
167 modeling by AlphaFold 2 and structural alignment conducted via the RCSB PDB interface (Fig.
168 2A, see methods). Supplementary Figure 2A presents these spatial protein structures in a
169 color-coded format, offering insights into their conformational similarities. Remarkably, the
170 model confidently predicts the spatial structure of UCP1 across all three species, while for
171 UCP2 and UCP3, the model confidence tends to be comparatively lower. Interestingly, within
172 the UCP proteins of all three species, short interdomain sequences exhibit notably low model
173 confidence (Suppl. Fig. 2A, red color). This context-dependent variability in spatial structure
174 may provide an opportunity to influence functionality under specific conditions, such as

175 modulating proton conductance or affecting the binding of fatty acids and purine nucleotides.
176 Detailed pairwise structural alignment measurements of the examined proteins, delineating
177 the extent of structural similarity with human UCP1, are outlined in Table 1. These analyses
178 reveal a consistent template modeling score (TM-score) of 0.75 or higher when compared to
179 human UCP1, with equivalent residue numbers notably aligning in all cases (Table 1). In finer
180 detail, the results underscore the remarkable similarity between mouse and rat UCP1 and
181 human UCP1, with an impressive TM score of 0.97. The homology is slightly lower for the UCP2
182 (0.84 TM score) and UCP3 (0.77 TM score) paralogous genes. It is worth noting that the amino
183 acid sequence similarities correspond closely with the result of sequence comparisons, both
184 underscoring the higher similarity in orthologous protein structures. However, the relevance
185 of the amino acid sequence homology in protein structure similarity is questioned by the fact
186 that the presence of the conserved UCP protein structure is a recurrent feature within the
187 SLC25 protein family (Hoang et al., 2013; Crichton et al., 2017), irrespective of homology level
188 between the constituent amino acids. For instance, the ADP/ATP translocase (SLC25A4 or
189 ANT1) exhibits a similar predicted 3D fold structure (Fig. 2B) and a substantial TM score of 0.8,
190 despite a modest 21% amino acid identity homology (Table 1). In conclusion, phylogenetic and
191 structural analyses of *UCP1*, *UCP2*, and *UCP3* across human, mouse, and rat genomes
192 demonstrate the closest relationships among orthologs. Paralogs exhibit distinct clustering
193 patterns based on gene and promoter sequences. The observed structural similarities,
194 irrespective of amino acid sequence homology, underscore the conserved nature of the UCP
195 protein structure within the SLC25 family and raise questions about the functional implications
196 of sequence variations in specific inter-domain regions.

## 197 2. Comparative Analysis of Transcription Factor Binding Sites of *UCP* Promoters in 198 Human, Mouse, and Rat

### 199 *2.1 Comparative Analysis of Putative Transcription Factor Response Elements in UCP1/2/3 200 Promoters Across Species Reveals Higher Similarities Between Human UCP1 and UCP2 than 201 UCP1 orthologs in mouse and rat*

202 Proceeding with a thorough exploration of the promoter sequences of the *UCP1*, *UCP2* and
203 *UCP3* genes in humans, mice and rats, we then focused on determining the specific
204 transcriptional regulation of the human *UCP1* gene. Exploiting the Eukaryotic Promoter
205 Database (EPD) and a diverse array of TF motifs derived from the JASPAR databases, we

206 conducted an in-depth analysis involving 579 TF motifs (Dreos et al., 2017). Employing
207 stringent significance thresholds ($p < 1e-4$), we identified varying numbers of TFs with putative
208 REs across the nine genes under study, ranging from 293 to 381 (Fig. 2C, Suppl. Fig. 2B, Suppl.
209 Table 1). Employing a stricter $p < 1e-5$ threshold resulted in the detection of 63 to 123 TFs with
210 REs in the promoter of the *UCP* genes across the three species (Suppl. Fig. 2C). The distribution
211 of TF REs within *UCP1* promoters exhibited considerable variations among regions and
212 species. While in the case of humans the proximal promoter shows the highest TF binding, in
213 mice and rats it is the enhancer region (Fig. 2D, 2E; Suppl. Table 2). Then, heat map
214 visualization and hierarchical clustering revealed that the regulatory TF RE profile of the
215 human *UCP1* gene closely resembled that of its paralog, human *UCP2* (Fig. 2F), contrary to
216 observations based on protein structure, amino acid composition, gene, and promoter
217 sequences (Fig. 1A, 2A), where orthologs were more akin to each other. This is also supported
218 by the comparison of relative gene expression profiles of human *UCP* genes within the cervical
219 derived adipocyte samples, as *UCP1* was most closely related to *UCP2* based on the correlation
220 using Spearman's rank similarity matrix (Fig. 2E).

221 *2.2 Comparative Analysis of Transcription Factor Response Elements Reveals Unique*
222 *Regulatory Sequences for Human UCP1 Gene Associated with Homeodomain Family Class TFs*

223 In our comprehensive investigation, a total of 364 TFs (including dimers, trimers and variants)
224 with putative REs were identified in the promoter region of the human *UCP1* gene (Figure 2C;
225 Suppl. Table 1). Interestingly, upon comparative analysis with the regulatory regions of the
226 human *UCP2* and *UCP3* genes, 319 of these TFs were found to be commonly present in one of
227 the investigated *UCP*s. However, 45 TFs (including dimers and variants) were exclusive to the
228 *UCP1* promoter. We examined these 45 *UCP1*-specific TFs in more depth, first comparing their
229 expression patterns to *UCP1* in cervical subcutaneous and deep neck preadipocytes, as well
230 as mature adipocytes (Figure 2H) (Tóth et al. 2020). Notably, our analysis revealed that 5 TFs,
231 including *AR*, *RXRA*, *HMBOX1*, *CREM*, and *HOXA5*, displayed a significantly positive correlation
232 with *UCP1* gene expression (Suppl. Table 3). Conversely, 17 TFs, with *JUNB*, *FOSL1/2*, *ATF1*,
233 *VDR*, and *NFE2L2* as the top six, exhibited a negative correlation. However, the RE of RXRA
234 only demonstrates specific occurrence in human *UCP1* promoter when dimerized with VDR.
235 Similarly, only a single variant of the JUNB TF, JUNB_var2, exhibits a specific binding sequence
236 occurrence in the human *UCP1* promoter and only when dimerized with FOSL1 or FOSL2.

237 When they are considered alone or dimerized with other TFs, their REs are identified in all
238 *UCP*s tested, thus they cannot be considered human *UCP1*-specific regulators.

239 Next, in comparison to TFs with REs detected in the regulatory regions of mouse and rat UCP1,
240 UCP2, and UCP3 genes, we identified 13 out of the 45 human UCP1-specific TFs that lacked
241 REs in the promoters of these rodent *UCP* genes, namely: ALX1, ALX4, DLX2, EN1, GSX1,
242 HOXA2, LHX8, MIXL1, NFIL3, PAX4, VSX1, VSX2, and HOXA5 (Figure 2F, red box). Except for
243 NFIL3, these TFs belong to the family of the class of homeodomain family TFs (HDF; JASPAR
244 database). Eukaryotic Promoter Database screening can detect most of these REs at the
245 strongest probability level ($p < 1e-5$) (Table 2A, Suppl. Table 2), supporting that these TFs may
246 potentially regulate thermogenic processes in a species-specific manner, which could add a
247 new layer of complexity to human *UCP1* regulation. Additionally, we examined the RE and the
248 nucleotide sequences at the binding positions identified by the Eukaryotic Promoter Database
249 for these 13 TFs and we noticed a distinctive 10 bp palindromic DNA nucleotide sequence, 5'-
250 TCTAATTAGA-3' positioned at -690 bp in the human *UCP1* promoter (EPD database; UCSC
251 genome browser GRCh 37/hg19: 141.490.801-141.490.810), serving as a RE for 9 out of the
252 13 human *UCP1* specific TFs (Table 2A). This sequence was indeed absent in *UCP2* and *UCP3*
253 promoters, as well as in mouse and rat *UCP*s' regulatory sequences based on EPD database
254 and NBCI BLAST search, highlighting its potentially unique regulatory role in humans. Further
255 data suggested a certain level of conservation of this sequence in primates (one base
256 difference), such as chimps, gorillas, orangutans, baboons, and green-monkey (Figure 1B;
257 Multiz Vertebrate Sequence Alignment in UCSC), broadening the range of possible species in
258 which this regulatory process may be implemented. In addition to the homeodomain family
259 TFs, the nuclear factor interleukin 3 (NFIL3), which belongs to the basic leucine zipper TF family
260 emerged as another potential human-specific *UCP1* regulatory factor, with its binding site
261 positioned at -881 bp in the human *UCP1* promoter, albeit not identified at a stringent
262 probability level ($p<1e-5$).

263 On the other hand, 93 TF REs were found to be consistently present in all UCP promoters from
264 the three investigated species suggesting a degree of shared regulation (Figure 2F Purple box,
265 Suppl. Table 1). These include TFs that have been found in previous studies to possess
266 phylogenetically well-conserved BS in the *UCP1* promoter, such as PPARA, PPARG, RARE,
267 TFAP2, SP, THR and SREBF1 (Gaudry and Campbell 2017).

268 *2.3 Predicting Transcription Factors Binding to the Unique Palindromic Nucleotide Sequence of*
269 *the Human UCP1 Regulatory Region*

270 Next, we aimed to predict the TFs that might bind to the unique nucleotide sequence "5'-
271 TCTAATTAGA-3'" situated at -690 bp in the human *UCP1* regulatory region. In addition to the
272 9 human specific homeodomain family TFs, the Eukaryotic Promoter Database also revealed
273 several homeodomain family TFs ($p < 1e-4$) that potentially recognize this RE but do not show
274 human *UCP1* specificity (Suppl. Table 2). To ascertain which of the proposed TFs are likely to
275 play a functional role and assess their significance in the adipocyte browning process, we
276 conducted a comprehensive analysis (Table 2B). Initially, we verified the presence of the
277 predicted homeodomain family 1 and 2 (HDF1-2) class TF REs using the ConTra v3 platform
278 (stringency: core=0.95; similarity matrix=0.85; http://bioit2.irc.ugent.be/contra/v3/#/step/1)
279 (Suppl. Fig. 3). In addition to validating the Eukaryotic Promoter Database results,
280 phylogenetic relationships can be explored based on the aligned sequences of many species.
281 This analysis, however, did not significantly reduce the set of TFs associated with the *UCP1*
282 promoter at -690 bp, as most of the HDF TFs were also identified by ConTra v3. Therefore, we
283 expanded our investigation to encompass all TFs of the HDF1 class based on JASPAR, as they
284 potentially could bind to the UCP1 promoter at this position. To narrow the range of potential
285 homeodomain family TFs, we examined their relative gene expression levels in our neck-
286 derived adipocyte samples and focused on those demonstrating a significant positive or
287 negative correlation with *UCP1* gene expression (Table 2B; Suppl. Fig. 4, Suppl. Table 3). For
288 some TFs, we did not have gene expression data and therefore could not exclude them from
289 the group of TFs potentially recognizing this specific RE. We gathered additional data on the
290 narrowed group of TFs, including ChipSeq data, Eukaryotic Promoter Database search results
291 using more stringent $p < 1e-5$ criteria, and position weight matrices (PWMs) from the JASPAR
292 database (Table 2A, 2B). Moreover, we utilized the AdipoNET platform (Tóth et al. 2021) to
293 categorize these TFs based on their potential involvement in the formation of brown, white,
294 or both adipocyte phenotypes as "linker" pathway genes. This categorization provided insights
295 into the possible roles of these proteins in specific adipocyte phenotypes. Additionally, we
296 conducted an automated search in the PubMed database to identify co-occurrences of these
297 TFs and *UCP1* in article abstracts, to explore what existing knowledge we have about these
298 TFs in the context of *UCP1* (Suppl. Table 4**,** Suppl. Fig. 5.). The data collected suggested that

299 Engrailed Homeobox 1 (EN1), Paired Related Homeobox 2 (PRRX2), Homeobox A5 (HOXA5),
300 Paired Box 4 (PAX4), and Visual System Homeobox 2 (VSX2) were the TFs most likely to bind
301 to the given sequence. *EN1* and *PRRX2* exhibited high expression in adipocytes and a strong
302 negative correlation with UCP1, while HOXA5 showed low expression but a positive
303 correlation. Although the expression profiles of *PAX4* and *VSX2* were not examined in Tóth et
304 al. (2020), the more rigorous Eukaryotic Promoter Database search also identified REs for
305 these TFs at -690 bp, reinforcing their potential for functional binding at this position.
306 Furthermore, using the ISMARA platform (Balwierz et al. 2014) we observed a high level of
307 activity of PAX4 motif in adipose tissue (Suppl. Fig. 6). Notably, the AdipoNET database and
308 publication data mining so far have not been associated with homeodomain family TFs with
309 adipocyte phenotype formation pathway or *UCP1* expression, thus here we present a novel
310 area for further investigation into the role of these TFs in browning processes.

311 *2.4 Examining the Vicinity of Unique TF REs of Human UCP1 Identifies Potential Regulatory*
312 *Hubs*

313 Homeodomain family TFs are well-established for their participation in embryonic
314 developmental processes, suggesting a plausible role in regulating tissue-specific *UCP1*
315 expression during the maturation of adipocytes (Gehring 1994; Banerjee-Basu and Baxevanis
316 2001). Their recognized function involves the modulation of gene expression through the
317 influence of epigenetic markers that impact chromatin accessibility, thereby controlling the
318 TFs binding to nearby REs. In addition, homo- or heterodimers are often formed as
319 cooperating complexes to enhance their DNA binding specificity (Slattery et al. 2011; Cain et
320 al. 2023). Consequently, our further investigation focused on TF regulatory elements in the
321 proximity to the identified human-specific genomic positions (within a range of +40 bp and -
322 40 bp) (Table 2A) to ascertain which TFs may be allowed to operate.

323 First, we uncovered TF REs close to the -690 position of the *UCP1* promoter. The Eukaryotic
324 Promoter Database, through a stringent search ($p < 1e-4$), identified the presence of REs for
325 B-cell lymphoma 6 (BCL6) at -702, Regulatory Factor X5 (RFX5) at -651/650, and REs for basic
326 helix-loop-helix (bHLH2) class TFs around -679 bp (Fig. 2I). Notably, *BCL6* and *RFX5* exhibited
327 high expression levels in our cervical adipocyte samples and a strong positive correlation with
328 *UCP1* (Suppl. Fig. 7; Suppl. Table 3). Of the bHLH2 TFs with putative RE at -679 bp, *ID4*, *SNAI2*
329 and *TCF3*/4/12 also displayed high gene expression levels, but a negative correlation with

330 *UCP1*. Moreover, through the utilization of the ISMARA platform (Balwierz et al. 2014), the
331 delineated primary interaction network of PAX4 includes TFs identified in the vicinity of its RE
332 at -690 bp, such as BCL6, TCF4, ID4, SNAI2 and RFX5, indicating the potential formation of a
333 regulatory complex (enhanceosome) at this position of the human *UCP1* promoter (Suppl. Fig.
334 8).

335 Additionally, from the identified 13 human UCP1-specific putative TF REs, we explored
336 regulatory elements associated with the other distinctive homeodomain family TFs within the
337 human *UCP1* promoter (Table 2A). The REs of HOXA5, HOXA2, and DLX2, situated at -4785, -
338 3959, and -2478 bp respectively, were present in primate genomes while being absent from
339 the *UCP1* promoters of mice and rats (Table 2A, Suppl. Fig. 3 Contra v3). The HOXA5 and DLX2
340 REs were also identified by the Eukaryotic Promoter Database at the more conservative p<1e-
341 5 probability level, however, only *HOXA5* gene expression showed a correlation with *UCP1*,
342 although all three TFs are expressed at low levels in cervical derived adipocytes. Notably, in
343 the genomic vicinity of HOXA5 RE, we identified the presence of the HIC1 TF RE (Suppl. Fig.
344 9A), which displayed high expression levels in preadipocyte samples but exhibited a significant
345 reduction in mature adipocytes, concurrently displaying a negative correlation with UCP1
346 expression (Suppl. Table 3). A context-dependent regulatory aspect was suggested by the
347 gene-expression profile of TEAD RE detected at -2496 bp, near the DLX2 RE (Suppl. Figure 9B).
348 *TEAD1*, *TEAD3*, and *TEAD4*, expressed at high levels in neck-derived adipocytes, demonstrated
349 a negative correlation with *UCP1* gene expression, while *TEAD2* exhibited a positive
350 correlation. At the same time, *NFKB2*, which has putative RE at -3976 bp and may form a
351 regulatory complex with HOXA2 (Suppl. Figure 9C), exhibited a negative gene-expression
352 correlation trend with *UCP1*, while *ZNF410* and *ZNF283*, with RE at -3947 and -3948
353 respectively, showed no apparent association (Suppl. Table 3). Importantly, these TF REs
354 proximal to the homeodomain family TF REs were also recognized under a more stringent
355 search threshold (p < 1e-5), underscoring the likelihood of their involvement in regulatory
356 complexes.

357 Additionally, our exploration extended to the vicinity of the unique NFIL3 RE identified at -881
358 bp within the human *UCP1* promoter (Table 2A). Here, the presence of the REL RE at -902 bp
359 was discerned (Suppl. Figure 9D, 9E), and also supporting the conservation of these *UCP1*

360 promoter regions in primates while indicating their absence in the mouse and rat genomes
361 (Suppl. Figure 9F).

362 Although gene expression profiles of these identified elements did not always reveal a clear
363 correlation with *UCP1*, this integrated analysis may unveil previously undescribed regulatory
364 mechanisms governing *UCP1* expression. Further research is required to elucidate their
365 precise roles in adipocyte browning processes.

366 *2.5 Common Regulatory Elements in UCP1 Promoters Across Species Occur at Different*
367 *Genomic Positions and Show no Consensus Sequence Conservation Among Human, Mouse and*
368 *Rat*

369 Exploration and comparison of the regulatory components within the human, mouse and rat
370 *UCP1* promoters depicted a set of TF REs occurring in the human *UCP1* promoter, including
371 ARID3B, ATF7, CREB3, MAF::NFE2, NFE2L2, POU6F1, RXRA::VDR, SOX21, TEAD4 and XBP1,
372 which are absent from the human *UCP2* and *UCP3* promoters but were found to also possess
373 REs in the mouse and rat *UCP1* promoters. This group is mostly composed of TFs known for
374 their involvement in *UCP1* gene expression, including well-characterized TFs like CREB/ATF
375 (Cao et al. 2001; Liu et al. 2019), RXR (Alvarez et al 1995), NFE2 and NFE2L2 (Rim and Kozak
376 2002). Upon closer examination of their gene expression profiles (Fig. 2J) and pathway
377 predictions using AdipoNET, it appears that their primary function might lean towards
378 promoting the white adipocyte phenotype, as indicated *by POU6F1*, *TEAD4*, *XBP1*, and *NFE2L2*
379 (Table 3A). Uniquely, *RXRA* exhibits a positive correlation with *UCP1* gene expression (Suppl.
380 Table 3), implying a potential enhancer role in browning processes. Consistently, predictions
381 of AdipoNET propose that RXRA, CREB3, and XBP1 could contribute to the development of
382 both, the white and brown adipocyte phenotypes. Intriguingly, among these, only POU6F1
383 from the homeodomain family TFs was exclusively associated with *UCP1* (absent in mouse and
384 rat *UCP2* and *UCP3*), indicating its potential role in the specific regulation of *UCP1* gene and
385 thermogenesis in these species. This observation would appear to suggest the presence of
386 evolutionarily conserved TF REs in the *UCP1* promoter across species. However, a
387 comparative analysis of *UCP1* regulatory sequences across species
388 (https://genome.ucsc.edu/) revealed that most of these BSs were recognizable in primates'
389 consensus sequences, but not in mice and rats, signifying that these REs were not conserved

390 across these species, but instead appeared in different positions within mouse and rat *UCP1*
391 promoter (Suppl. Figure 10A, 10B; for TF RE positions in different species see Suppl. Table 2).

392 *2.6 Common Species-Specific bHLH TF Binding Sites in the Proximal Promoter of Human UCP1-*
393 *2-3 Genes*

394 Finally, our comparative analysis unveiled four TF regulatory elements (HES1, MAX, MNT,
395 TCFL5) exclusive to human *UCP* genes (Suppl. table 1). These TFs possess binding sequences
396 within the regulatory domains of human *UCP1*, *UCP2*, and *UCP3* genes but lack representation
397 in the promoters of any of the investigated *UCP* genes within the mouse and rat (Suppl. Fig.
398 11A). Interestingly, all of these TFs have an RE in the proximal promoter region of the human
399 *UCP1* at -263 bp. This region pertains to the basic helix-loop-helix (bHLH) family TFs, and the
400 Eukaryotic Promoter Database assigns additional TFs to this nucleotide sequence in over the
401 above four (Suppl. Table 2). Therefore, further studies are needed to discern which context-
402 dependent regulatory mechanisms might allow TFs to be actual users of this RE.

403 2.7 *Predicted and Experimentally Validated TF Binding Sites in the Human UCP1 Promoter*
404 *Exhibit little overlap*

405 In our investigation, we conducted a detailed comparative analysis of putative TF REs
406 identified through the Eukaryotic Promoter Database (https://epd.epfl.ch) in the regulatory
407 region of the human *UCP1* gene, spanning from -5000 to +1000 bases from the transcription
408 start site, at a significance level of $p < 1e-4$. Subsequently, we compared these findings to
409 experimentally validated binding proteins/TFs for the same DNA sequence using the ChiPBase
410 (https://rnasysu.com) and TFlink (https://tflink.net/) databases. Intriguingly, while we
411 identified 364 putative TF REs using the Eukaryotic Promoter Database, ChiPBase recognized
412 170 and TFlink 131 experimentally validated TFs, respectively (Fig. 3A). Typically, only a small
413 fraction of the experimentally identified binding TFs had a matching binding sequence
414 recognized by the Eukaryotic Promoter Database (ChiPBase: 23 TFs,14%; TFlink: 41 TFs, 31%).
415 Conversely, we found that only 53 TFs (15%) of those with REs according to the Eukaryotic
416 Promoter Database were associated with laboratory validation based on these two databases.
417 Notably, just 11 TFs were identified from all three databases (GATA2, HIF1A, JUN, KLF4, MXI1,
418 MYC, MYCN, NFKB1, RBPJ, TEAD4 and USF1) emphasising their potential role in regulating
419 human *UCP1* gene expression. The gene expression patterns of these TFs in neck-derived

420 adipocyte samples (Tóth et al. 2020) further support this, with most exhibiting a strong
421 correlation with *UCP1*, mainly negative, while Upstream Transcription Factor 1 (*USF1*) showed
422 a positive correlation (Figure 3B). USF1, with REs identified in all UCPs of the three species
423 studied, seems to act as a general regulator of *UCP*s' gene expression rather than a specific
424 regulator of *UCP1* and browning processes (Suppl. Table 2) (Laurila et al. 2016). AdipoNET
425 predictions also indicate the involvement of most of these TFs in white adipocyte phenotype
426 development, while JUN and NFKB1 may have roles in both white and brown adipocyte
427 phenotypes (Table 3B). Response elements of HIF1A, JUN, RBPJ, and TEAD4 were identified at
428 more stringent significance levels ($p < 1e-5$), strengthening their potential involvement in
429 regulating *UCP1* gene expression. Interestingly, four of these 11 TFs (HIF1A, MYC, MYCN and
430 MXI1) share a common binding location around -263 bp in the proximal promoter region,
431 which already stood out in our investigation as a potential site for fine-tuned, state-dependent
432 regulation of the human *UCP*s' gene-expression (Suppl. Figure 11A). This RE is not
433 evolutionarily conserved in mouse and rat *UCP* genes, highlighting the unique control aspects
434 of *UCP1* in humans and the potential for the formation of a regulatory module with SP2 and
435 EGR1 TFs, which have RE in the vicinity (Suppl. Table 2), based on data revealed by ConTra v3
436 (Suppl. Fig. 11B, 11C, 11E).

437 **DISCUSSION**

438 **1. Evolutionary Dynamics of the UCP1 Genes and its Regulatory Elements Differs in**
439 **Humans and Rodents**

440 *1.1 Research Trends and Major Differences Between Human and Rodent UCP1 Promoter*

441 *In silico* studies based on DNA sequences have emerged as a fundamental approach to
442 understanding the evolutionary dynamics of gene-specific functions. Generally, phylogenetic
443 analyses have focused primarily on coding sequences, and by now, comparative analysis of
444 orthologous genes for *UCP1* surpasses hundred species, providing insights into complex
445 phylogenetic relationships (Cassard et al. 1990; Klingenspor et al. 2008; McGaugh and
446 Schwartz 2017, Hughes et al. 2009; Jastroch et al. 2005, 2008, 2018; Berg et al. 2006; Hou et
447 al. 2017; Gaudry et al. 2017). Meanwhile, evolutionary studies of transcriptional regulation
448 remained strikingly limited. Notably, Gaudry and Campbell (2017) undertook a comparative
449 analysis of *UCP1* transcriptional regulatory elements across mammals, although only focusing

450 on described REs, particularly in the core promoter and distal enhancer. Their work has shed
451 light on key evolutionarily conserved sites such as the TATA box, RARE-3 and PPRE motifs that
452 are expected to be functionally relevant in Eutherian phylogeny (Gaurdy and Campbell, 2017;
453 McGaugh and Schwartz, 2017). Our investigations substantiate the ancient origin of these
454 regulatory elements, indicating a remarkable level of conservation. According to Eukaryotic
455 Promoter Database searches, these elements are prevalent in the promoter regions of *UCP*
456 genes across all three studied species. This prevalence implies a fundamental role in gene
457 expression regulation within this gene family. However, due to their extensive conservation
458 and distribution, these elements may be less amenable as specific targets for precise
459 modification of *UCP1* gene transcription, which is the ultimate objective of controlling
460 metabolic disorders. Nevertheless, the conclusions taken from Gaudry and Campbell's (2017)
461 results suggested a limited degree of conservation of transcriptional control of gene
462 expression within the mammalian clade. This observation points to the possible existence of
463 novel regulatory and functional elements within human *UCP1* promoters that potentially
464 accommodate and coordinate for species-specific conditions and needs, and allow the
465 development of more precisely targeted interventions.

466 *1.2 Differences in the regulatory region of UCP1 between humans and rodents*

467 Several interfaces are currently being developed for predicting DNA binding of TFs with known
468 Position Weight Matrices based on the promoter nucleotide sequence, such as the Eukaryotic
469 Promoter Database we used, TRANSFAC2.0 (https://genexplain.com/transfac-2-0/), GTRD
470 (Yevshin et al. 2017), or ConTra v3, which is also available for promoter alignment analysis for
471 identification of TF REs across species. Our comparative analysis of the upstream regulatory
472 segments in the *UCP1* genes of humans, mice, and rats brings to light the intricate interplay
473 between sequence conservation and divergence characterizing the evolutionary trajectory of
474 these genes. While a substantial fraction of the identified cis-acting regulatory elements in
475 these segments are shared across rodent and human *UCP1* genes (174 out of 364; Suppl. Fig.
476 2B), in broad concurrence with earlier findings (Collins et al., 2004), our sequence similarity
477 analysis uncovers large disparities (Fig. 1B), which opens up the possibility for the emergence
478 of different regulatory modules (Saito et al. 2008). Notably, a significant portion of the distal
479 enhancer region of the human *UCP1* promoter is absent from the mouse and rat genomes.
480 This may be attributed to secondary losses in mouse and rat ancestors following the

481 appearance of an evolutionarily novel segment after the divergence of the opossum lineage.
482 Remarkably, these DNA sequence losses did not lead to species extinction, indicating that its
483 existence was not vital or that potential compensatory mechanisms enable sufficient TF
484 binding to other parts of the regulatory region. Our study supports the possibility of divergent
485 gene-expression regulation among these species, as the hierarchical cluster analysis relying
486 on identified putative REs in the *UCP* promoters reveals distinctions between human and
487 rodent *UCP*s. Specifically, the regulatory profile of human *UCP1* exhibits the highest proximity
488 to the human *UCP2*, while mouse *UCP1* demonstrate a closer relationship to rat *UCP1*.
489 Detailed analysis also identified unique sequence elements that could potentially form the
490 basis of human-specific regulation. Recent studies indicate interspecies variability in the
491 regulation of *UCP1* expression too, revealing differential processing of the 3'-untranslated
492 region (3'-UTR) of *UCP1* mRNA between mice and humans that quantitatively affects *UCP1*
493 synthesis and thermogenic activity (Lu et al., 2021). These disparities call for caution when the
494 interpretation of results gained from research carried out on rodent models is transferred to
495 humans.

496 *1.3 Repetitive Elements*

497 Moreover, we identified numerous repeat elements within the human *UCP1* promoter region,
498 including those embedded within evolutionarily novel sequences, missing in the
499 corresponding regulatory regions of rodents (Fig. 1B). Additionally, BLAST searches revealed
500 the presence of these repeats at multiple loci across the human genome (Suppl. Fig. 1H).
501 However, the mechanisms underlying their integration remain ambiguous, particularly
502 regarding whether these mobile elements integrate randomly or are directed by pre-existing
503 genomic sequences. Integration of this repetitive element in the human genome may organize
504 the expression of nearby genes into a regulatory circuit. Notably, our observations indicate
505 the occurrence of short palindromic sequences flanking the novel DNA segments and regions
506 recognized as repetitive sequences, some of which are annotated as TF REs by the Eukaryotic
507 Promoter Database. Palindromic sequences in DNA are traditionally recognized as specific
508 binding sites for nucleotides or proteins, resulting in functional consequences (Pingoud and
509 Jeltsch 2001; Linheiro and Bergman 2008). This suggests that these binding motifs may
510 represent ancient genetic elements, that may have served integration into the genome and
511 potentially repurposed by eukaryotic cells as regulatory modules governing gene expression

512 (Thompson et al. 2016). An intriguing observation related to the identified 5'-TCTAATTAGA-'3
513 palindrome sequence at -690 bp in the human *UCP1* promoter supports this hypothesis. This
514 nucleotide sequence binding TFs of the homeodomain family was previously identified by Li
515 and co-workers (2005) as a distinctive DNA fragment located at one endpoint of the 1165 bp
516 sequence differentiating between two genotypic variants of Mamestra configurata
517 nucleopolyhedrovirus-A (MacoNPV-A). These variants include the lower virulence variant 90/4
518 and the archetypal variant 90/2. They concluded that the recognised palindrome sequence
519 raises the possibility that the distinction between these viral variants is due to the insertion or
520 deletion of a transposable element. Other studies also propose that transposable elements
521 and short tandem repeats can influence gene expression regulation within the genome
522 (Horton et al., 2023; Chuong et al., 2016; Kunarso et al., 2010). For example, Scarpato and
523 colleagues (2015) have suggested that some repetitive elements, such as SINEs, are highly
524 prone to hybridize with microRNAs thereby modifying the transcriptional activity of nearby
525 genes. Together, these support the possibility that species-specific differences in transposable
526 and repetitive elements in the *UCP1* promoters underlie the evolution of specific regulatory
527 mechanisms.

528 **2. Comparative Analyses of Putative RE in *UCP1*, *UCP2* and *UCP3* promoters in Humans,**
529 **Mice and Rats Pinpoint the Possible Regulatory Role of Homeodomain Family TFs.**

530 *2.1 The Human UCP1 Specific 5'*-TCTAATTAGA-3' *Binding Sequence*

531 Our study key objective was to identify unique regulatory elements within the *UCP1* promoter
532 specific to the human genome that may have been hidden in " exploratory" studies, usually
533 performed in rodents. The majority of the recognised putative regulatory factors of human
534 *UCP1* are general regulators of the investigated *UCPs*, classifying them as less specific targets.
535 Uncovering the distinct regulatory elements opens the opportunity for more adequate
536 targeted interventions for influencing metabolic processes by promoting adipocyte browning.
537 Our investigation, focused on the identified human *UCP1* unique putative homeodomain
538 family TF REs and their vicinity, mostly a revealed palindromic nucleotide sequence, 5'-
539 TCTAATTAGA-3', located at the -690/689 bp position in the human *UCP1* gene (Fig 2I, Fig. 4).
540 It provides an opportunity for homeodomain family TFs to bind to both DNA strands, allowing
541 for homo- or heterodimer formation and stronger tethering (Jolma et al. 2013; Cain et al.
542 2023; Dror et al. 2014). It is also possible that several homeodomain family TFs may

543 alternatively use this BS depending on the circumstances, and these TFs might differ in
544 whether they promote or block the expression of the gene (Foulkes et al 1991). The
545 homeodomain family is one of the largest groups of TFs in metazoans, consisting of almost
546 200 family members in humans (Lambert et al. 2018). Their recognised functions include
547 regulating developmental and homeostatic processes by modulating the availability of DNA.
548 (Waldén et al. 2012; Hilton et al. 2015; Cantile et al. 2003), yet their potential for species-specific
549 regulatory roles in adipocyte phenotype remodelling has been underappreciated. These TFs can bind
550 highly similar AT-rich monomer sequences, with minor differences in the nucleotides still
551 sufficient for specific interaction and modulate DNA openness to allow or deny nearby REs to
552 be accessed by TFs (Park et al. 2009). In our analysis, regulatory elements such as bHLH type
553 2 TFs, BCL6 and RFX5, which were identified in proximity to this palindromic sequence unique
554 to human *UCP1*, may further enhance the probability of specific attachment by forming a
555 complex/enhanceosome with homeodomain family TFs via a process called "epistatic
556 capture" (Emera and Wagner 2012). In a previous study that analyzed the pathways enriched
557 by genes with differential expression in cervical subcutaneous and deep-neck samples, both
558 in preadipocytes and in differentiated mature adipocytes, the homeodomain family TF
559 pathway was found to be significantly dominant in subcutaneous adipocytes, indicating their
560 possible active involvement in permitting browning processes (Tóth et al. 2020).

561 *2.2 Candidate Binding TFs of the Human-Specific Nucleotide Sequence* '5-TCTAATTAGA-3'

562 Among the homeodomain family TFs, we propose several potential binding candidates. EN1,
563 which is abundant in human neck-derived adipocytes with gene expression negatively
564 correlated with *UCP1* has been previously shown to be a positive regulator of brown
565 adipogenesis in mice (Zhang et al. 2016; Singh et al. 2015; Atit et al. 2006), however, it was
566 downregulated after the animals were exposed to cold. Previous research indicated its
567 upregulation in the obesity-prone phenotype of human adipocytes carrying the FTO
568 rs1421085 T-to-C single-nucleotide polymorphism (SNP) (Tóth et al. 2020, unpublished data),
569 suggesting a diverse, species-specific role in the regulation of *UCP1* expression. Conversely,
570 another candidate, HOXA5, exhibited a gene expression profile which may suggest its positive
571 effect on browning in cervical-derived adipocytes (Fig. 2H, Suppl. table 3). In line with this
572 study, it has been reported that HOXA5 exhibited elevated expression in rodents and had an
573 anti-inflammatory role in enhancing adipose tissue browning (Cao et al. 2018; Kato et al. 2021;

574 Singh et al. 2016). At the same time, a human study showed elevated expression in white
575 adipocytes derived from induced pluripotent cells (Mohsen-Kanson et al. 2014). This
576 discrepancy also suggests that HOXA5 may have context-dependent roles in shaping the
577 adipocyte phenotype in each species and fat depots.

578 Regarding PAX4, its known role in pancreatic beta cell development and function is associated
579 with maintaining glucose homeostasis (Brun et al. 2008; Xu et al. 2017), and abnormalities in
580 its expression are linked to metabolic diseases like diabetes (Lau et al. 2023). However, its role
581 in adipocyte function remains less understood. PAX4 is expressed in adipocyte progenitor
582 mesenchymal cells (Xu et al. 2017). Furthermore, our study suggests its potential involvement
583 in adipogenesis regulation, including balancing between white and brown adipocyte
584 formation (Suppl. Fig. 6, Suppl. Fig. 8). As for the other homeodomain family TFs PRRX2 and
585 VSX1/2, their roles in the context of thermogenic response have not been previously
586 characterized. Nonetheless, single-nucleus-based RNA sequencing studies have identified
587 PRRX2 with differential activity in brown versus white adipogenesis in human cells (Gupta et
588 al 2023), indicating a potential role during thermogenic adipocyte formation.

589 Generally, TF REs identified in the human *UCP1* promoter regions are often evolutionarily
590 conserved in primates, but not in mice and rats. Even the identified common TF REs in *UCP1*
591 orthologous promoters, for example, POU6F1 TF from the homeodomain family TFs, are
592 located at evolutionary novel promoter segments and do not maintain evolutionary
593 conservation in consensus genomic sequences. It indicates a kind of "mobility" of regulatory
594 elements and possible convergent evolutionary processes. Nonetheless, POU6F1, based on
595 our data mining efforts, has not been previously linked to adipocyte function. Recent research
596 has identified its role in lung biology and cancer-related processes unveiling its interaction
597 with RORA in the lung, leading to the downregulation of *HIF1A* gene expression and potential
598 inhibition of adenocarcinoma cell proliferation (Xiao et al., 2022). Nevertheless, certain POU
599 factors, as demonstrated in previous studies (Takahashi et al., 2007), serve as effective
600 mediators of cell reprogramming. Given its substantial expression in mature adipocytes and a
601 notable negative correlation with *UCP1* gene expression (Table 3A, Suppl. Table 3), it is
602 plausible that POU6F1 may play a role in adipocyte phenotype switching and that this function
603 is conserved throughout mammalian evolution.

604 *2.3 Limitations in Identifying Functional Regulatory Elements in the Human UCP1 Promoter*

605 The comparative analysis revealed a small overlap between the TF REs predicted by our study
606 based on DNA sequence analyses and those identified experimentally in the human *UCP1*
607 promoter using ChipSeq highlighting the challenges inherent in identifying functional
608 regulatory elements. TFs, like other proteins, possess the ability for non-specific DNA binding,
609 which is essential for efficient target site localization through facilitated linear diffusion
610 (Pingoud and Jeltsch 2001). Such non-specific binding usually involves interactions with the
611 DNA backbone rather than the bases. In contrast, specific binding is characterized by a close
612 interplay between direct (interaction with the bases) and indirect (interaction with the
613 backbone) readout. The frequent occurrence of non-functional TF binding can introduce
614 significant ambiguity in ChipSeq laboratory experiments concerning functional DNA-TF
615 binding. Moreover, the infinite combinations of biological tissues, developmental stages, and
616 specific conditions significantly influence *ad hoc* TF binding to specific DNA segments, making
617 results generalization challenging. While position weight matrices (PWMs), used for *in silico*
618 prediction of regulatory factors, are instrumental in predicting TF binding sites within
619 promoter sequences, they also come with numerous limitations. One is the random
620 occurrence of the short binding sequence in the genome, which highlights the need for
621 rigorous statistical analysis of PWM applications to distinguish between biologically significant
622 TF binding events and chance matches, as well as screening approaches to narrow down gene
623 specificity. The Eukaryotic Promoter Database used employs robust statistical tests to predict
624 PWM in a DNA sequence, thus providing the best possible prediction. Further limitations of
625 the *in silico* approach are our incomplete knowledge of the binding sequences of TFs and the
626 inability of PWMs to fully capture the sequence flexibility inherent in TF binding sites. TFs can
627 tolerate variations in the DNA sequence at their binding sites, leading to a degree of
628 "sloppiness" or degeneracy. This sequence flexibility allows TFs to recognize and bind to a
629 range of related sequences, which may differ from the PWM-defined consensus motif.
630 Therefore, PWMs may miss functional TF binding sites that deviate from the consensus
631 sequence or incorrectly predict non-functional sites that closely match the consensus. In
632 addition, PWMs do not account for the context-dependent nature of TF binding, where
633 neighbouring DNA sequences or chromatin structures can influence binding specificity. As a
634 result, both predictions may overlook biologically relevant binding events or produce false
635 positives, highlighting the importance of integrating experimental and complementary
636 computational approaches to improve accuracy in identifying TF binding sites.

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

638 Prediction of human *UCP1*-specific TFs determined by Eukaryotic Promoter Database and from
639 ChipSeq experiments converge to highlight specific DNA regions with higher TF RE density
640 within the human *UCP1* promoter (Fig. 4). Such clustering of regulatory binding sites may give
641 rise to enhanceosomes, where molecular elements, such as TFs and cofactors form complex
642 multilevel regulatory networks that finely orchestrate gene expression. Notably, in the
643 enhancer region of the human *UCP1* promoter, emerging cluster-forming nucleotide
644 sequences are observed at various homeodomain family binding sites, often located at the
645 beginning or end of regions identified as repetitive elements. This observation hints at a
646 potential role for these segments in ancient transposition events. Moreover, in the proximal
647 promoter, multiple REs appear close to the bHLH TF family RE identified at -262/-263 bp,
648 suggesting the possibility of regulatory complex formation (see Suppl. Fig. 11B, 11C, 11D). For
649 instance, the close binding sites for EGR1 and SP2 offer a scenario where dynamic regulation
650 could occur through interaction with HIF1A, potentially modulating *UCP1* expression based on
651 cellular conditions. This proposed mechanism shares similarities with those detailed in non-
652 small-cell lung carcinoma, where interactions between HIF1A and EGR1, as well as HIF1A and
653 SP1, are pivotal to the regulation of EPOR gene expression in response to changing hypoxic
654 conditions (Su et al. 2019). Given the well-established role of hypoxia and HIF1A in adipose
655 tissue function, the potential involvement of HIF1A in *UCP1* transcription through promoter
656 binding and interaction with other TFs may be an interesting area of research (Tóth et al. 2021;
657 Basse et al. 2017). Furthermore, the involvement of EGR1 in metabolic diseases has already
658 been demonstrated, indicating its capacity to influence adipocyte insulin sensitivity and can
659 tilt PI3K/Akt and MAPK signaling pathways (Zhang et al. 2013). It is conceivable that minor
660 variations in the levels of various components within *UCP1* enhanceosomes may
661 synergistically interact, resulting in significant differences in *UCP1* expression, as previously
662 proposed in comparative studies in mice (Xue et al. 2005). Among potentially related
663 cofactors, p300/CBP-associated factor, with well-described histone acetyltransferase (HAT)
664 activity stands out as a frequently described partner of TFs with a putative RE in the human
665 *UCP1* promoter, indicating an essential role for them in shaping the adipocyte phenotype (Fig.
666 4). Determining the prominent complex components, delineating their roles and

667 understanding which components can be substituted or omitted during the initiation of
668 human *UCP1* transcription under a specific condition a promising future research direction.

669 **III. METHODOLOGY**

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

671 To generate a combined human, mouse and rat UCP1, UCP2, and UCP3 amino acid, coding
672 sequence and promoter sequence phylogenetic tree, the amino acid and the mRNA data set
673 was downloaded from the National Library of Medicine (https://www.ncbi.nlm.nih.gov),
674 whereas the sequences of gene promoters were downloaded from the EPD database (link:
675 https://epd.epfl.ch/). Accession numbers of the promoters' FASTA sequences: human *UCP1*
676 FP007435, *UCP2_1* FP016503, *UCP2_2* FP016504, *UCP3_1* FP016505, *UCP3_2* FP016506;
677 mouse *UCP1* FP011827, *UCP2* FP010492, *UCP3* FP010490; rat *UCP1* FP011518, *UCP2*
678 FP000767, *UCP3* FP000766; Accession numbers of the mRNAs' FASTA sequences: human UCP1
679 NM_021833.5, UCP2 NM_001381943.1, UCP3 NM_003356.4; mouse UCP1 NM_009463.3,
680 UCP2 NM_011671.6, UCP3 NM_009464.3; rat UCP1 NM_012682.2, UCP2 NM_019354.3,
681 UCP3 NM_013167.2; Accession numbers of the amino-acids' FASTA sequences: human UCP1
682 NP_068605, UCP2 AAC51336, UCP3 AAC51367.1; mouse UCP1 NP_033489, UCP2 AAB17666,
683 UCP3 AAB87084; rat UCP1 NP_036814, UCP2 BAA25698, UCP3 AAB71523. The resulting
684 amino acid, exon and promoter sequences were aligned using MAFFT 7.505 (Katoh &
685 Standley, 2013), and a maximum likelihood tree was constructed using IQtree2 2.0.7 (Minh et
686 al., 2020) with automatic model selection ('-m MFP') turned on and using 1 000 aLRT and 1
687 000 rapid bootstrap replications to test the robustness of the analysis. The alignment of the
688 mRNA sequences had a length of 3,398 bp, of which 1536 were variable and 878 informative,
689 whereas the promoters had 7,956 variables and 5,412 informative sites on an alignment
690 length of 12585 bp. Of the 313 aligned amino acid positions, we identified 172 polymorphic
691 and 150 informative characters. Phylogenetic trees were visualized with packages ape 5.7.1
692 (Paradis & Schliep, 2019) and phangorn 2.11.1 (Schliep, 2011) using R 4.2.2 (R Foundation for
693 Statistical Computing, n.d.). Pairwise genetic distances of *UCP* promoters and exons were
694 calculated with the R package pegas 1.2 (Paradis, 2010) using the "K80" model of nucleotide
695 substitution (Kimura's 2-parameters distance metric; Kimura, 1980). Pairwise distances were
696 visualized as a neighbour-joining phylogram with pegas, as a UPGMA dendrogram with

697 phangorn, and as a heatmap with pheatmap 1.0.12 (https://cran.r-
698 project.org/web/packages/pheatmap/index.html).

699 **2. 3D Protein Structure Prediction, Alignment and Pairwise Structural Comparison**

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

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

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

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

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

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

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

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

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

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

739 **4. Computational Identification of Putative Transcription Factor Response Elements in**
740 **Gene Regulatory Sequences Using the EPD Interface**

741 The identification of putative TF REs within the regulatory sequences of the genes under study
742 was performed using the Eukaryotic Promoter Database (EPD) interface (https://epd.epfl.ch)
743 (Périer et al. 2000). EPD's approach to identify TF REs relies on the JASPAR Transcription Motifs
744 Library (specifically JASPAR CORE 2018 vertebrates) (Sandelin et al. 2004), which includes a
745 diverse set of TF motifs, including generic promoter motifs such as the TATA box, Initiator, GC
746 box and CCAAT box. This library contains a large body of information represented as a position
747 weight matrix (PWM) for 579 TF-binding sequences. Many approaches have been developed
748 to model and learn a protein's DNA-binding specificity, however, comparative analyses
749 indicate that simple models based on mononucleotide PWMs trained by the best methods
750 perform similarly to more complex models (Weirauch et al. 2013). To determine whether a
751 given binding sequence is present within the regulatory sequence of the target gene, the EPD
752 interface uses a probability-based algorithm with four different p-value thresholds (p-value <
753 0.01, 0.001, 0.0001 and 0.00001). In this study, we focused on the two most stringent p-value
754 thresholds ($p < 0.0001$ and $p < 0.00001$), as empirical evidence from the literature suggests
755 that more permissive probabilities can lead to high false positive rates (Benita et al., 2009).

756 However, our exploratory studies showed that for some genes with experimentally proven
757 regulatory TFs, the software accurately identified the binding sequence only at $p < 0.0001$, as
758 exemplified by the occurrence of the HIF1A TF RE for genes such as *VEGFA* and *GLUT1*. In
759 addition, considering that every 4000 nucleotides, a given RE of average length 12 bp can be
760 expected to be found randomly (Wunderlich and Mirny, 2009), the identified occurrence itself
761 has limited meaning. It rather suggests that TFs for which we did not identify the BS are less
762 likely to be involved in the regulation of *UCP* gene expression. Consequently, we presented
763 our findings primarily utilizing the less stringent $p < 1e\text{-}4$ criteria, albeit occasionally
764 referencing $p < 1e\text{-}5$ outcomes, thus reducing the number of false negative hits, which is more
765 valuable information in our comparative analysis. In this study, we analysed regulatory
766 sequences 5 kb upstream and 1 kb downstream of the TSS for the genes under study (eg.:
767 human *UCP1*: GRCh37/hg19; Chr. 4: 141,489,115-141,495,114; TSS: Chr. 4:141,489,115). The
768 relevant base-pair boundaries of the upstream regulatory DNA segment may vary among
769 genes, therefore the widest DNA region allowed by the EPD interface was analysed. Data
770 retrieval was automated using a custom script, which we have made publicly available on
771 GitHub, increasing transparency and reproducibility.

772 **5. Comparative Analysis of the Putative TFs in the Regulatory Sequences of Human,**
773 **Mouse and Rat *UCP1*, *UCP2* and *UCP3* Genes**

774 Data analysis and screening were conducted in an R interactive environment. Initially, data
775 were extracted from the EPD (https://epd.epfl.ch) interface as individual data files (text files),
776 corresponding to each target gene. EPD utilizes the JASPAR database, which contains PWMs
777 of TF REs for examining their presence within specific DNA sequences
778 (https://jaspar.genereg.net). Subsequently, the data extracted for each gene were
779 consolidated into a unified database that compiled occurrence information for the 579 TF
780 binding sequences within the promoter regions of three genes: *UCP1*, *UCP2*, and *UCP3* in
781 humans, mice, and rats. This resulting matrix was then utilized to identify distinct clusters of
782 TF REs based on specific criteria. These criteria included identifying TF REs that were specific
783 to human *UCP1* (not present in human *UCP2* or *UCP3* but potentially present in mouse and rat
784 *UCP1*, *UCP2*, or *UCP3*), unique to human *UCP1* (exclusive to human *UCP1*, absent in human
785 *UCP2*, *UCP3* as well as in mouse and rat *UCP1*, *UCP2*, and *UCP3*), specific to *UCP1* (present in
786 both human and mouse and rat *UCP1*, absent in human *UCP2* and *UCP3* but potentially

787 present in mouse and rat *UCP2* and *UCP3*), common to human, mouse, and rat *UCP1* (absent
788 in human, mouse, and rat *UCP2* and *UCP3*), and general human *UCP* TF REs (occurring in both
789 human *UCP1* and *UCP2* and *UCP3* but not in the promoter regions of mouse and rat *UCP1*,
790 *UCP2*, and *UCP3* genes).

791 **6. Chromosome position of the investigated *UCP1*, *UCP2* and *UCP3* promoters and**
792 **genes in the human genom**

793 *UCP1* promoter in human: GRCh 37/hg19 Chr 4:141,489,115 – 141,495,114 size: 6000 bases;
794 *UCP1* gene in human: GRCh 37/hg19 Chr 4:141,480,585 – 141,490,115 size: 9531 bases, minus
795 strand; *UCP2* promoter in human: GRCh 37/hg19 Chr11:73,693,248 - 73,699,247 size: 6000
796 bases; *UCP2* gene in human: GRCh 37/hg19 Chr11:73,685,717 - 73,694,247 size: 8531 bases;
797 *UCP3* promoter in human: GRCh 37/hg19 Chr11:73,719,130 - 73,725,129 size: 6000 bases;
798 *UCP3* gene in human: GRCh 37/hg19 chr11:73,711,322 - 73,720,130; size: 8809 bases.

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

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

813 **8. Heat Map Visualization and Hierarchical Cluster Analyses of the TF RE Profile of the**
814 **UCP genes and the Transcriptomic Data of Human Adipocytes**

815 Hierarchical cluster analyses and heat map visualization were performed with the
816 Morpheus web tool (https://software.broadinstitute.org/morpheus/) using Spearman rank
817 correlation of rows and columns and complete linkage based on the presence or absence of
818 the TF RE in the given gene upstream regulatory region. The same tools were used to visualise
819 relative gene-expression profile of human neck-derived adipocytes (Tóth et al. 2020). RNAseq
820 data is deposited to Sequence Read Archive (SRA) database
821 (**https://www.ncbi.nlm.nih.gov/sra**) under accession number PRJNA607438. Heatmaps
822 depict the z-scores calculated from the filtered and normalized transcriptomic data of isolated
823 and maintained preadipocytes from subcutaneous and deep neck adipose stromal cells and
824 lipid-laden mature adipocytes differentiated *in vitro* with white and brown protocols
825 described in Tóth et al. (2020).

826 **9. Comparison of the Identified Putative TF REs in the Regulatory Sequence of UCP1**
827 **with the Corresponding Data Found in the Chip Based Databases**

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

834 **10. Identification of Repeat Elements in the Genome**

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

839 **11. Literature Mining Using a Custom Script**

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

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

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

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

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

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

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

874 **12. Identification of TF Motif Activity in Different Tissues and Exploration of TF**
875 **Interaction Partners Using the ISMARA Online Platform.**

876 We obtained RNA-seq expression profiles from 16 human cell types using the Illumina Body
877 Map 2 dataset (GEO accession GSE30611), accessible within the Integrated System for Motif
878 Activity Response Analysis (ISMARA) online platform
879 (https://ismara.unibas.ch/supp/dataset1_IBM_v2/ismara_report/) (Balwierz et al. 2014). We
880 employed ISMARA to analyze the motif activity of putative species-specific and species-
881 common homeodomain family TFs across different human tissues for identification of TF motif
882 activity. By leveraging the ISMARA platform, we investigated the motif activity patterns of
883 these TFs, providing insights into their potential roles in tissue-specific gene regulation.
884 Furthermore, we explored potential TFs and their firs-level TF interaction partners using
885 ISMARA to uncover regulatory networks and complex formations, such as enhanceosomes.

886 **13. Statistical Analyses**

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

892 **Figure legends**

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

902 **Figure 2. 1A**. RCSB PBD aligned, AlphaFold 2 predicted protein structures of human, mouse and rat UCP1, UCP2
903 and UCP3 from three different perspectives. **1B.** AlphaFold 2 predicted H-UCP1 aligned with H-ANT1. **1C.** Number
904 of TFs with RE in human, mouse and rat *UCP1*, *UCP2* and *UCP3* Promoter: -5000 bp – +1000 bp (p<1e-4) (Venn
905 program in R environment). **1D.** TF RE frequency in human, mouse and rat *UCP1* gene promoters -5000 bp –

906 +1000 bp (p<1e-4) (Excel). **1F.** Bars show the frequency of TF RE per 10 bp in different parts of the UCP1 promoter
907 in humans, mouse and rat -5000 bp – +1000 bp (p<1e-4). Distal enhancer -5000 bp – -1001 bp; Proximal and Core
908 promoter-1000 bp – TSS; Intronic promoter TSS – +1000 bp (Excel). **2F**. Hierarchical cluster analysis of the human,
909 mouse and rat *UCP1*, *UCP2* and *UCP3* genes based on the TFs detected in their promoters by RE. The red box
910 highlights the 13 TFs that are unique to the human *UCP1* promoter, the purple box shows the 93 TFs that are
911 common to the examined UCPs, while the green box shows the clusters with the most similar TF RE profile to
912 *UCP1* (Morpheus Interface). **2G** Heatmap shows the gene expression profile of the six *UCP* genes in adipocyte
913 samples of cervical origin. Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies
914 *UCP*s with similar gene expression profiles to *UCP1* (Morpheus Interface). **2H** Heatmap shows the gene
915 expression profile of 45 TFs with RE in the human *UCP1* promoter in adipocyte samples of cervical origin. For
916 comparison, *UCP1* is also shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix
917 identifies TFs with similar gene expression profiles to *UCP1* (Morpheus Interface). **2I** Comparison of the
918 nucleotide sequence of the HDF1-class TF RE at the human *UCP1* promoter -690 bp with sequences from other
919 species using the Multiz Alignment of 100 Vertebrates database in the UCSC genome browser (red box). In the
920 vicinity of HDF1 RE the TF RE sequences of RFX5, MYOD1 and BCL6 identified by EPD are highlighted with brown,
921 green and blue boxes, respectively. **2J** Heatmap shows the gene expression profile of 12 TFs with RE common in
922 the human, mouse and rat *UCP1* promoter in adipocyte samples of cervical origin. For comparison, *UCP1* is also
923 shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix identifies TFs with similar gene
924 expression profiles to *UCP1* (Morpheus Interface). Adipocytes were differentiated using a white (W) or brown
925 (B) differentiation protocol (Tóth et al. 2020).

926 **Figure 3. 3A** The Venn diagrams show the number of TFs with putative RE in the human *UCP1* promoter (-5000
927 – +1000 BP) based on EPD, ChiPBase and TFlink databases. TFs commonly found in the databases are named in
928 the side panels (Venn program in R environment). **3B** The heatmap shows the gene expression profiles of 11 TFs
929 in human adipocytes of cervical origin that have been predicted or experimentally shown to have RE in the human
930 *UCP1* promoter based on EPD, ChiPBase and TFlink interfaces. For comparison, *UCP1* is also shown. Hierarchical
931 cluster analysis based on the Spearman Rank similarity matrix identifies TFs with similar gene expression profiles
932 to *UCP1* (Morpheus interface).

933 **Figure 4.** This figure provides a comprehensive overview of the transcription factor response elements identified
934 and filtered by our study within the human *UCP1* promoter. These REs exhibit gene or species specificity.
935 Additionally, several previously reported TF REs, which did not display such specificity, are included for reference.
936 The figure highlights specific promoter segments in the human *UCP1* promoter that are absent in the mouse and
937 rat *UCP1* promoters (indicated by the red line). The yellow line represents repeat elements identified by
938 RepeatMasker. Numerical annotations within the promoter denote base pair positions relative to the
939 transcription start site (TSS). For clarity, a list of TF abbreviations is provided in the "List of abbreviations" section
940 at the end of the article. H: human, M: mouse, R: rat. HDF: homeodomain family, bHLH: basic helix-loop-helix
941 family (PowerPoint).

942 **Supplementary Figures Legends**

943 **Supplementary Figure 1.** Heatmaps show the pairwise genetic distances of the nucleotid sequences of the *UCP1*,
944 *UCP2* and *UCP3* genes' regulatory and coding regions in humans, mice and rats (pheatmap program in R
945 environment). **1B** The phylogenetic tree shows the presence and the absence (red box) of the human UCP1
946 promoter DNA fragment spans between -2083 bp and -3502 bp (GRCh37/hg19: Chr4 141,492,160-141,493,617)
947 in *UCP1* promoter of different vertebrate species according to the Multiz Alignment of 100 Vertebrates database
948 in the UCSC genome browser. No information is available for the three genera marked with blue question marks.
949 **1C – 1G** Alignments of the *UCP1* promoter sequences of different species highlight the palindromic nucleotide
950 sequences of the beginning and end of the DNA fragments (1419 bp and 658 bp) missing from the mouse and rat
951 *UCP1* promoter according to the Multiz Alignment of 100 Vertebrates database in the UCSC genome browser.
952 **1H** Graphical summary showing the human genomic occurrence of the *UCP1* promoter DNA fragment (1419 bp)
953 absent from the mice and rat *UCP1* promoter (https://blast.ncbi.nlm.nih.gov/). The graphic is an overview of the
954 database sequences aligned to the query sequence. These are represented by horizontal bars colored coded by
955 score and showing the extent of the alignment on the query sequence. Separate aligned regions on the same
956 database sequence are connected by a thin grey line.

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

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

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

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

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

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

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

995 **Supplementary Figure 9.** Alignment of *UCP1* promoter sequences from different species at different positions
996 on the ConTra v3 platform. Human *UCP1*-specific TF REs are highlighted in different colors, and nearby TF REs
997 are outlined in a blue box. **9A** shows the presence of the HOXA5 TF RE, unique to the human *UCP1* promoter at
998 -4785 bp, and the nearby HIC1 RE at -4797 (EPD interface data). **9B** shows the presence of the DLX2 TF RE, unique
999 to the human *UCP1* promoter at -2478 bp, and the nearby TEAD1-4 RE at -2496 (EPD interface data). **9C** shows
1000 the presence of the HOXA2 TF RE, unique to the human *UCP1* promoter at -3959 bp, and the nearby NFKB RE at
1001 -3978, and ZNF410 and ZNF282 at - 3948 bp (EPD interface data). **9D** shows the presence of the NFIL3 TF RE,
1002 unique to the human *UCP1* promoter at -881 bp, and **9E** the nearby REL RE at -902 bp (EPD interface data). **9F**
1003 shows the presence of the NFIL3 TF RE at -881 bp and the nearby REL RE at -902 bp across species in UCSC
1004 genome browser (ConTra v3 and EPD interface data).

1005 **Supplementary Figure 10.** Alignment of the *UCP1* promoter sequences of different species **10A** at -4715 bp
1006 position and **10B** at -2913 bp position to investigate the presence of POU6F1 TF-binding sequences in human,
1007 mouse and rat promoter according to the Multiz Alignment of 100 Vertebrates database in the UCSC genome
1008 browser.

1009 **Supplementary Figure 11**. **11A** Heatmap showing the relative gene expression profile of common TFs identified
1010 in the promoters of human *UCP1*, *UCP2* and *UCP3* around -264 bp missing the binding sequence in the promoters
1011 of mouse and rat orthologous genes using EPD interface (p<1e-4) in cervical derived adipocyte samples. For
1012 comparison, *UCP1* is also shown. Hierarchical cluster analysis based on the Spearman Rank similarity matrix
1013 identifies TFs with similar gene expression profiles to *UCP1*. (Morpheus Interface) **11B** Phylogenetic footprinting:
1014 the green color/red box highlights the HIF1A binding site in the *UCP1* promoter at -264 (ConTra v3 interface).

1015 **11C** Phylogenetic footprinting: Blue color/green box highlights the EGR1 RE at -257 and and red box the HIF1A
1016 RE at -264 bp in the *UCP1* promoter (ConTra v3 interface). **11D** Phylogenetic footprinting: the green color/purple
1017 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
1018 at -264 bp in the *UCP1* promoter (ConTra v3 interface). Adipocytes were differentiated using a white (W) or
1019 brown (B) differentiation protocol.

1020 **Table legends**

1021 **Table 1.** Results of pairwise structure alignments, measures describing the extent of structural similarity to
1022 human UCP1 of the human, mouse and rat UCP1, UCP2 and UCP3 proteins. The structure alignments are based
1023 on protein structures predicted by AlphaFold 2 and aligned by RCSP-PBD. See Methodology for the interpretation
1024 of the measures.

1025 **Table 2.** Tables 2A and 2B summarize our data on TFs with an RE in the human *UCP1* promoter identified by EPD
1026 (p<1e-4) and filtered according to specific criteria. These include the position of the RE relative to the TSS, their
1027 gene expression pattern relative to *UCP1*, their gene expression levels measured in neck-derived preadipocytes
1028 and *in vitro* differentiated adipocytes, their ChipSeq data from the ChiPBase and TFlink databases, their
1029 classification into TF class and family based on JASPAR databases, their PWM, their involvement in the adipocyte
1030 phenotype formation pathway according to the AdipoNET interface, EPD TF RE search results at p<1e-5
1031 significance level in the human *UCP1* promoter. Bold highlights TF or value outstanding in the related category.
1032 **Table 2A** highlights the 13 putative regulatory TFs that possess REs only in the human *UCP1* promoter and are
1033 absent from the other *UCP* genes examined. **Table 2B** summarizes our data for 17 TFs belonging to the
1034 homeodomain Family 1 and 2 TF classes whose RE matches the human-specific binding sequence identified in
1035 the *UCP1* promoter ("TCTAATTAGA") the expression of which is correlated with *UCP1* or for which we do not
1036 have gene expression data and therefore cannot exclude them. GE: Gene Expression, Corr.: Correlation.

1037 **Table 3.** Tables 3A and 3B summarize our data on TFs with RE in the human *UCP1* promoter identified using EPD
1038 (p<1e-4) and filtered according to specific criteria. These include the position of their RE relative to the TSS, their
1039 gene expression pattern relative to *UCP1*, their gene expression levels measured in neck-derived preadipocytes
1040 and in vitro differentiated adipocytes, and their ChipSeq. data from the ChiPBase and TFlink databases, their
1041 classification into TF class and family based on JASPAR databases, their PWM, their involvement in the adipocyte
1042 phenotype formation pathway according to the AdipoNET interface, EPD TF RE search results at p<1e-5
1043 significance level in the human *UCP1* promoter. Bold highlights TF or value outstanding in the related category.
1044 **3A** The table summarizes the 12 TFs with RE found to be common in human, mouse and rat *UCP1.* **3B.** The table
1045 summarizes the data of TFs with RE (EPD p<1e-4) in the human *UCP1* promoter that also has ChipSeq data from
1046 both ChiPBase and TFlink databases. GE: Gene Expression, Corr.: Correlation.

1047 **Supplementary Table Legends**

1048 **Supplementary Table 1** The table lists TFs identified through RE analysis in the promoters of human, mouse, and
1049 rat UCP1, UCP2, and UCP3 genes, categorized by species and genes, employing the EPD interface (p<1e-4) for
1050 identification. Subgroup categorization is based on the Venn diagram representation presented in Figure 1D.

1051 **Supplementary Table 2** The table presents the list of TFs identified through RE analysis in the promoters of
1052 human, mouse, and rat UCP1, UCP2, and UCP3 genes. The listing includes the TFs' relative positions to the TSS,
1053 categorized by species and genes, utilizing the EPD interface (p<1e-4, p<1e-5).

1054 **Supplementary Table 3** The table presents correlation coefficients depicting the associations between the gene
1055 expression profile of human UCP1 and the TFs identified within human UCP1 promoters using putative REs. The
1056 data utilized for this analysis were derived from RNA sequencing of human adipocytes obtained from the cervical
1057 region.

1058 **Supplementary Table 4.** The table shows the TF abbreviations and the number of abstracts where they occur
1059 together with the keyword UCP1 in publications in the PubMed database.

1060 **DECLARATIONS**

1061 **Ethics approval and consent to participate**

1062 Not Applicable

1063 **Consent for Publication**

1064 Not Applicable

1065 **Availability of data and material**

1066 RNAseq data have been deposited to Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra)
1067 under accession number PRJNA607438. We automated the retrieval of EPD data and the screening of abstracts
1068 from the PubMed database using custom scripts that were made publicly available on GitHub
1069 (https://github.com/DEpt-metagenom/TFRE-in-UCP1-promoter). The main and supplementary figures and
1070 tables are available in the Mendeley database: https://data.mendeley.com/preview/z5b6stz3zx?a=828a4650-
1071 5239-487a-b17f-43c667b02454

1072 **Competing interests**

1073 The authors declare no conflict of interest.

1074 **Funding**

1075 During this research BBT, LL, EV and GP was funded by the European Union and the European Regional
1076 Development Fund (GINOP-2.3.4-15-2016-00002 „A felsőoktatás és az ipar együttműködése az
1077 egészségiparban"), GH was supported by the European Union and the European Regional Development Fund
1078 (GINOP-2.3.4-15-2020-00008 „Komplex Egészségipari Multidiszciplináris Kompetencia Központ kialakítása a
1079 Debreceni Egyetemen új innovatív termékek és technológiák fejlesztése érdekében"), and we are thankful for
1080 this. BBT was also supported by the University of Debrecen Scientific Research Bridging Fund (DETKA). The
1081 funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and
1082 in writing the manuscript.

1083 **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. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.

1156 20. Collins S, Cao W, Robidoux J. Learning new tricks from old dogs: beta-adrenergic receptors teach new lessons
1157 on firing up adipose tissue metabolism. Mol Endocrinol. 2004 Sep;18(9):2123-31. doi: 10.1210/me.2004-
1158 0193. Epub 2004 Jul 8. PMID: 15243132.
1159 21. Crichton PG, Lee Y, Kunji ER. The molecular features of uncoupling protein 1 support a conventional
1160 mitochondrial carrier-like mechanism. Biochimie. 2017 Mar;134:35-50. doi: 10.1016/j.biochi.2016.12.016.
1161 Epub 2017 Jan 3. PMID: 28057583; PMCID: PMC5395090.
1162 22. del Mar Gonzalez-Barroso M, Pecqueur C, Gelly C, Sanchis D, Alves-Guerra MC, Bouillaud F, Ricquier D,
1163 Cassard-Doulcier AM. Transcriptional activation of the human ucp1 gene in a rodent cell line. Synergism of
1164 retinoids, isoproterenol, and thiazolidinedione is mediated by a multipartite response element. J Biol Chem.
1165 2000 Oct 13;275(41):31722-32. doi: 10.1074/jbc.M001678200. PMID: 10921912.
1166 23. Dreos R, Ambrosini G, Groux R, Cavin Périer R, Bucher P. The eukaryotic promoter database in its 30th year:
1167 focus on non-vertebrate organisms. Nucleic Acids Res. 2017;45(D1):D51–
1168 5. https://doi.org/10.1093/nar/gkw1069 Epub 2016 Nov 28. PMID: 27899657; PMCID: PMC5210552.
1169 24. Dror I, Zhou T, Mandel-Gutfreund Y, Rohs R. Covariation between homeodomain transcription factors and
1170 the shape of their DNA binding sites. Nucleic Acids Res. 2014 Jan;42(1):430-41. doi: 10.1093/nar/gkt862.
1171 Epub 2013 Sep 27. PMID: 24078250; PMCID: PMC3874178.
1172 25. Emera D, Wagner GP. Transformation of a transposon into a derived prolactin promoter with function during
1173 human pregnancy. Proc Natl Acad Sci U S A. 2012 Jul 10;109(28):11246-51. doi: 10.1073/pnas.1118566109.
1174 Epub 2012 Jun 25. PMID: 22733751; PMCID: PMC3396485.
1175 26. Enerbäck S, Jacobsson A, Simpson EM, Guerra C, Yamashita H, Harper ME, Kozak LP. Mice lacking
1176 mitochondrial uncoupling protein are cold-sensitive but not obese. Nature. 1997 May 1;387(6628):90-4. doi:
1177 10.1038/387090a0. PMID: 9139827.
1178 27. Feldmann HM, Golozoubova V, Cannon B, Nedergaard J. UCP1 ablation induces obesity and abolishes diet-
1179 induced thermogenesis in mice exempt from thermal stress by living at thermoneutrality. Cell Metab. 2009
1180 Feb;9(2):203-9. doi: 10.1016/j.cmet.2008.12.014. PMID: 19187776.
1181 28. Foulkes NS, Borrelli E, Sassone-Corsi P. CREM gene: use of alternative DNA-binding domains generates
1182 multiple antagonists of cAMP-induced transcription. Cell. 1991 Feb 22;64(4):739-49. doi: 10.1016/0092-
1183 8674(91)90503-q. PMID: 1847666.
1184 29. Gaudry MJ, Campbell KL. Evolution of *UCP1* Transcriptional Regulatory Elements Across the Mammalian
1185 Phylogeny. Front Physiol. 2017 Sep 20;8:670. doi: 10.3389/fphys.2017.00670. PMID: 28979209; PMCID:
1186 PMC5611445.
1187 30. Gaudry MJ, Jastroch M, Treberg JR, Hofreiter M, Paijmans JLA, Starrett J, et al. Inactivation of thermogenic
1188 UCP1 as a historical contingency in multiple placental mammal clades. *Sci Adv.* (2017) 3:e1602878
1189 10.1126/sciadv.1602878
1190 31. Gehring WJ, Affolter M, Bürglin T. Homeodomain proteins. Annu Rev Biochem. 1994;63:487-526. doi:
1191 10.1146/annurev.bi.63.070194.002415. PMID: 7979246.
1192 32. Gupta A, Efthymiou V, Kodani SD, Shamsi F, Patti ME, Tseng YH, Streets A. Mapping the transcriptional
1193 landscape of human white and brown adipogenesis using single-nuclei RNA-seq. Mol Metab. 2023

Aug;74:101746. doi: 10.1016/j.molmet.2023.101746. Epub 2023 Jun 5. PMID: 37286033; PMCID: PMC10338377.

33. Harms MJ, Lim HW, Ho Y, Shapira SN, Ishibashi J, Rajakumari S, Steger DJ, Lazar MA, Won KJ, Seale P. PRDM16 binds MED1 and controls chromatin architecture to determine a brown fat transcriptional program. Genes Dev. 2015 Feb 1;29(3):298-307. doi: 10.1101/gad.252734.114. PMID: 25644604; PMCID: PMC4318146.
34. Héder, M., Rigó, E., Medgyesi, D., Lovas, R., Tenczer, S., Török, F., Farkas, A., Emődi, M., Kadlecsik, J., Mező, G., Pintér, Á., & Kacsuk, P. (2022). The Past, Present and Future of the ELKH Cloud. Információs Társadalom, 22(2), 128.
35. Hilton C, Karpe F, Pinnick KE. Role of developmental transcription factors in white, brown and beige adipose tissues. Biochim Biophys Acta. 2015 May;1851(5):686-96. doi: 10.1016/j.bbalip.2015.02.003. Epub 2015 Feb 8. PMID: 25668679.
36. Hoang T, Smith MD, Jelokhani-Niaraki M. Expression, folding, and proton transport activity of human uncoupling protein-1 (UCP1) in lipid membranes: evidence for associated functional forms. J Biol Chem. 2013 Dec 20;288(51):36244-58. doi: 10.1074/jbc.M113.509935. Epub 2013 Nov 6. PMID: 24196960; PMCID: PMC3868739.
37. Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, Shah N, Suzuki PH, Shrikumar A, Afek A, Greenleaf WJ, Gordân R, Zeitlinger J, Kundaje A, Fordyce PM. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. Science. 2023 Sep 22;381(6664):eadd1250. doi: 10.1126/science.add1250. Epub 2023 Sep 22. PMID: 37733848.
38. Hou L, Shi J, Cao L, Xu G, Hu C, Wang C. Pig has no uncoupling protein 1. Biochem Biophys Res Commun. 2017 Jun 10;487(4):795-800. doi: 10.1016/j.bbrc.2017.04.118. Epub 2017 Apr 23. PMID: 28442343.
39. Hughes J, Criscuolo F. Evolutionary history of the UCP gene family: gene duplication and selection. BMC Evol Biol. 2008 Nov 3;8:306. doi: 10.1186/1471-2148-8-306. PMID: 18980678; PMCID: PMC2584656.
40. Hughes DA, Jastroch M, Stoneking M, Klingenspor M. Molecular evolution of UCP1 and the evolutionary history of mammalian non-shivering thermogenesis. BMC Evol Biol. 2009 Jan 7;9:4. doi: 10.1186/1471-2148-9-4. PMID: 19128480; PMCID: PMC2627829.S
41. Inagaki T, Sakai J, Kajimura S. Transcriptional and epigenetic control of brown and beige adipose cell fate and function. Nat Rev Mol Cell Biol. 2016 Aug;17(8):480-95. doi: 10.1038/nrm.2016.62. Epub 2016 Jun 2. Erratum in: Nat Rev Mol Cell Biol. 2017 Aug;18(8):527. PMID: 27251423; PMCID: PMC4956538.
42. Jastroch M, Oelkrug R, Keipert S. Insights into brown adipose tissue evolution and function from non-model organisms. *J Exp Biol.* (2018) 221(Pt Suppl 1):jeb169425. 10.1242/jeb.169425.
43. Jastroch M, Withers KW, Taudien S, Frappell PB, Helwig M, Fromme T, Hirschberg V, Heldmaier G, McAllan BM, Firth BT, Burmester T, Platzer M, Klingenspor M. Marsupial uncoupling protein 1 sheds light on the evolution of mammalian nonshivering thermogenesis. Physiol Genomics. 2008 Jan 17;32(2):161-9. doi: 10.1152/physiolgenomics.00183.2007. Epub 2007 Oct 30. PMID: 17971503.
44. Jastroch M, Wuertz S, Kloas W, Klingenspor M. Uncoupling protein 1 in fish uncovers an ancient evolutionary history of mammalian nonshivering thermogenesis. Physiol Genomics. 2005 Jul 14;22(2):150-6. doi: 10.1152/physiolgenomics.00070.2005. Epub 2005 May 10. PMID: 15886331.

45. Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J. DNA-binding specificities of human transcription factors. Cell. 2013 Jan 17;152(1-2):327-39. doi: 10.1016/j.cell.2012.12.009. PMID: 23332764.
46. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
47. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 2000 Sep;16(9):418-20. doi: 10.1016/s0168-9525(00)02093-x. PMID: 10973072.
48. Kato H, Ario T, Kishida T, Tadano M, Osawa S, Maeda Y, Takakura H, Izawa T. Homeobox A5 and C10 genes modulate adaptation of brown adipose tissue during exercise training in juvenile rats. Exp Physiol. 2021 Feb;106(2):463-474. doi: 10.1113/EP089114. Epub 2021 Jan 5. PMID: 33369800.
49. Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010
50. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980 Dec;16(2):111-20. doi: 10.1007/BF01731581. PMID: 7463489.
51. Klingenspor M, Fromme T, Hughes DA Jr, Manzke L, Polymeropoulos E, Riemann T, Trzcionka M, Hirschberg V, Jastroch M. An ancient look at UCP1. Biochim Biophys Acta. 2008 Jul-Aug;1777(7-8):637-41. doi: 10.1016/j.bbabio.2008.03.006. Epub 2008 Mar 18. PMID: 18396149.
52. Kottke FJ, Phalen JS, et al. Effect of hypoxia upon temperature regulation of mice, dogs, and man. Am J Physiol. 1948 Apr;153(1):10-5. doi: 10.1152/ajplegacy.1948.153.1.10. PMID: 18867027.
53. Kreft L, Soete A, Hulpiau P, Botzki A, Saeys Y, De Bleser P. ConTra v3: a tool to identify transcription factor binding sites across species, update 2017. Nucleic Acids Res. 2017 Jul 3;45(W1):W490-W494. doi: 10.1093/nar/gkx376. PMID: 28472390; PMCID: PMC5570180.
54. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010 Jul;42(7):631-4. doi: 10.1038/ng.600. Epub 2010 Jun 6. PMID: 20526341.
55. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The Human Transcription Factors. Cell. 2018 Feb 8;172(4):650-665. doi: 10.1016/j.cell.2018.01.029. Erratum in: Cell. 2018 Oct 4;175(2):598-599. PMID: 29425488.
56. Lau HH, Krentz NAJ, Abaitua F, Perez-Alcantara M, Chan JW, Ajeian J, Ghosh S, Lee Y, Yang J, Thaman S, Champon B, Sun H, Jha A, Hoon S, Tan NS, Gardner DS, Kao SL, Tai ES, Gloyn AL, Teo AKK. PAX4 loss of function increases diabetes risk by altering human pancreatic endocrine cell development. Nat Commun. 2023 Sep 30;14(1):6119. doi: 10.1038/s41467-023-41860-z. PMID: 37777536; PMCID: PMC10542369.
57. Laurila PP, Soronen J, Kooijman S, Forsström S, Boon MR, Surakka I, Kaiharju E, Coomans CP, Van Den Berg SA, Autio A, Sarin AP, Kettunen J, Tikkanen E, Manninen T, Metso J, Silvennoinen R, Merikanto K, Ruuth M, Perttilä J, Mäkelä A, Isomi A, Tuomainen AM, Tikka A, Ramadan UA, Seppälä I, Lehtimäki T, Eriksson J,

1270 Havulinna A, Jula A, Karhunen PJ, Salomaa V, Perola M, Ehnholm C, Lee-Rueckert M, Van Eck M, Roivainen
1271 A, Taskinen MR, Peltonen L, Mervaala E, Jalanko A, Hohtola E, Olkkonen VM, Ripatti S, Kovanen PT, Rensen
1272 PC, Suomalainen A, Jauhiainen M. USF1 deficiency activates brown adipose tissue and improves
1273 cardiometabolic health. Sci Transl Med. 2016 Jan 27;8(323):323ra13. doi: 10.1126/scitranslmed.aad0015.
1274 PMID: 26819196.
1275 58. Li L, Li Q, Willis LG, Erlandson M, Theilmann DA, Donly C. Complete comparative genomic analysis of two
1276 field isolates of Mamestra configurata nucleopolyhedrovirus-A. J Gen Virol. 2005 Jan;86(Pt 1):91-105. doi:
1277 10.1099/vir.0.80488-0. PMID: 15604435.
1278 59. Li Y, Schwalie PC, Bast-Habersbrunner A, Mocek S, Russeil J, Fromme T, Deplancke B, Klingenspor M.
1279 Systems-Genetics-Based Inference of a Core Regulatory Network Underlying White Fat Browning. Cell Rep.
1280 2019 Dec 17;29(12):4099-4113.e5. doi: 10.1016/j.celrep.2019.11.053. PMID: 31851936.
1281 60. Linheiro RS, Bergman CM. Testing the palindromic target site model for DNA transposon insertion using the
1282 Drosophila melanogaster P-element. Nucleic Acids Res. 2008 Nov;36(19):6199-208. doi:
1283 10.1093/nar/gkn563. Epub 2008 Oct 1. PMID: 18829720; PMCID: PMC2577343.
1284 61. Liu Y, Maekawa T, Yoshida K, Muratani M, Chatton B, Ishii S. The Transcription Factor ATF7 Controls
1285 Adipocyte Differentiation and Thermogenic Gene Programming. iScience. 2019 Mar 29;13:98-112. doi:
1286 10.1016/j.isci.2019.02.013. Epub 2019 Feb 18. PMID: 30826729; PMCID: PMC6402263.
1287 62. Lu WH, Chang YM, Huang YS. Alternative Polyadenylation and Differential Regulation of *Ucp1*: Implications
1288 for Brown Adipose Tissue Thermogenesis Across Species. Front Pediatr. 2021 Feb 9;8:612279. doi:
1289 10.3389/fped.2020.612279. PMID: 33634052; PMCID: PMC7899972.
1290 63. McGaugh S, Schwartz TS. Here and there, but not everywhere: repeated loss of uncoupling protein 1 in
1291 amniotes. Biol Lett. 2017 Jan;13(1):20160749. doi: 10.1098/rsbl.2016.0749. PMID: 28052937; PMCID:
1292 PMC5310580.
1293 64. Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R.
1294 (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.
1295 Molecular Biology and Evolution, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015
1296 65. Mohsen-Kanson T, Hafner AL, Wdziekonski B, Takashima Y, Villageois P, Carrière A, Svensson M, Bagnis C,
1297 Chignon-Sicard B, Svensson PA, Casteilla L, Smith A, Dani C. Differentiation of human induced pluripotent
1298 stem cells into brown and white adipocytes: role of Pax3. Stem Cells. 2014 Jun;32(6):1459-67. doi:
1299 10.1002/stem.1607. PMID: 24302443.
1300 66. Nedergaard J, Bengtsson T, Cannon B. Unexpected evidence for active brown adipose tissue in adult humans.
1301 Am J Physiol Endocrinol Metab. 2007 Aug;293(2):E444-52. doi: 10.1152/ajpendo.00691.2006. Epub 2007
1302 May 1. PMID: 17473055.
1303 67. Nedergaard J, Cannon B. The 'novel' 'uncoupling' proteins UCP2 and UCP3: what do they really do? Pros and
1304 cons for suggested functions. Exp Physiol. 2003 Jan;88(1):65-84. doi: 10.1113/eph8802502. PMID:
1305 12525856.
1306 68. Nedergaard J, Cannon B. The browning of white adipose tissue: some burning issues. Cell Metab. 2014 Sep
1307 2;20(3):396-407. doi: 10.1016/j.cmet.2014.07.005. Epub 2014 Aug 7. PMID: 25127354.

1308 69. Nedergaard J, Golozoubova V, Matthias A, Asadi A, Jacobsson A, Cannon B. UCP1: the only protein able to
1309 mediate adaptive non-shivering thermogenesis and metabolic inefficiency. Biochim Biophys Acta. 2001 Mar
1310 1;1504(1):82-106. doi: 10.1016/s0005-2728(00)00247-4. PMID: 11239487.
1311 70. Nicholls DG. Mitochondrial proton leaks and uncoupling proteins. Biochim Biophys Acta Bioenerg. 2021 Jul
1312 1;1862(7):148428. doi: 10.1016/j.bbabio.2021.148428. Epub 2021 Mar 31. PMID: 33798544.
1313 71. Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach.
1314 Bioinformatics, 26(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696
1315 72. Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary
1316 analyses in R. Bioinformatics, 35(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633
1317 73. Park MJ, Kim HY, Kim K, Cheong J. Homeodomain transcription factor CDX1 is required for the transcriptional
1318 induction of PPARgamma in intestinal cell differentiation. FEBS Lett. 2009 Jan 5;583(1):29-35. doi:
1319 10.1016/j.febslet.2008.11.030. Epub 2008 Dec 4. PMID: 19059241.
1320 74. Périer RC, Praz V, Junier T, Bonnard C, Bucher P. The eukaryotic promoter database (EPD). Nucleic Acids Res.
1321 2000 Jan 1;28(1):302-3. doi: 10.1093/nar/28.1.302. PMID: 10592254; PMCID: PMC102462.
1322 75. Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. Nucleic Acids Res. 2001 Sep
1323 15;29(18):3705-27. doi: 10.1093/nar/29.18.3705. PMID: 11557805; PMCID: PMC55916.
1324 76. R Foundation for Statistical Computing. (n.d.). R: A language and environment for statistical computing.
1325 Vienna, Austria.
1326 77. Razzoli M, Bartolomucci A. The Dichotomous Effect of Chronic Stress on Obesity. Trends Endocrinol Metab.
1327 2016 Jul;27(7):504-515. doi: 10.1016/j.tem.2016.04.007. Epub 2016 May 5. PMID: 27162125; PMCID:
1328 PMC4912918.
1329 78. Ricquier D. Uncoupling protein 1 of brown adipocytes, the only uncoupler: a historical perspective. Front
1330 Endocrinol (Lausanne). 2011 Dec 28;2:85. doi: 10.3389/fendo.2011.00085. PMID: 22649389; PMCID:
1331 PMC3355862.
1332 79. Rim JS, Kozak LP. Regulatory motifs for CREB-binding protein and Nfe2l2 transcription factors in the
1333 upstream enhancer of the mitochondrial uncoupling protein 1 gene. J Biol Chem. 2002 Sep
1334 13;277(37):34589-600. doi: 10.1074/jbc.M108866200. Epub 2002 Jun 25. PMID: 12084707.
1335 80. Rosen ED, Walkey CJ, Puigserver P, Spiegelman BM. Transcriptional regulation of adipogenesis. Genes Dev.
1336 2000 Jun 1;14(11):1293-307. PMID: 10837022.
1337 81. Rousset S, Alves-Guerra MC, Mozo J, Miroux B, Cassard-Doulcier AM, Bouillaud F, Ricquier D. The biology of
1338 mitochondrial uncoupling proteins. Diabetes. 2004 Feb;53 Suppl 1:S130-5. doi:
1339 10.2337/diabetes.53.2007.s130. PMID: 14749278.
1340 82. Saito S, Saito CT, Shingai R. Adaptive evolution of the uncoupling protein 1 gene contributed to the
1341 acquisition of novel nonshivering thermogenesis in ancestral eutherian mammals. *Gene*. (2008) 408:37–44.
1342 10.1016/j.gene.2007.10.018.
1343 83. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. JASPAR: an open-access database for
1344 eukaryotic transcription factor binding profiles. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D91-4. doi:
1345 10.1093/nar/gkh012. PMID: 14681366; PMCID: PMC308747.

1346 84. Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. Bioinformatics, 27(4), 592–593.
1347 https://doi.org/10.1093/bioinformatics/btq706
1348 85. Scarpato M, Angelini C, Cocca E, Pallotta MM, Morescalchi MA, Capriglione T. Short interspersed DNA
1349 elements and miRNAs: a novel hidden gene regulation layer in zebrafish? Chromosome Res. 2015
1350 Sep;23(3):533-44. doi: 10.1007/s10577-015-9484-6. PMID: 26363800.
1351 86. Seale P., Kajimura S., Yang W., et al. Transcriptional control of brown fat determination by PRDM16. *Cell*
1352 *Metabolism.* 2007;6(1):38–54. doi: 10.1016/j.cmet.2007.06.001.
1353 87. Shapira SN, Seale P. Transcriptional Control of Brown and Beige Fat Development and Function. Obesity
1354 (Silver Spring). 2019 Jan;27(1):13-21. doi: 10.1002/oby.22334. PMID: 30569639; PMCID: PMC6309799.
1355 88. 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,
1356 Li X, Chen Y, Cao Y. Brown-fat-mediated tumour suppression by cold-altered global metabolism. Nature.
1357 2022 Aug;608(7922):421-428. doi: 10.1038/s41586-022-05030-3. Epub 2022 Aug 3. PMID: 35922508;
1358 PMCID: PMC9365697.
1359 89. Singh S, Rajput YS, Barui AK, Sharma R, Datta TK. Fat accumulation in differentiated brown adipocytes is
1360 linked with expression of Hox genes. Gene Expr Patterns. 2016 Mar;20(2):99-105. doi:
1361 10.1016/j.gep.2016.01.002. Epub 2016 Jan 26. PMID: 26820751.
1362 90. Singh S, Rajput YS, Barui AK, Sharma R, Grover S. Expression of developmental genes in brown fat cells grown
1363 in vitro is linked with lipid accumulation. In Vitro Cell Dev Biol Anim. 2015 Nov;51(10):1003-11. doi:
1364 10.1007/s11626-015-9930-y. Epub 2015 Jul 7. PMID: 26148884.
1365 91. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS.
1366 Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell. 2011 Dec
1367 9;147(6):1270-82. doi: 10.1016/j.cell.2011.10.053. PMID: 22153072; PMCID: PMC3319069.
1368 92. Su T, Liu P, Ti X, Wu S, Xue X, Wang Z, Dioum E, Zhang Q. HIF1α, EGR1 and SP1 co-regulate the erythropoietin
1369 receptor expression under hypoxia: an essential role in the growth of non-small cell lung cancer cells. Cell
1370 Commun Signal. 2019 Nov 21;17(1):152. doi: 10.1186/s12964-019-0458-8. PMID: 31752873; PMCID:
1371 PMC6869211.
1372 93. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent
1373 stem cells from adult human fibroblasts by defined factors. Cell. 2007 Nov 30;131(5):861-72. doi:
1374 10.1016/j.cell.2007.11.019. PMID: 18035408.
1375 94. Tattersall GJ, Milsom WK. Hypoxia reduces the hypothalamic thermogenic threshold and thermosensitivity.
1376 J Physiol. 2009 Nov 1;587(Pt 21):5259-74. doi: 10.1113/jphysiol.2009.175828. Epub 2009 Sep 21. PMID:
1377 19770191; PMCID: PMC2790263.
1378 95. Tayanloo-Beik A, Nikkhah A, Alaei S, Goodarzi P, Rezaei-Tavirani M, Mafi AR, Larijani B, Shouroki FF, Arjmand
1379 B. Brown adipose tissue and alzheimer's disease. Metab Brain Dis. 2023 Jan;38(1):91-107. doi:
1380 10.1007/s11011-022-01097-z. Epub 2022 Nov 2. PMID: 36322277.
1381 96. Thompson PJ, Macfarlan TS, Lorincz MC. Long Terminal Repeats: From Parasitic Elements to Building Blocks
1382 of the Transcriptional Regulatory Repertoire. Mol Cell. 2016 Jun 2;62(5):766-76. doi:
1383 10.1016/j.molcel.2016.03.029. PMID: 27259207; PMCID: PMC4910160.

97. 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.
98. 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.
99. 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.
100. 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.
101. 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.
102. 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.
103. 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.
104. 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.
105. 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.
106. 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.
107. 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.

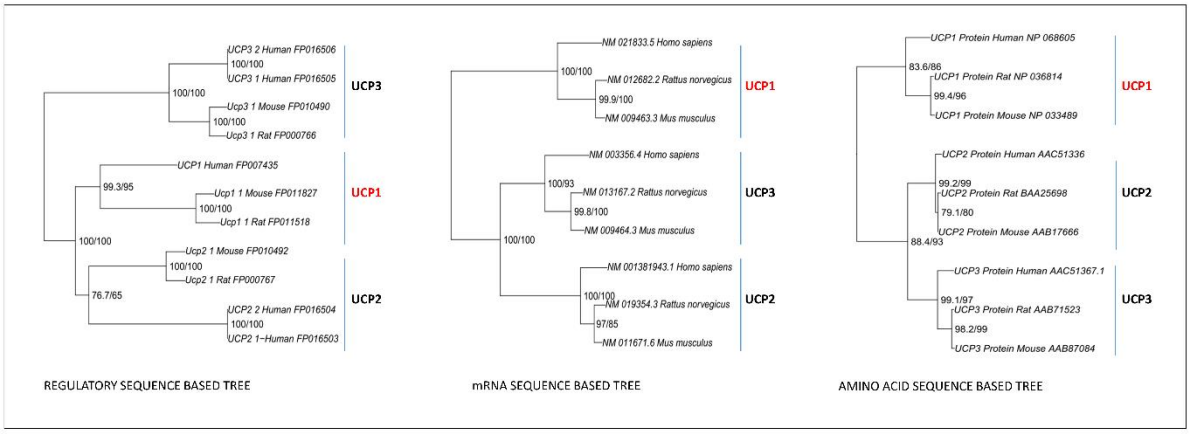
1421 108.Xu L, Xu C, Zhou S, Liu X, Wang J, Liu X, Qian S, Xin Y, Gao Y, Zhu Y, Tang X. PAX4 promotes PDX1-induced
1422 differentiation of mesenchymal stem cells into insulin-secreting cells. Am J Transl Res. 2017 Mar 15;9(3):874-
1423 886. PMID: 28386318; PMCID: PMC5375983.
1424 109.Xue B, Coulter A, Rim JS, Koza RA, Kozak LP. Transcriptional synergy and the regulation of Ucp1 during brown
1425 adipocyte induction in white fat depots. Mol Cell Biol. 2005 Sep;25(18):8311-22. doi:
1426 10.1128/MCB.25.18.8311-8322.2005. PMID: 16135818; PMCID: PMC1234324.
1427 110.Yevshin I, Sharipov R, Valeev T, Kel A, Kolpakov F. GTRD: a database of transcription factor binding sites
1428 identified by ChIP-seq experiments. Nucleic Acids Res. 2017 Jan 4;45(D1):D61-D67. doi:
1429 10.1093/nar/gkw951. Epub 2016 Oct 24. PMID: 27924024; PMCID: PMC5210645.
1430 111.Zhang X, Wang C, Liu M; 77-OR: Targeting JunB Induces a Shift from Low- to High-Thermogenic Adipocytes
1431 and Ameliorates Diet-Induced Insulin Resistance. *Diabetes* 20 June 2023; 72 (Supplement_1): 77–OR. doi:
1432 10.2337/db23-77-OR
1433 112.Zhang C, Weng Y, Shi F, Jin W. The Engrailed-1 Gene Stimulates Brown Adipogenesis. Stem Cells Int.
1434 2016;2016:7369491. doi: 10.1155/2016/7369491. Epub 2016 Apr 11. PMID: 27148369; PMCID:
1435 PMC4842372.
1436 113.Zhang J, Zhang Y, Sun T, Guo F, Huang S, Chandalia M, Abate N, Fan D, Xin HB, Chen YE, Fu M. Dietary obesity-
1437 induced Egr-1 in adipocytes facilitates energy storage via suppression of FOXC2. Sci Rep. 2013;3:1476. doi:
1438 10.1038/srep01476. PMID: 23502673; PMCID: PMC3600596.

1439 **FIGURES**

1440 **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.

1441

1442

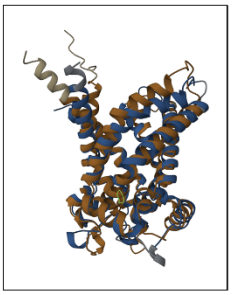
1443 **Figure 2.**

2A

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

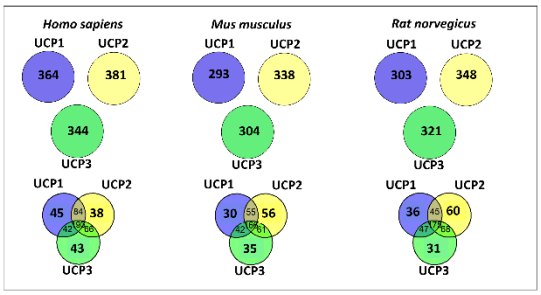
2B

H-UCP1 aligned with H-ANT1



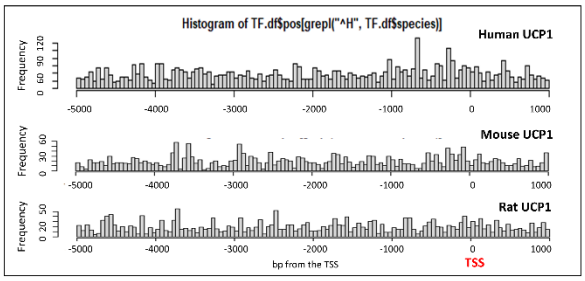
2C

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



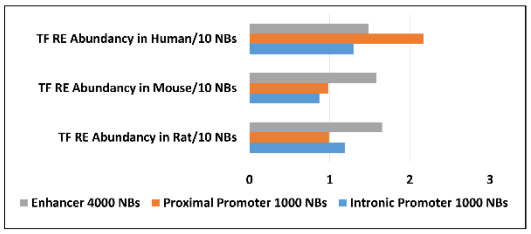

2D

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



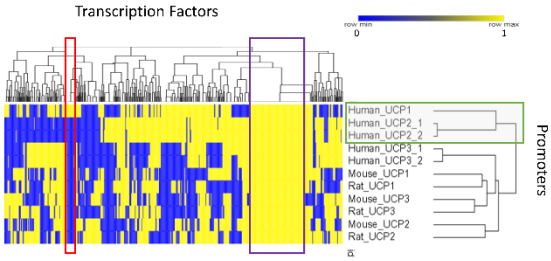

2E

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



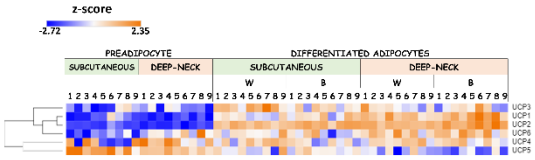

2F

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



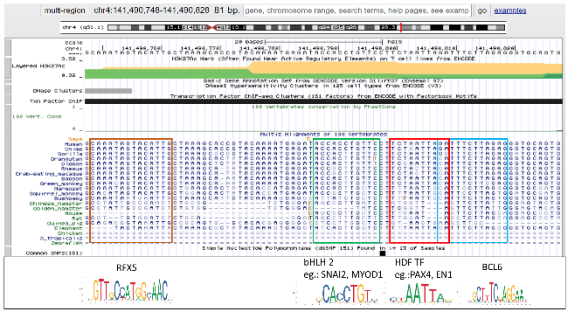

2G

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



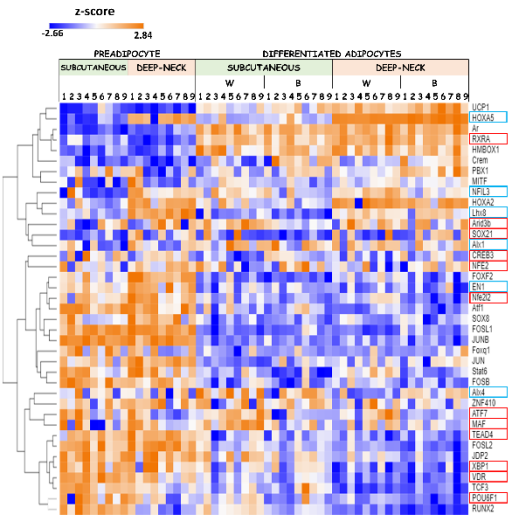

2I

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



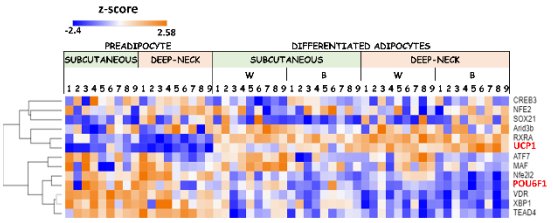

2H

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.




1444

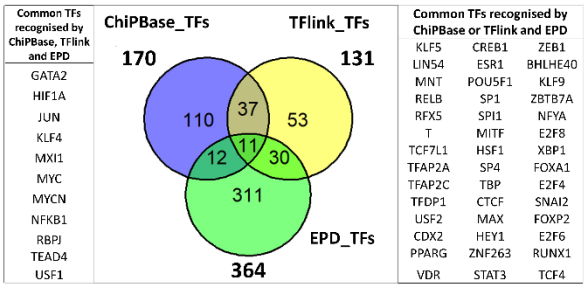
1445

1446 **Figure 3.**

**Figure 3.**

3A

Comparing Predicted and Experimentally Validated TF Binding Sites in the Human UCP1 Promoter.

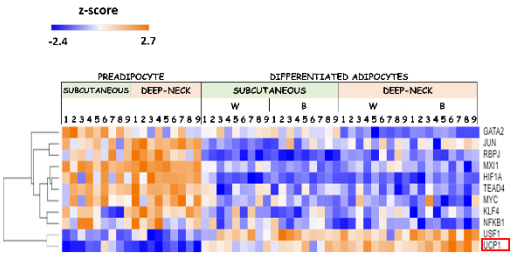



| Common TFs recognised by ChiPBase, TFlink and EPD |
|---|
| GATA2 |
| HIF1A |
| JUN |
| KLF4 |
| MXI1 |
| MYC |
| MYCN |
| NFKB1 |
| RBPJ |
| TEAD4 |
| USF1 |

| Common TFs recognised by ChiPBase or TFlink and EPD | | |
|---|---|---|
| KLF5 | CREB1 | ZEB1 |
| LIN54 | ESR1 | BHLHE40 |
| MNT | POU5F1 | KLF9 |
| RELB | SP1 | ZBTB7A |
| RFX5 | SPI1 | NFYA |
| T | MITF | E2F8 |
| TCF7L1 | HSF1 | XBP1 |
| TFAP2A | SP4 | FOXA1 |
| TFAP2C | TBP | E2F4 |
| TFDP1 | CTCF | SNAI2 |
| USF2 | MAX | FOXP2 |
| CDX2 | HEY1 | E2F6 |
| PPARG | ZNF263 | RUNX1 |
| VDR | STAT3 | TCF4 |

3B

Gene-expression profile of TFs common in human adipocytes of cervical origin that were all predicted by EPD, ChiPBase and Tflink interface

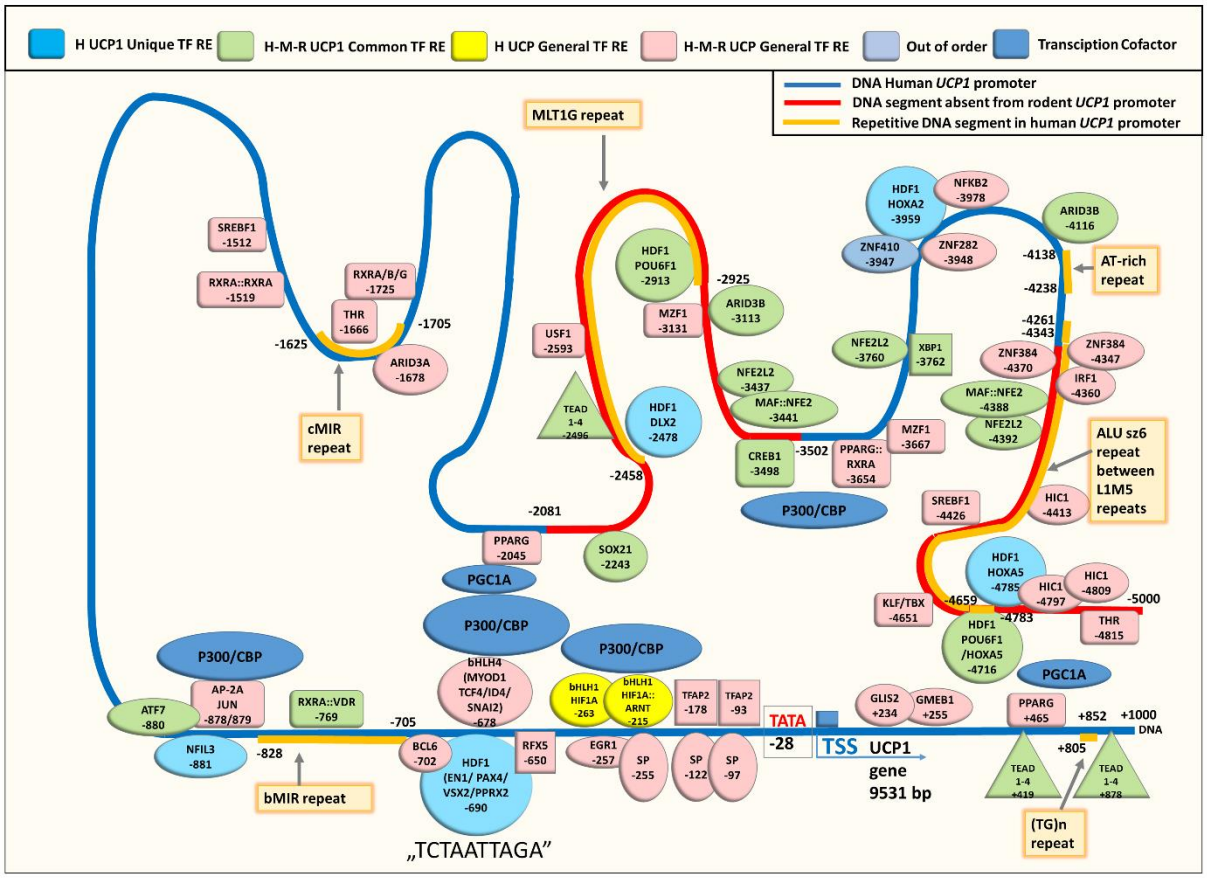



1447

1448

1449 **Figure 4.**

Figure 4.




1450

1451

1452 **TABLES**

1453 **Table 1.**

1454 **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 |

1455

1456

1457

1458

1459

1460

1461

1462

1463

1464

1465

1466 **Table 2.**

1467 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 |

1468 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 |

1469

1470

1471

1472

**Table 3.**

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

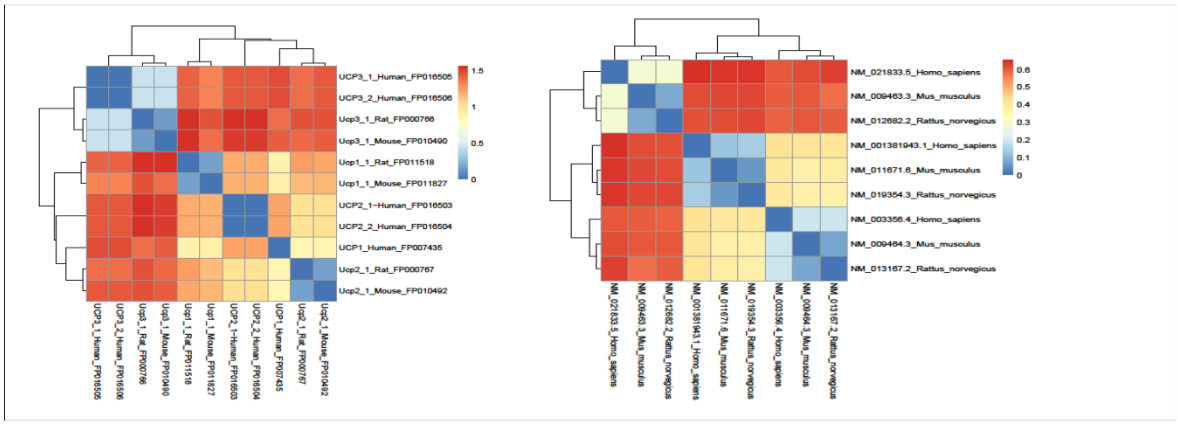
| TF | RE position in human UCP1 | GE Corr. TF vs.UCP1 | RNA Expression in Adipocytes | ChipSeq data | PubMED abstract with UCP1 | PWM in JASPAR | TF Class/Family | AdipoNET | EPD p<1e-5 |
|---|---|---|---|---|---|---|---|---|---|
| **POU6F1** | -4716 -4715 -2913 -2912 | NEG. ** | MEDIUM | NO | NO | | 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).

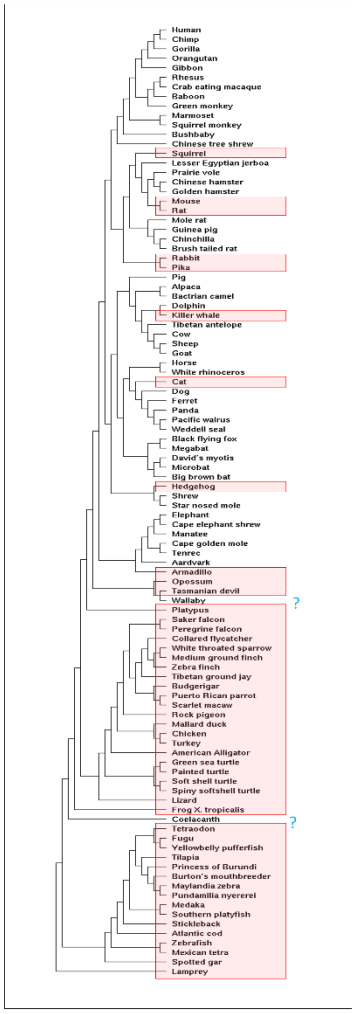
| TF | RE position in human UCP1 | GE. Corr. TF vs.UCP1 | RNA Expression in Adipocytes | ChipSeq data | PubMED abstract with UCP1 | PWM in JASPAR | TF Class/Family | AdipoNET | EPD p<1e-5 |
|---|---|---|---|---|---|---|---|---|---|
| **GATA2** | -2789 | Neg. ** | Low | ChiPBase/ TFlink | Yes | | Other 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 |

# SUPPLEMENTARY FIGURES

1A

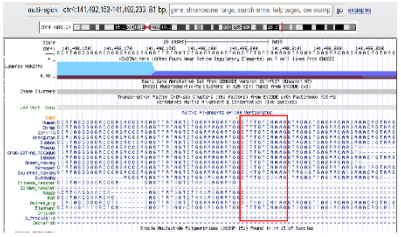



1B



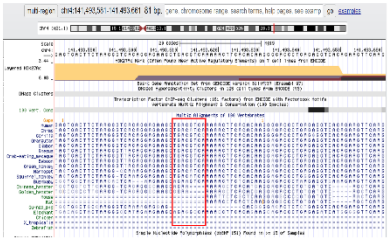

1C

Alignments of the *UCP1* promoter sequences (-2081 bp)



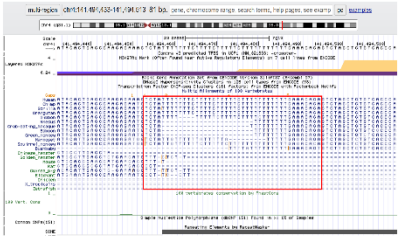
1D

Alignments of the *UCP1* promoter sequences (-3502 bp)



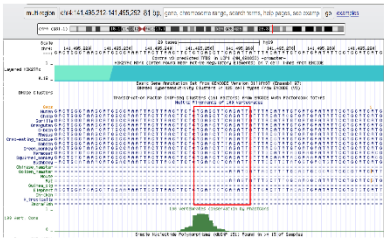
1E

Alignments of the *UCP1* promoter sequences (-4343 bp)



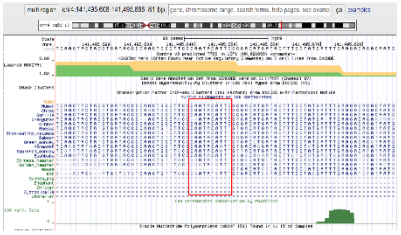
1F

Alignments of the *UCP1* promoter sequences (-5139 bp)



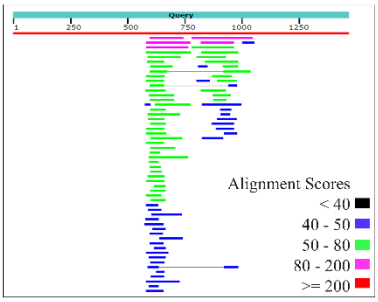
1G

Alignments of the *UCP1* promoter sequences (-5533 bp)



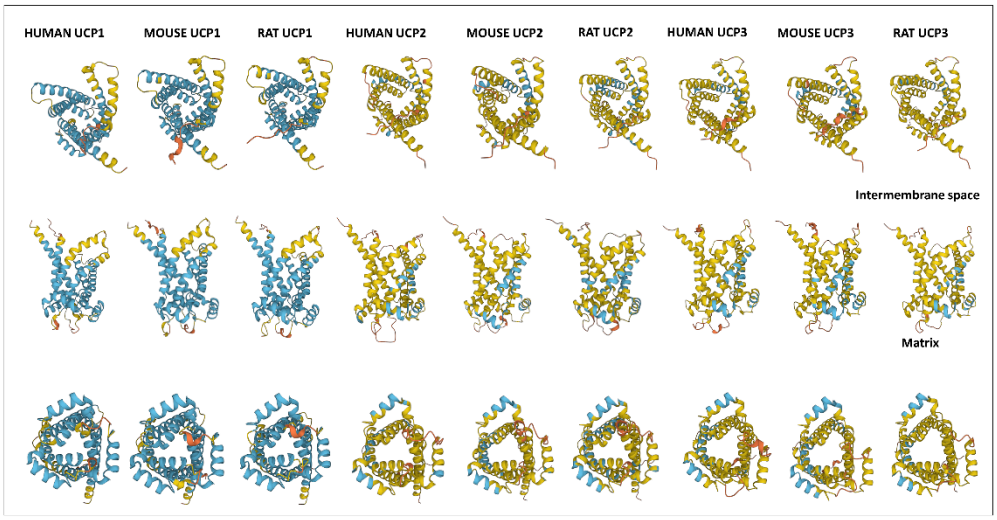
1H

Distribution of the Top 82 Blast Hits on 79 Subjects Sequences

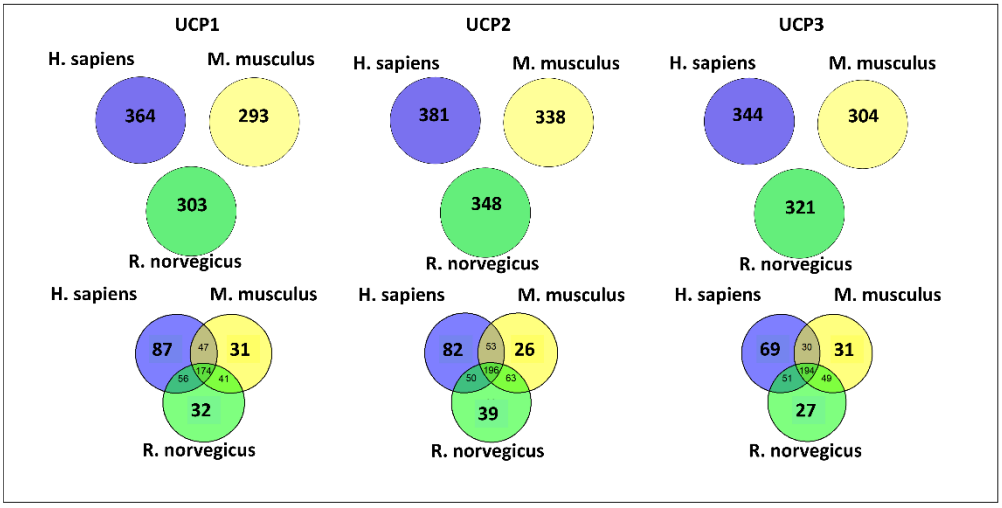


1485 **Supplementary Figure 2.**

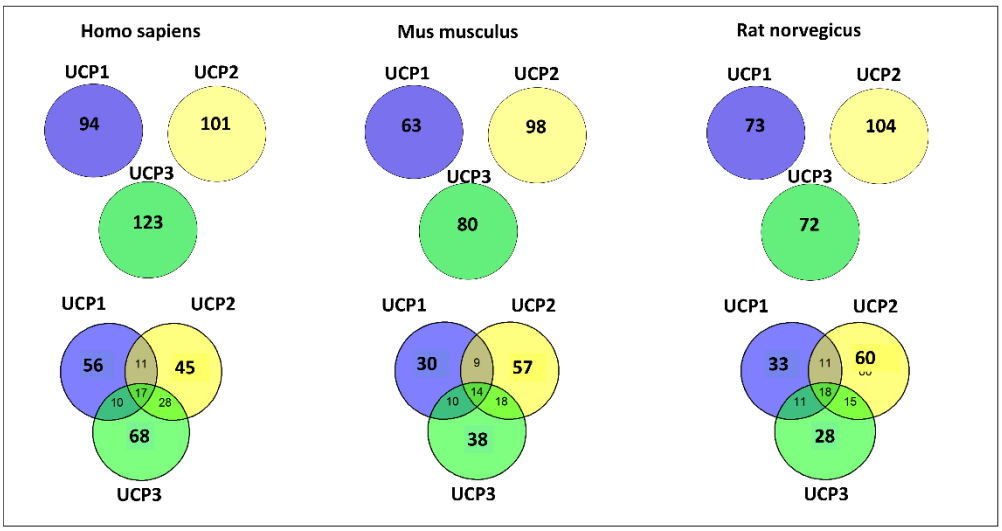
2A




2B



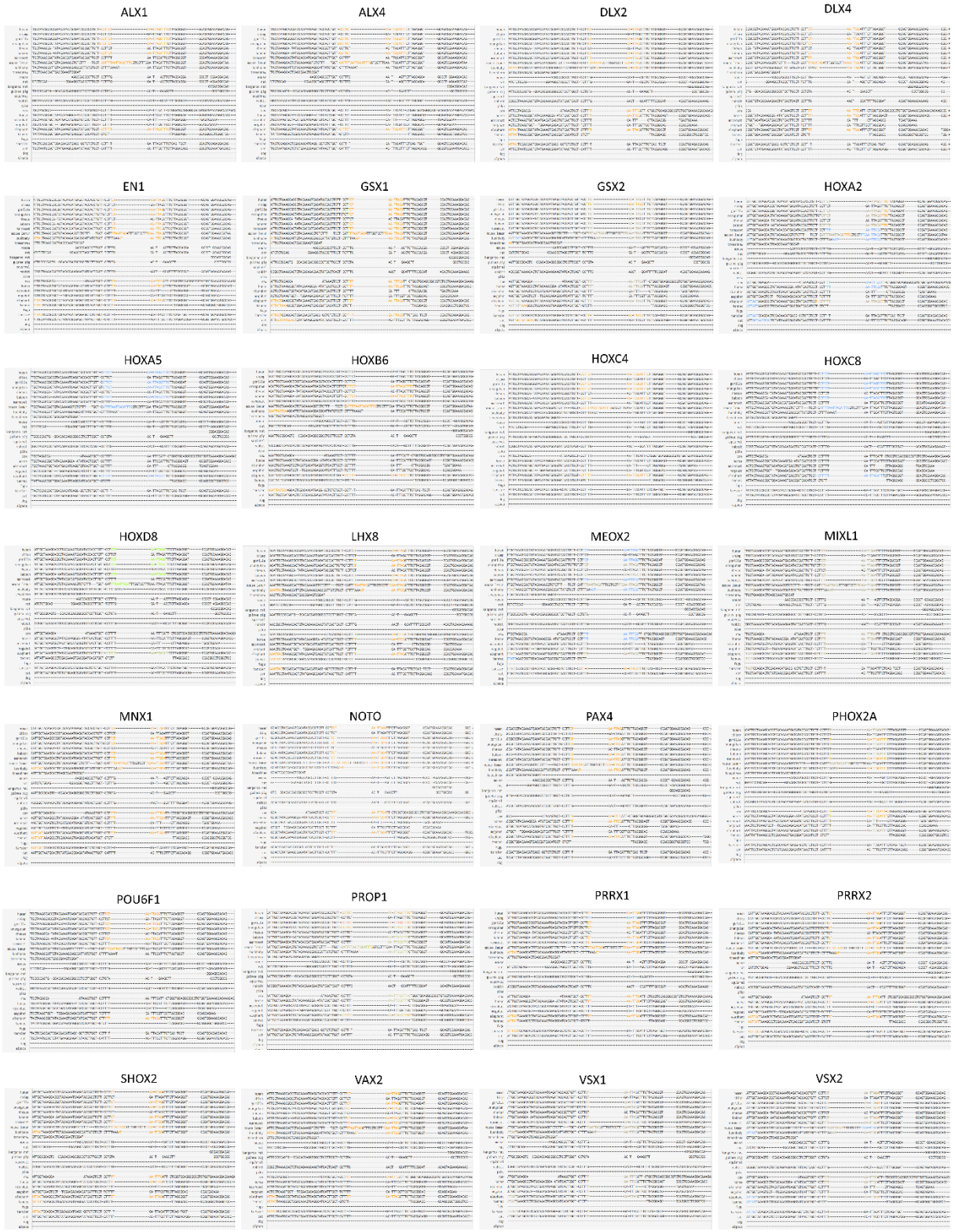

2C



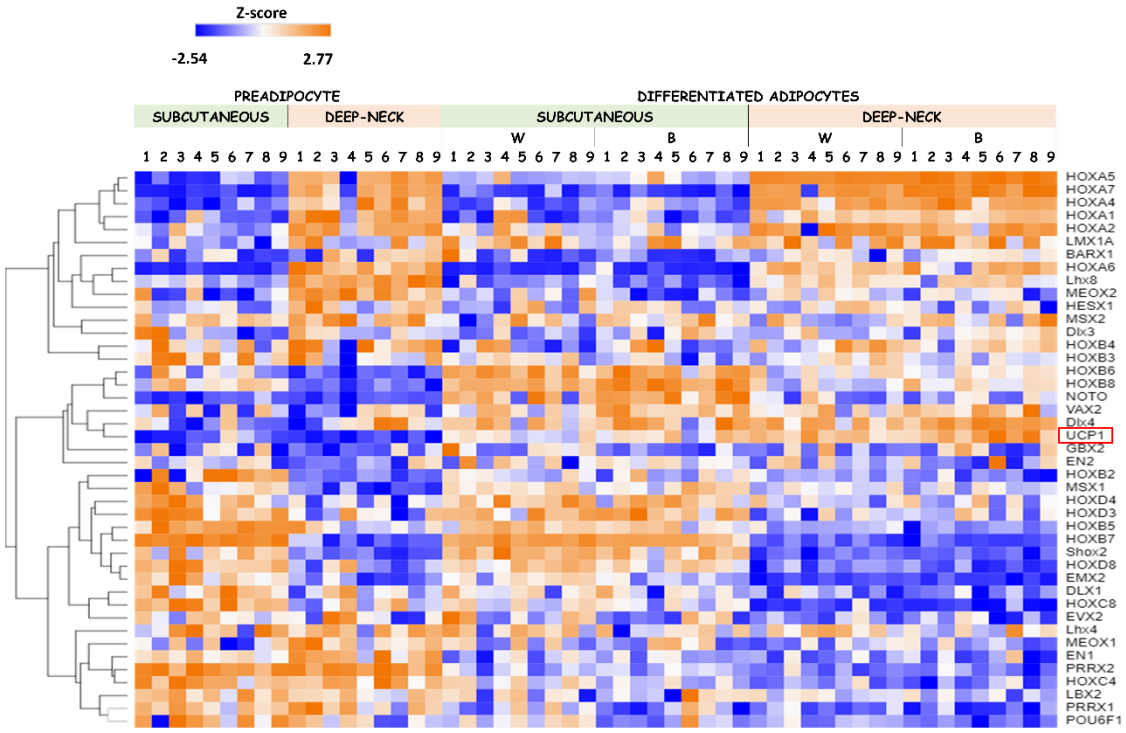

1486

1487

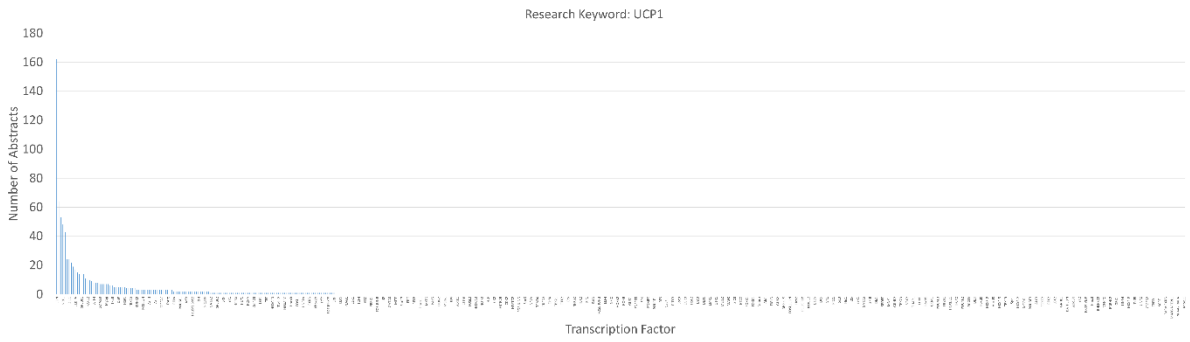
**Supplementary Figure 3.**



**Supplementary Figure 4.**



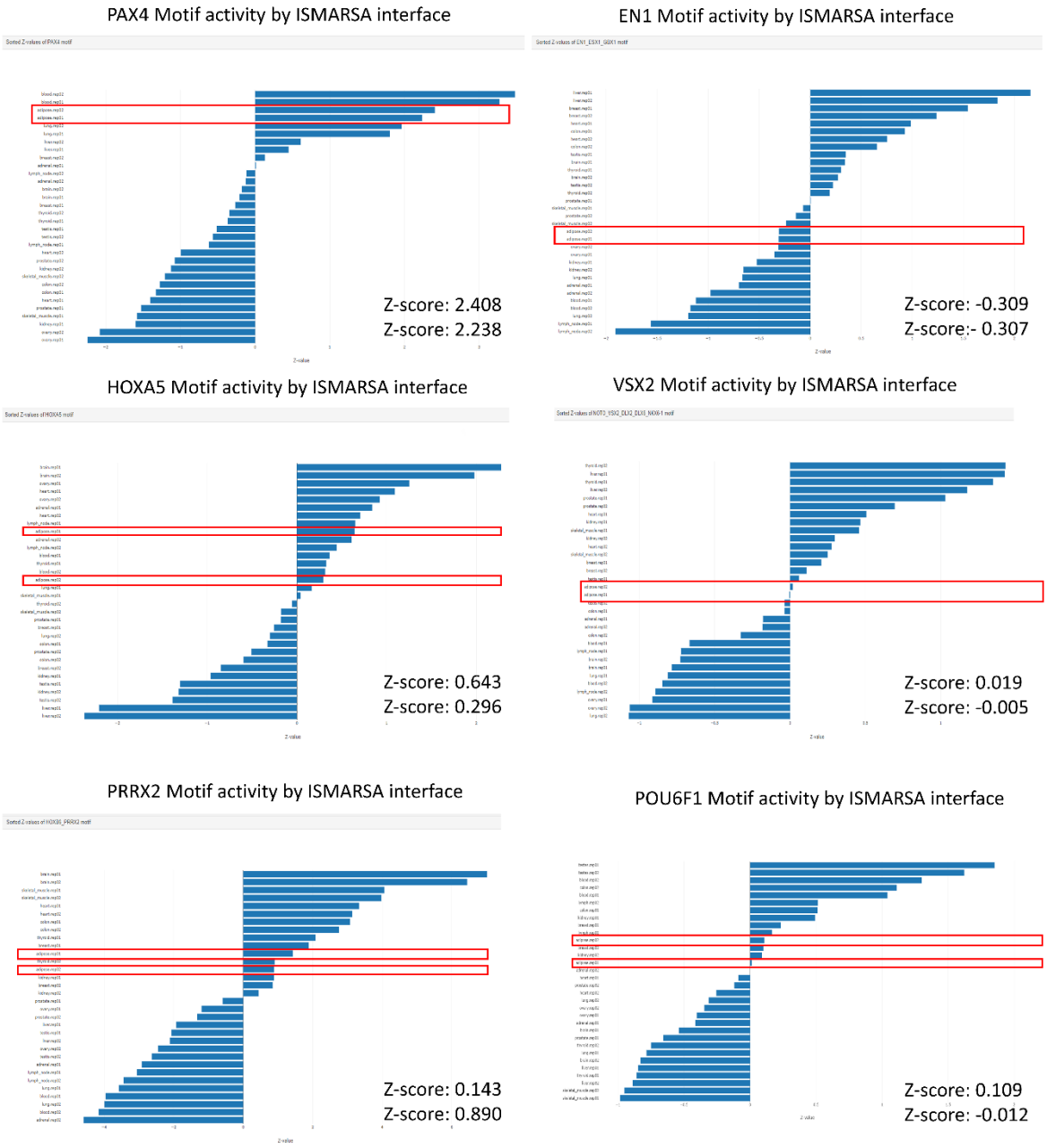
1494 **Supplementary Figure 5.**




1495

1496

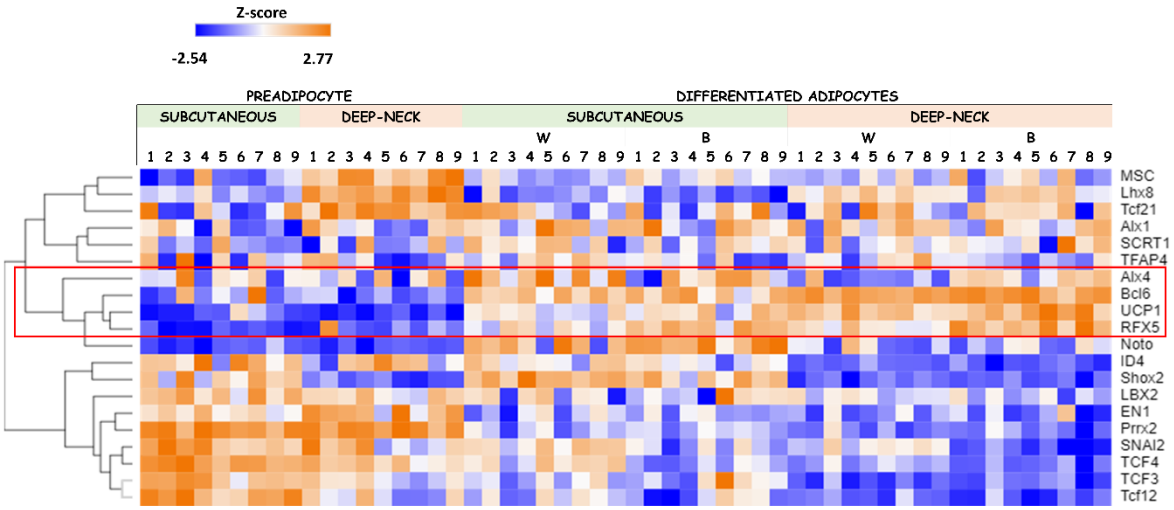
1497 **Supplementary Figure 6.**




1498

1499

1500 **Supplementary Figure 7.**

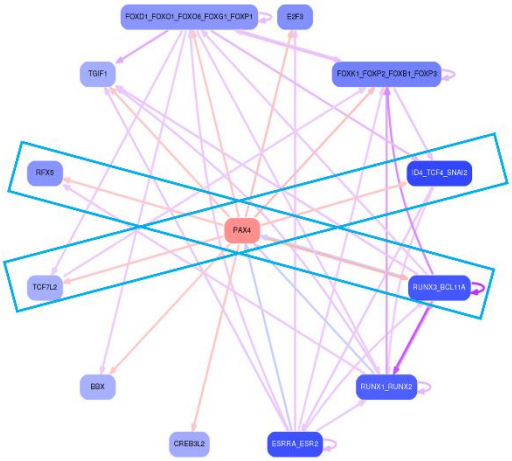



1501

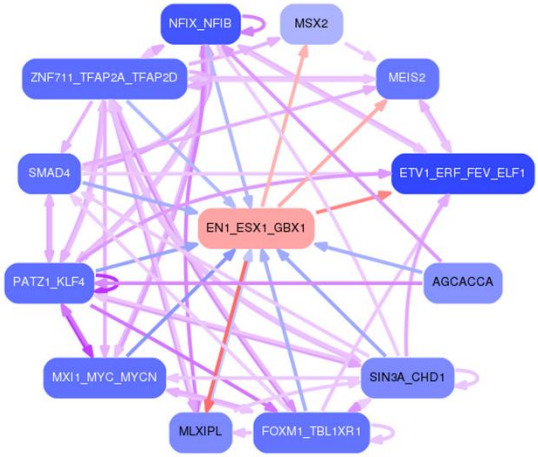
1502

**Supplementary Figure 8.**

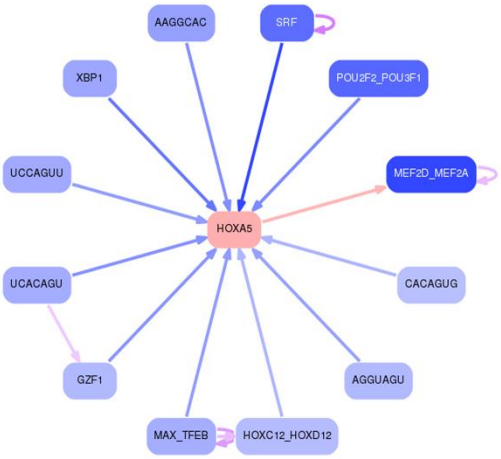
First level regulatory network of PAX4



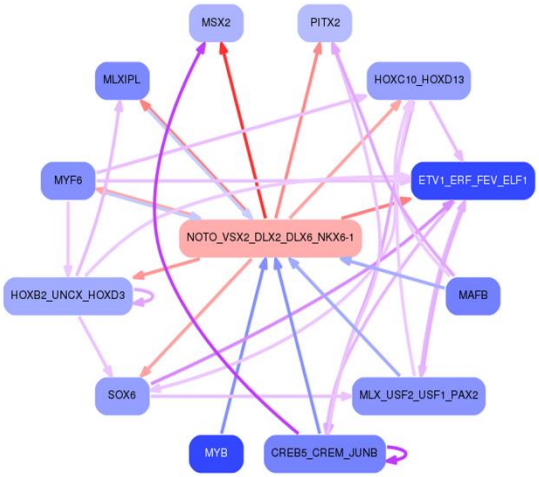

First level regulatory network of EN1



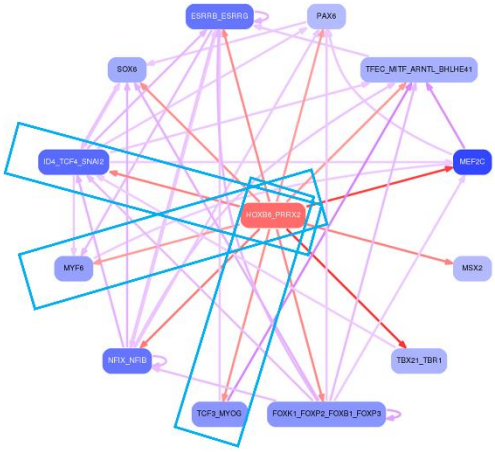

First level regulatory network of HOXA5



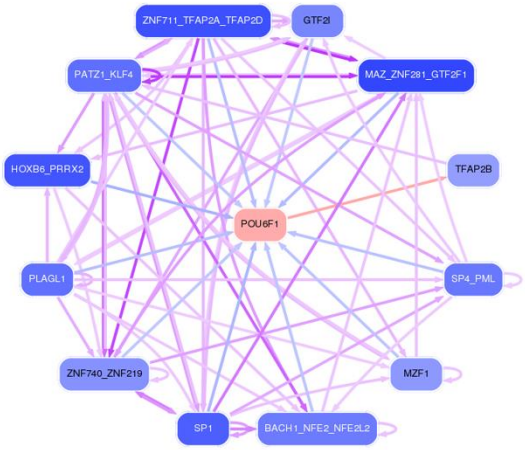

First level regulatory network of VSX2




First level regulatory network of PRRX2



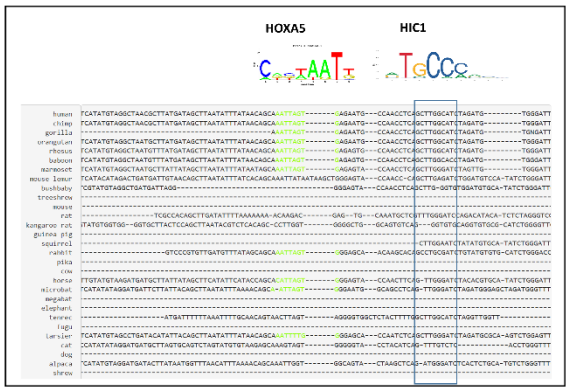

First level regulatory network of POU6F1



1506 **Supplementary Figure 9.**

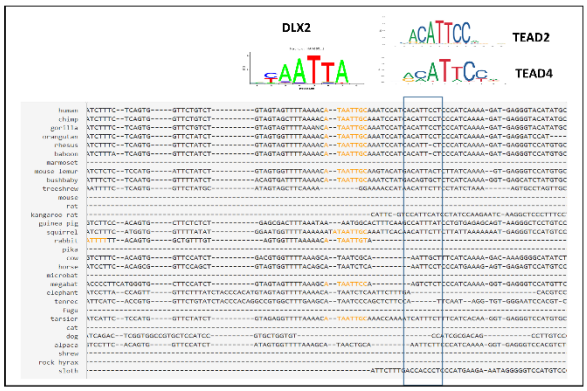
9A

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



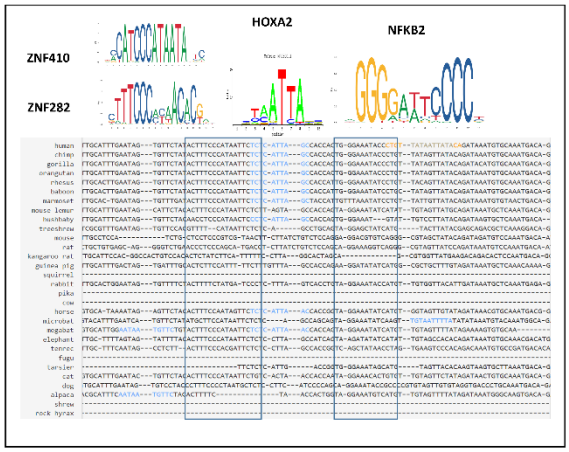

9B

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



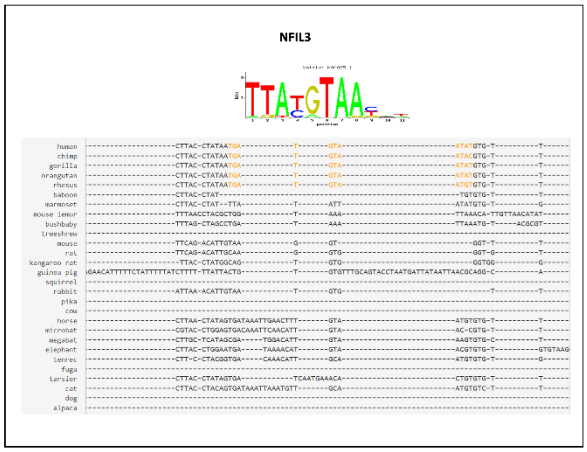

9C

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



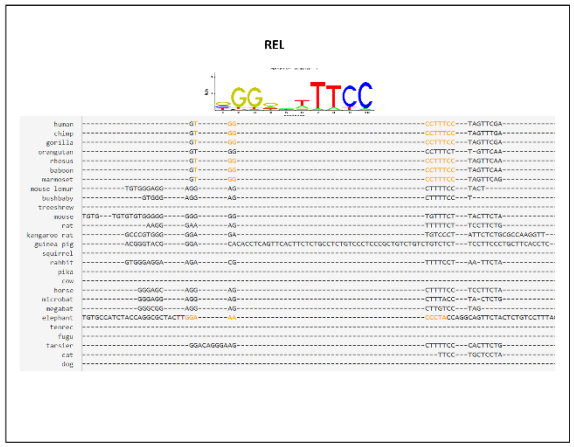

9D

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



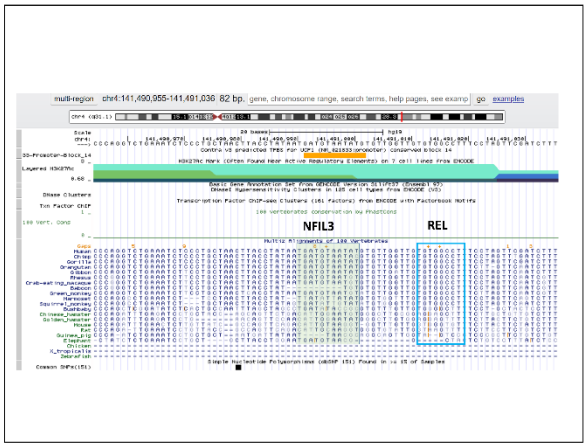

9E

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



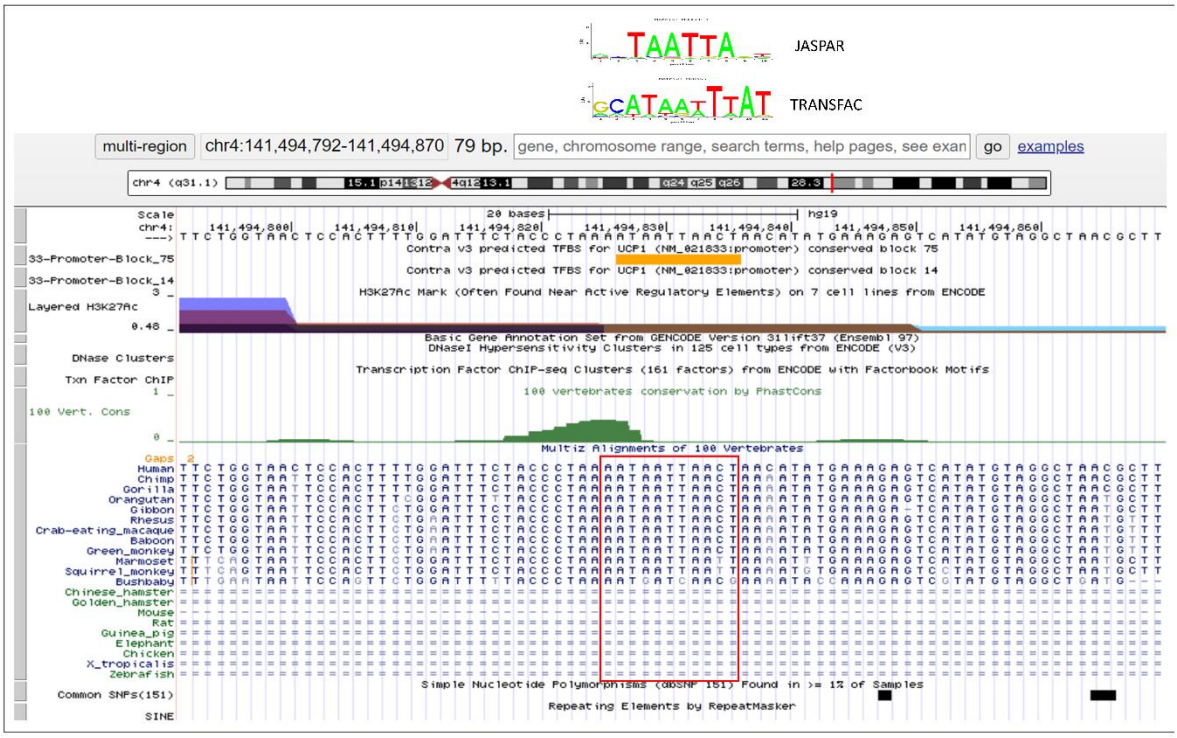

1507

1508

## Supplementary Figure 10.

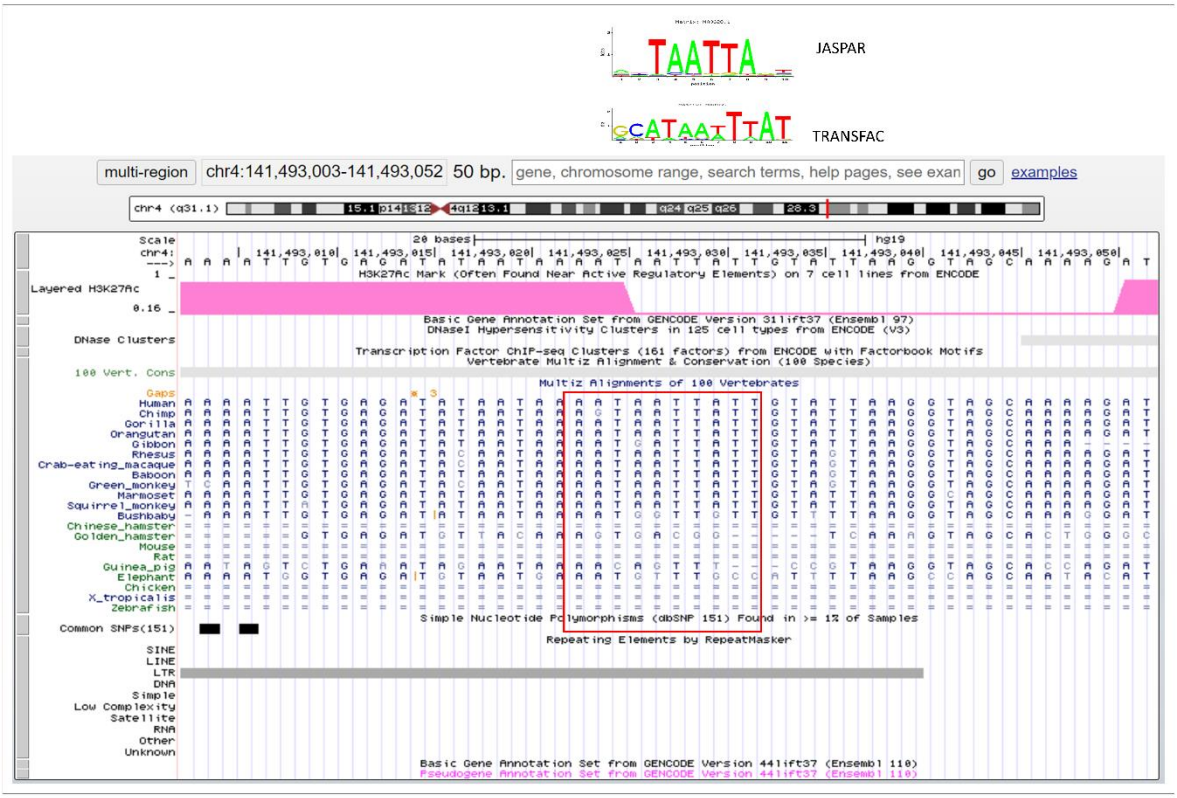
10A

POU6F1 RE at -4715 bp in human UCP1 promoter



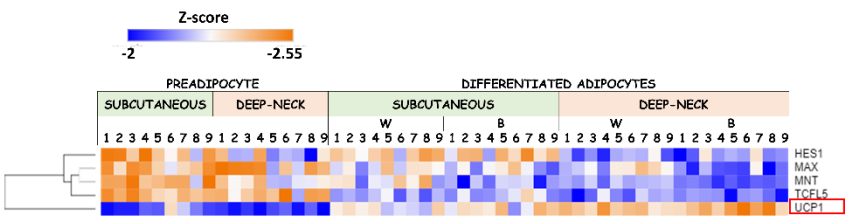

10B

POU6F1 RE at -2913 bp in human UCP1 promoter

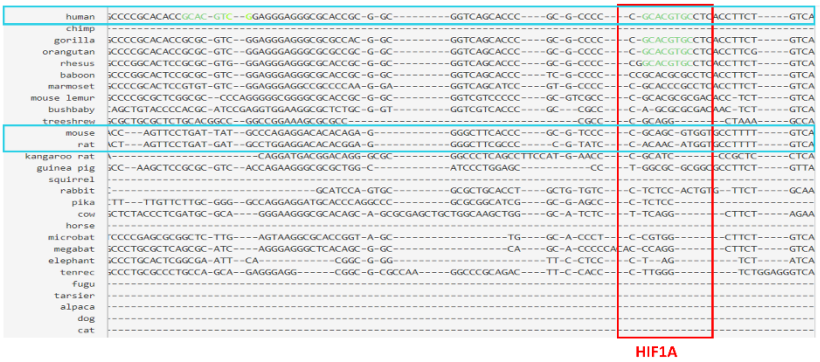


## Supplementary Figure 11.

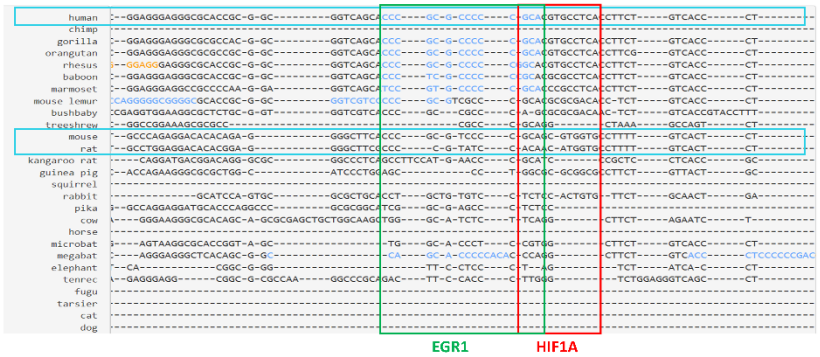
11A

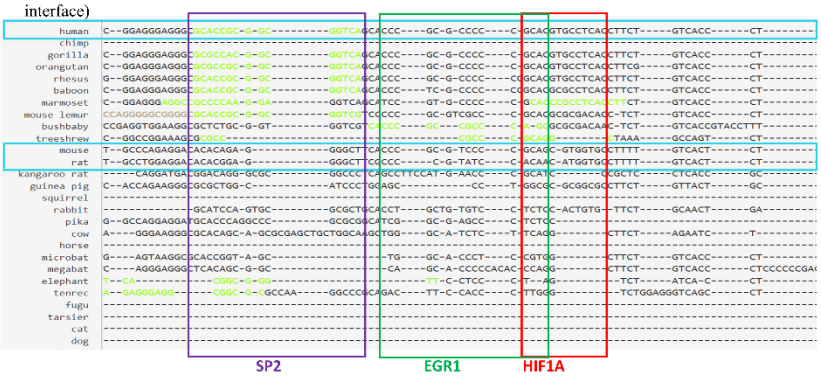



11B Phylogenetic footprinting: The HIF1A binding site in the human UCP1 promoter (ConTra 3v interface)




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)