- 1 $\,$ Transcriptional network analysis in muscle reveals AP-1 as a partner of PGC-1 $\!\alpha$ in the
- 2 regulation of the hypoxic gene program

3

Mario Baresic^{1,#}, Silvia Salatino^{1,2,3,#}, Barbara Kupr¹, Erik van Nimwegen^{2,3,*}, and Christoph
 Handschin^{1,*}

6

- 7 ¹Focal Area Growth & Development and ²Focal Area Computational & Systems Biology,
- 8 Biozentrum, University of Basel, Basel 4056, Switzerland
- 9 ³Swiss Institute of Bioinformatics, Basel 4056, Switzerland
- ¹⁰ [#]These authors contributed equally to this manuscript
- 11
- *Correspondence to: Christoph Handschin, Biozentrum, University of Basel,
 Klingelbergstrasse 70, CH-4056 Basel, Switzerland, email: <u>christoph.handschin@unibas.ch</u>
 +41 (0)61 267 23 78 or Erik van Nimwegen, Biozentrum, University of Basel,
 Klingelbergstrasse 70, CH-4056 Basel, Switzerland, email: <u>erik.vannimwegen@unibas.ch</u> +41
 (0) 61 267 15 76.

17

18 **Short title:** Transcriptional network regulation by PGC-1 α

19

- 20 **Keywords:** skeletal muscle; PGC-1α; coregulators; transcriptional regulation; transcription
- 21 factor binding site prediction; transcriptional networks; motif activity response analysis;
- 22 principal component analysis; exercise

23

25 ABSTRACT

Skeletal muscle tissue shows an extraordinary cellular plasticity, but the underlying 26 molecular mechanisms are still poorly understood. Here we use a combination of 27 experimental and computational approaches to unravel the complex transcriptional network 28 of muscle cell plasticity centered on the peroxisome proliferator-activated receptor γ 29 coactivator 1α (PGC- 1α), a regulatory nexus in endurance training adaptation. By integrating 30 data on genome-wide binding of PGC-1a and gene expression upon PGC-1a over-expression 31 with comprehensive computational prediction of transcription factor binding sites (TFBSs), 32 33 we uncover a hitherto underestimated number of transcription factor partners involved in 34 mediating PGC-1 α action. In particular, principal component analysis of TFBSs at PGC-1 α 35 binding regions predicts that, besides the well-known role of the estrogen-related receptor α 36 (ERR α), the activator protein-1 complex (AP-1) plays a major role in regulating the PGC-1 α controlled gene program of hypoxia response. Our findings thus reveal the complex 37 38 transcriptional network of muscle cell plasticity controlled by PGC-1a.

40 INTRODUCTION

A sedentary life style can lead to an imbalance between energy intake and expenditure and 41 42 favors the development of a number of chronic diseases like obesity and type 2 diabetes. 43 Regular exercise on the other hand is an effective way to reduce the risk for these lifestylerelated pathologies (1). The health benefits of exercise are at least in part induced by 44 changes in skeletal muscle tissue. Muscle cells exhibit a high plasticity and thus a remarkably 45 complex adaptation to increased contractile activity. For example, endurance training 46 47 induces mitochondrial biogenesis, increases capillary density and improves insulin sensitivity 48 (1, 2). To achieve such a complex plastic response, a number of different signaling pathways 49 are activated in an exercising muscle, for example p38 MAPK-mediated protein phosphorylation events, increased intracellular calcium levels or the activation of the 50 metabolic sensors AMP-dependent protein kinase (AMPK) and sirtuin-1 (SIRT1) (3). While 51 the temporal coordination of the numerous inputs is not clear, all of the major signaling 52 53 pathways converge on the peroxisome proliferator-activated receptor y coactivator 1α (PGC- 1α) to either induce *Pparqc1a* gene expression, promote post-translational modifications of 54 55 the PGC-1 α protein, or by doing both (4, 5). Upon activation, PGC-1 α mediates the muscular adaptations to endurance exercise by coactivating various transcription factors (TFs) 56 involved in the regulation of diverse biological programs such as mitochondrial biogenesis, 57 58 angiogenesis, ROS detoxification or glucose uptake (3). Accordingly, transgenic expression of PGC-1a in mouse skeletal muscle at physiological levels not only induces mitochondrial 59 biogenesis but also drives a fiber type conversion towards a more oxidative, slow-twitch 60 phenotype (6) while muscle-specific Ppargc1a knockout animals exhibit several symptoms of 61 pathological inactivity (7, 8). 62

63 Coregulators are part of multicomponent regulatory protein complexes that are well suited to translate external stimuli into changes in promoter and enhancer activities by combining 64 various enzymatic activities to modulate histones and chromatin structure, and recruit other 65 66 TFs (9). Thus, dynamic assembly of distinct coregulator complexes enables the integration of many different signaling pathways leading to a coordinated and specific regulation of entire 67 biological programs by multiple TFs (10, 11). For example, PGC-1 α not only recruits histone 68 69 acetylases (12), the TRAP/DRIP/Mediator (13) as well as the SWI/SNF protein complexes (14), but also binds to and coactivates a myriad of different transcription factors, even 70

though a systematic inventory of TF binding partners has not been compiled yet (15). Thus, the specific control exerted by the PGC-1 α -dependent transcriptional network might provide an explanation for the dynamic and coordinated muscle adaptation to exercise. Since PGC-1 α in skeletal muscle not only confers a trained phenotype, but also ameliorates several different muscle diseases (16), the unraveling of the PGC-1 α -controlled transcriptional network in skeletal muscle would be of great interest to identify putative therapeutic targets within this pathway.

Therefore, we aimed at obtaining a global picture of the co-regulatory activity of PGC-1 α in 78 79 skeletal muscle cells. More precisely, by combining data on the genome-wide binding 80 locations of PGC-1a and the gene expression profiles in response to PGC-1a over-expression 81 with comprehensive computational prediction of transcription factor binding site (TFBS) occurrence, we sought to unveil the biological processes that are regulated by PGC-1 α , to 82 identify the transcription factors that partner with PGC-1 α , and to determine the 83 84 mechanistic details of PGC-1 α -regulated transcription. We not only mapped the locations on the DNA where PGC-1 α was bound, but also delineated the target genes whose expression is 85 either directly or indirectly affected by PGC-1 α and identified novel putative transcription 86 factor partners that mediated PGC-1 α 's action. In particular, our results strongly suggest that 87 the activator protein-1 (AP-1) complex is a major regulatory partner of PGC-1 α , with AP-1 88 89 and PGC-1 α together regulating the hypoxic response gene program in muscle cells in vitro 90 and in vivo.

92 MATERIALS AND METHODS

93 Cell culture and siRNA transfection

94 C2C12 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 95 10% fetal bovine serum (FBS), 100 Units/ml penicillin and 100ug/ml streptomycin. To obtain 96 myotubes, the C2C12 myoblasts were allowed to reach 90% confluence and the medium was 97 changed to DMEM supplemented with 2% horse serum (differentiation medium) for 72 98 hours.

99 The siRNAs for the knockdown of NFE2L2, FOS, JUN, ATF3, NFYC, ZFP143, GTF2I, the nontargeting siRNA pool and the DharmaFECT1 transfection reagent were purchased from Dharmacon (Fisher Scientific) and the siRNA transfection was performed according to the Thermo Scientific DharmaFECT Transfection Reagents siRNA Transfection Protocol. Briefly, after three days of differentiation, the respective siRNAs (50nM final concentration) was added to the medium. 24h after siRNA transfection, the cells were infected with either the PGC-1α or GFP adenovirus. Then, 48h after adenoviral infection, the cells were harvested.

Differentiated C2C12 cells were infected with adenoviral (AV) shERRa (kindly provided by Dr. 106 Anastasia Kralli, Scripps Research Institute, La Jolla, CA, USA) to knockdown and inactivate 107 ERRa or shGFP as a control. The infected cells were kept in culture for 4 days. Afterwards, 108 109 cells were infected with the AV-flag-PGC-1 α or AV-GFP and kept in culture two additional days. As a supplement to the previously infected AV shERR α cells, 2 μ M of the ERR α inverse 110 111 agonist XCT-790 were added. To the remaining cells, 0.02% DMSO as a vehicle were added to the differentiated medium. All the experiments have been performed in biological 112 triplicates. For RNA isolation, TRIzol was used according to the TRIzol reagent RNA isolation 113 protocol (Invitrogen). Three conditions were used for further analysis: AV-shGFP + AV-GFP + 114 vehicle, AV-shGFP + AV-flag-PGC-1 α + vehicle, AV-shERR α + AV-flag-PGC-1 α + 2 μ M XCT-790. 115

116

117 ChIP and ChIP Sequencing

ChIP was performed according to the Agilent Mammalian ChIP-on-chip Protocol version
 10.0. For each immunoprecipitation, approximately 1x10⁸ C2C12 cells were differentiated
 into myotubes and infected with AV-flag-PGC-1α. For cross-linking protein complexes to DNA

121 binding elements, the cells were incubated in a 1% formaldehyde solution for 10 minutes, 122 followed by the addition of glycine to a final concentration of 125mM to quench the effect of the formaldehyde. The cells were rinsed in 1xPBS, harvested in ice-cold 1xPBS using a 123 silicone scraper and pelleted by centrifugation. The pelleted cells were either used 124 immediately or flash frozen and stored for later. The cells were then lysed at 4°C using two 125 126 lysis buffers containing 0.5% NP-40/0.25% Triton X-100 and 0.1% Na-deoxycholate/0.5% N-127 lauroylsarcosine, respectively. The chromatin was then sheared by sonication to obtain DNA fragments of about 100-600bp in length. 50µl of the sonicated lysate was saved as input 128 129 DNA. The immunoprecipitation was performed overnight at 4°C using magnetic beads (Dynabeads® Protein G, Invitrogen), which were previously coated with monoclonal 130 antibodies like the monoclonal ANTI-FLAG® M2 Antibody, Sigma for the ChIP of PGC-1a or 131 with the monoclonal anti-c-Fos (9F6) rabbit antibody #2250, Cell Signaling for the ChIP of 132 133 FOS. The beads carrying the precipitate were washed five times for the c-Fos antibody and six times for the flag antibody with RIPA buffer and once with TE that contained 50mM NaCl 134 to eliminate unspecific binding of DNA to the beads. For elution, the beads were 135 136 resuspended in elution buffer containing 1% SDS, placed in 65°C water bath for 15 minutes and vortexed every 2 minutes. To reverse the cross-links, the samples were incubated at 137 65°C overnight. The following day, the RNA and the cellular proteins were digested using 138 139 RNase A and proteinase K. The DNA was precipitated and the success of the chromatin 140 immunoprecipitation was validated by semiquantitative real-time PCR. The ChIP experiments were performed in triplicates. The ChIP of PGC-1 α was further used for 141 sequencing. The ChIP-Seq experiment of over-expressed PGC-1 α in C2C12 cells was 142 performed in biological duplicates. At the joint Quantitative Genomics core facility of the 143 University of Basel and the Department of Biosystems Science and Engineering (D-BSSE) of 144 the ETH Zurich in Basel, DNA libraries were prepared using the standard Illumina ChIP-Seq 145 146 protocol, as described by the manufacturer, and the immunoprecipitated samples 147 sequenced on the Genome Analyzer II. In order to keep only high quality data, the sequenced reads were filtered based on the quality score of each read and its alignments. 148 149 Read were retained when Phred score >= 20, read length >= 25 bps and number of wrongly called nucleotides (Ns) <= 2. Those reads that passed the filter, (6'711'717 for the first 150 immunoprecipitated sample (IP), 36'580'431 for the second IP, 17'899'074 for the first 151 152 Whole Cell Extract (WCE), and 35'525'221 for the second WCE), were aligned to the mouse genome, UCSC mm9 assembly, using Bowtie version 0.12.7 (17) using parameters --best -strata -a --m 100. The number of aligned reads equaled 5'699'648 for the first IP sample,
16'053'370 for the first WCE, 21'448'059 for the second IP, and 32'244'584 for the second
WCE.

157

158 Identification of bound regions

To identify regions that were significantly enriched in the ChIP, we passed a 200 bps long sliding window along the genome, sliding by 25 bps between consecutive windows, and estimated the fraction of all ChIP reads f_{IP} that fall within the window, as well as the fraction f_{WCE} of reads from the whole cell extract that fall in the same window (which we estimate from a 2000 bps long window centered on the same genomic location). A Z-score quantifying the enrichment in the ChIP of each window was computed as:

$$Z = \frac{f_{IP} - f_{WCE}}{\sqrt{\sigma^2_{IP} + \sigma^2_{WCE}}}$$

where σ^2_{IP} and σ^2_{WCE} are the variances of the IP and WCE read frequencies, which are given by:

167
$$\sigma_{IP}^2 = \frac{f_{IP} * (1 - f_{IP})}{N_{IP}} \text{ and } \sigma_{WCE}^2 = \frac{f_{WCE} * (1 - f_{WCE})}{N_{WCE}}$$

168 respectively.

The enrichments were reproducible across biological replicates. Using only the first sequencing dataset, we called peaks at a Z cutoff of 4.5; we then compared these with the Z scores from the corresponding regions of the second dataset and the Pearson correlation coefficient was found to be 0.778. Similarly, we called peaks at a Z cutoff of 4.5 using only the second sequencing dataset; when we compared these peaks with the Z scores of the corresponding regions from the first dataset, the Pearson correlation coefficient was found to be 0.782.

To obtain a final set of binding peaks, we combined the reads from the two biological replicates computing the Z score of each window was computed as: $Z = \frac{f_{IP1} + f_{IP2} - f_{WCE1} - f_{WCE2}}{\sqrt{\sigma^2_{IP1} + \sigma^2_{IP2} + \sigma^2_{WCE1} + \sigma^2_{WCE2}}}$

179 We conservatively considered all windows with a Z-score larger than 4.5 as were considered 180 significantly enriched (False Discovery Rate 0.6%). The final binding peaks were obtained by merging consecutive windows that all passed the cut-off and by considering the "peak" to 181 182 correspond to the top scoring window, i.e. corresponding to the summit of the ChIP-Seq 183 signal. To determine the PGC-1α distribution genome-wide, peaks were annotated according in relation to their closest Mus musculus RefSeg transcripts. We defined peaks as: "Intronic" 184 185 (peak center lying inside an intron); "Exonic" (peak center lying inside an exon); "Upstream of TSS" (peak center lying within -10 to 0 kb from the closest TSS); "Downstream of TES" 186 187 (peak center lying within 0 to 10 kb from the closest TES); "Intergenic" (peak center located farer than 10 kb from the nearest transcript). Moreover, we computed the ratio between 188 189 observed and expected peak location distributions, obtained by generating 100 peak sets 190 composed of 7512 random peaks each.

191

178

192 Motif finding and TFBSs over-representation

193 The binding peak regions were aligned to orthologous regions from other 6 mammalian 194 species – human (hg18), rhesus macaque (rheMac2), dog (canFam2), horse (equCab1), cow (bosTau3) and opossum (monDom4) - using T-Coffee (18). A collection of 190 mammalian 195 196 regulatory motifs (position weight matrices or WMs) representing the binding specificities of approximate 350 mouse TFs (in many cases, sequence specificities of multiple closely-related 197 198 TFs were represented with the same WM) were downloaded from the SwissRegulon website (19). TFBSs for all known motifs were predicted using the MotEvo algorithm (20) on the 199 alignments of all the 7512 peak sequences. Only binding sites with a posterior probability >= 200 201 0.1 were considered for the further steps of the analysis. In order to create a background set 202 of regions to assess the overrepresentation of binding sites within our regions, we created randomized alignments by shuffling the multiple alignment columns, maintaining both the 203 204 gap patterns and the conservation patterns of the original alignments. TFBSs were predicted 205 on the shuffled alignments using the same MotEvo settings as for the original peak 206 alignments. Over-representation of motifs in the PGC-1 α binding peaks was calculated by comparing total predicted TFBS occurrence within binding peaks with the predicted TFBS 207

208 occurrence in the shuffled alignments. We evaluated the enrichment of TFBSs for each motif 209 *x* by collecting the sum n_x of the posterior probabilities of its predicted sites in the peak 210 alignments as well as the corresponding sum n'_x in the shuffled alignments, and computed a 211 Z-score:

$$Z = \frac{f_x - f'_x}{\sqrt{\frac{f_x * (1 - f_x)}{L_x} + \frac{f'_x * (1 - f'_x)}{L'_x}}}$$

where L_x and L'_x are the total lengths of the original and shuffled alignments, respectively, while f_x and f'_x are given by the equations:

214
$$n_x * l_x = f_x * L_x$$
 and $n'_x * l_x = f'_x * L'_x$

215 with l_x the length of motif *x*.

216

217 Principal Component Analysis of TFBS occurrence in binding peaks

The input matrix N for the Principal Component Analysis (PCA) contained the total number 218 219 of predicted binding sites N_{pm} in each of the 7512 binding peaks p (rows) for each of the 190 mammalian regulatory motifs m (columns). After mean centering the columns of this matrix, 220 $\widetilde{N}_{pm} = N_{pm} - \langle N_m \rangle$, i.e. subtracting the average site count for each motif, Singular Value 221 Decomposition (SVD) was used to factorize this matrix: $\tilde{N} = U \cdot S \cdot V^T$, where U is a $P \times M$ 222 matrix whose columns are the left singular vectors of \tilde{N} ; S is a $M \times M$ diagonal matrix 223 containing the singular values, and V^T (the transpose of V) is an $M \times M$ matrix whose rows 224 are the right singular vectors, with P the number of peaks, and M the number of motifs. The 225 SVD was performed using the "svd" package of the "R" programming language. 226

227

228 Gene expression arrays

229 Whole-gene expression after 48 hours of transfection with adenovirus was measured in 230 C2C12 cells with Affymetrix GeneChip[®] Mouse Gene 1.0 ST microarrays at the Life Science 231 Training core facility of the University of Basel. Raw probe intensities were corrected for 232 background and unspecific binding using the Bioconductor package "affy" (21). 233 Subsequently, probes were classified as expressed or non-expressed by using the "Mclust" R package (22) and, after removal of non-expressed probes, the intensity values were quantile 234 normalized across all samples. Using mapping of the probes to the UCSC collection of mouse 235 236 mRNAs, probes were then associated to a comprehensive collection of mouse promoters available from the SwissRegulon database (19). The log2 expression level of a given 237 238 promoter was calculated as the weighted average of the expression levels of all probes 239 associated to it. Log2 expression levels were then compared between over-expressed PGC- 1α and the control GFP sample; for each promoter, the change in expression level across the 240 two conditions was measured by log2 fold change (log2FC), computed as the difference 241 242 between the mean of the log2 values in PGC-1 α and the mean of the log2 values in GFP. The 243 significance of the expression change was assessed by a Z score, which was computed as:

$$Z = \frac{\overline{E}_{PGC1\alpha} - \overline{E}_{GFP}}{\sqrt{\frac{\sigma^2_{PGC1\alpha}}{n} + \frac{\sigma^2_{GFP}}{n}}}$$

where n = 3 was the number of replicate samples, $\bar{E}_{PGC1\alpha}$ is the mean log2 expression across the PGC-1 α samples, \bar{E}_{GFP} is the mean log2 expression across the GFP samples, and $\sigma^2_{PGC1\alpha}$ and σ^2_{GFP} are the variances of log2 expression levels across the replicates for the PGC-1 α and control samples, respectively. Promoters were considered significantly up-regulated when log2FC >= 1 and Z >= 3, and significantly down-regulated when log2FC <= -1 and Z <= -3.

Peaks were assigned to promoters by proximity. To assign each peak to a promoter, we calculated the distance from the center of the peak to the center of neighboring promoters; whenever the peak was closer than 10 kb from at least one promoter, it was assigned to the nearest promoter and, thus, to its associated gene.

254

255 Gene Ontology enrichment analysis

Gene IDs were extracted from differentially regulated promoters and divided in four groups: up-regulated promoters with an assigned binding peak, up-regulated promoters without an assigned binding peak, down-regulated promoters with an assigned peak, and downregulated promoters without an assigned peak. These four gene sets were used as input for the functional analysis tool FatiGO (23) to identify significantly over-represented Gene
 Ontology (GO) categories compared to all Mus musculus genes. Only GO terms having an
 FDR-adjusted p-value <= 0.05 were considered significant.

263

264 Motif activity at direct and indirect targets of PGC-1α

265 To integrate the information from the PGC-1 α binding peaks, we extended MARA (24) to model the direct and indirect regulatory effects of PGC-1 α . Given the input expression data 266 and the computationally predicted binding sites, MARA infers, for each of 190 regulatory 267 motifs m, the activity A_{ms} of the motif in each sample s when the motif occurs outside of a 268 region of PGC-1 α , and the activities A_{ms}^* of the motifs when they occur within a PGC-1 α 269 270 binding peak. That is, changes in the motif activities A_{ms} upon over-expression of PGC-1 α 271 indicate indirect regulatory effects of PGC-1 α on each motif *m*, whereas changes in the motif activities A^*_{ms} reflect direct regulatory effects of PGC-1 α as mediated by each motif m. For 272 273 each promoter p that was not associated with any PGC-1 α binding peak (which we denote 274 indirect targets), we modeled its log-expression in sample s, eps, in terms of the predicted number of TFBSs N_{pm} that occur in the proximal promoter region (running from -500 to +500 275 276 relative to TSS) for each regulatory motif *m*. That is, MARA assumes the linear model:

$$e_{ps} = c_p + \tilde{c}_s + \sum_m N_{pm} A_{ms}$$

where c_p is the basal expression of promoter p, \tilde{c}_s is a sample-dependent normalization constant, and A_{ms} is the regulatory activity of motif m in sample s, which is inferred by the model. Formally, A_{ms} quantifies amount by which the expression of promoter p in sample swould be reduced if a binding site for motif m were to be deleted from the promoter.

For each "direct target" promoter *p* that has an associated PGC-1 α binding peak, which we defined as promoters with a peak within 1 kb or with a peak within 100 kb that was highly conserved according to PhastCons score of the region (25), we model its expression in terms of the predicted TFBSs in the binding peak, i.e.:

$$e_{ps} = c_p + \tilde{c}_s + \sum_m N_{pm}^* A_{ms}^*$$

where N_{pm}^* is the number of predicted TFBSs for motif *m* in *the peak* associated with promoter *p*, and A_{ms}^* is the motif activity of regulator *m* in sample *s* when this motif occurs in the context of PGC-1 α binding. That is, the inferred motif activities A_{ms} quantify the activities of regulatory motifs when they occur independent of PGC-1 α binding, and the motif activities A_{ms}^* quantify the activities of motifs when they occur in a PGC-1 α binding peak, i.e. the latter activities reflect direct effects of a PGC-1 α while the former reflect indirect effects.

292 MARA predicts activities for 190 different mammalian regulatory motifs, associated with 293 roughly 350 mouse TFs. Besides motif activities MARA also calculates error-bars δ_{ms} for each 294 motif *m* in each sample *s*. Using these, MARA calculates, for each motif *m*, an overall 295 significance measure for the variation in motif activities across the samples analogous to a z-296 statistic:

$$z_m = \sqrt{\frac{1}{S} \sum_{s=1}^{S} \left(\frac{A_{ms}}{\delta_{ms}}\right)^2}$$

297 For each motif we calculate both a z-score z_m associated with its indirect activity changes, 298 and a z-score z_m^* associated with its direct activity changes. MARA also ranks the confidence 299 on predicted target promoters of each motif by a Bayesian procedure that quantifies the contribution of that factor to explaining the promoter's expression variation by a Chi-300 squared value (for details, see (24)). The parameters used for motif stratification were: (i) 301 302 the Z score z_m^* for direct activity changes, (ii) the Z score z_m for indirect motif activity 303 changes, (iii) the Z score \bar{z}_m^* for direct motif activity changes, computed by averaging the sample replicates and (iv) the Z score \bar{z}_m for indirect motif activity changes, computed by 304 averaging the sample replicates. The latter two measures were used to show which direction 305 306 the motif activity changes when over-expressing PGC-1 α with respect to the control condition. All motifs *m* for which either the direct or indirect motif activities were changing 307 308 significantly ($z \ge 2$) were subsequently selected.

309

310 De novo motif finding

PhyloGibbs (26) was used to identify *de novo* motifs across the 200 top enriched PGC-1 α peaks. The parameters used were -D 1 -z 1 -y 200 -m 10, corresponding to searching on multiple alignments for a single motif of length 10 with a total of 200 sites. The resulting motif was scanned for similarity to the other known motifs from our dataset using STAMP (27), with settings: Pearson Correlation Coefficient for column comparison metric, Smith-Waterman for the alignment method, penalty of 0.5 and 0.25 for gap opening and gap extension, respectively.

318

319 Real-time PCR and target gene validation

Putative target genes of distinct transcription factor-PGC-1α combinations were chosen according to three criteria: first, positive transcriptional regulation by PGC-1α by more than 2 fold, second, presence of a PGC-1α binding peak within a 10 kb distance from the TSS and third, prediction of targeting by MARA with a positive Chi-squared score. The sequences of the primers used in real-time PCR experiments are depicted in Suppl. Table 1. Relative mRNA was quantified by qPCR on a StepOnePlus system (Applied Biosystems) using Power SYBR Green PCR Master Mix (Applied Biosystems).

The values are presented as the mean +/- SEM. A Student's t-test was performed and a pvalue < 0.05 was considered as significant (*p<0.05, **p<0.01, ***p<0.001).

329

330 Animals

Mice were housed in a conventional facility with a 12-h night/12-h day cycle with free access to chow diet pellet and water. For the experiments, 22-23 week-old skeletal muscle-specific HSA-PGC-1 α knockout (MKO) male mice and 8 week-old PGC-1 α muscle-specific transgenic (TG) male mice were used as previously described (6-8). All experiments were performed according to the criteria outlined for the care and use of laboratory animals and with approval of the veterinary office of the canton Basel and the Swiss authorities.

337

338 Treadmill running

Treadmill running was performed with the TG mice on the Columbus Instruments motorized treadmill with an electric shock grid. The mice were acclimatized to the treadmill and then let run till exhaustion. The running protocol is as follows: 10m/min for 5min with an increase by 2m/min every 5min until 26m/min and an inclination of 5 degrees. The speed of 26m/min was kept until exhaustion of the mice (7, 28, 29). Mice were killed and tissues were collected 3h after exercise.

345

346 **RNA isolation of muscle tissue**

347 Gastrocnemius and quadriceps were used to isolate RNA by TRIzol[®] according to the TRIzol[®]

348 reagent RNA isolation protocol (Invitrogen).

349

351 RESULTS

352 Broad recruitment of PGC-1α to the mouse genome

353 PGC-1a-dependent gene transcription has been studied in many different experimental 354 contexts. In isolation, gene expression arrays however are unable to distinguish direct from indirect targets, or to reveal the genomic sites where PGC-1 α is recruited to enhancer and 355 356 promoter elements, i.e. by coactivating TFs that directly bind to the DNA. Thus, we first 357 performed chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) of PGC-358 1α in differentiated C2C12 mouse myotubes to identify the locations where PGC- 1α is bound 359 to the genome. To identify genomic regions that are significantly enriched in the ChIP, we 360 slid a 200 bp window across the genome comparing the local ChIP read density with the read density from a background whole cell extract sample. We selected all regions with a Z-361 362 statistic larger than 4.5 as significantly enriched (FDR 0.6%, Suppl. Fig. S1A). Using this stringent cutoff, we identified 7512 binding regions for PGC-1 α via interaction with a TF 363 364 genome-wide, which include binding regions in the promoters of known PGC-1 α target genes (Fig. 1A) such as medium-chain specific acyl-CoA dehydrogenase (Acadm) and 365 cytochrome c (Cycs) (30, 31). The enrichment of immunoprecipitated DNA fragments from 366 367 the ChIP-Seq was validated for these and other PGC-1 α target genes by semiquantitative real-time PCR (Fig. 1B). In absolute terms, the distribution of the ChIP-Seq peaks revealed 368 that PGC-1 α is mostly recruited at distal sites from the assigned targets and, to a lesser 369 370 extent, to proximal regions of the gene or within an intronic sequence (Fig. 1C). However, 371 when compared to randomly selected DNA regions of equal size and number, PGC-1 α 372 binding peaks occur twice as often within 10 kb upstream of the transcription start site (TSS).

373 In parallel to the ChIP-Seg experiment, we furthermore analyzed gene expression patterns in differentiated muscle cells both in control condition and under PGC-1 α over-expression. 374 375 Using a reference set of mouse promoters (19) and associating microarray probes to promoters by mapping to known transcripts, we found 1566 promoters (corresponding to 376 377 984 genes) to be significantly up-regulated (log2 fold change >= 1; Z score >= 3) and 1165 promoters (corresponding to 727 genes) to be significantly down-regulated (log2 fold change 378 <= -1; Z score <= -3). Thus, similar to previous reports, PGC-1 α induced and repressed the 379 380 transcription of almost the same number of genes, respectively, indicating that the 381 physiological function of PGC-1 α includes both the activation and inhibition of substantial 382 numbers of genes.

383 To combine the DNA binding results from the ChIP-Seq with the data of the gene expression 384 arrays, we then assigned ChIP-Seq peaks to the closest promoter (and the associated gene) within a maximum distance of 10 kb. In this way, about 30% of all peaks (2295 of 7512) could 385 be associated with a target promoter. Inversely, for about 35% of all significantly up-386 387 regulated genes (341 of 984), a PGC-1 α binding peak is found within 10 kb of the promoter. 388 Since some of the up-regulated promoters may be regulated by more distal peaks, this is 389 only a lower bound on the fraction of genes that are directly regulated. In stark contrast, 390 only about 5% of all repressed genes harbor one or more PGC-1a DNA recruitment peaks in their vicinity (36 of 727) opposed to 95% indirectly down-regulated PGC-1 α target genes 391 (691 genes) (Fig. 1D). Moreover, the distribution of the distances between PGC-1 α peaks and 392 their associated promoters revealed a tight cluster of 532 peaks close to promoter regions 393 394 for up-regulated, direct PGC-1 α target genes (Fig. 1E) whereas the distribution of the 43 peaks associated to down-regulated genes was much wider, raising the possibility that the 395 396 association of peaks to transcriptionally repressed genes might be spurious (Fig. 1F). In summary, the strong enrichment of binding peaks near up-regulated genes and the almost 397 398 complete absence of binding peaks near down-regulated genes suggest that direct 399 regulation of transcription by PGC-1 α is almost exclusively activating. We note that there is a 400 large fraction of binding peaks (75%) that are associated to target genes that do not 401 significantly alter their expression. These peaks may have been wrongly assigned, their 402 functionality may be dependent on additional factors not active in these cells, or they may simply be spurious binding events that are not functional. 403

We next used this stratification of peaks and genes to study whether direct (i.e. with an 404 associated binding peak) and indirect PGC-1 α target genes exert different biological function 405 and identified Gene Ontology (GO) terms that were over-represented in any of the four 406 407 categories. First, we observed that the most significantly enriched functional categories for directly and indirectly up-regulated genes were those related to mitochondria, oxidative 408 409 phosphorylation and energy production (Fig. 1G and Suppl. Fig. S1B). In contrast, GO analysis 410 of indirectly down-regulated PGC-1 α target genes revealed a high prevalence of terms related to inflammation and immune response (Fig. 1H and Suppl. Fig. S1C). Assuming that 411

the assignment of peaks to repressed genes is not spurious, the few directly repressed PGC1α targets exhibit an enrichment in functions related to muscle contraction, in particular for
genes that are linked to contractile and metabolic properties of glycolytic, fast-twitch muscle
fibers (Fig. 1H and Suppl. Fig. S1D), as would be expected from the observed shift from
glycolytic to oxidative fibers mediated by PGC-1α in muscle (6).

417

418 Modeling the direct and indirect gene regulatory effects of PGC-1α

As a next step, we rigorously modeled the effects of PGC-1 α on its target genes in terms of 419 420 the occurrence of TFBSs for a large collection of mammalian regulatory motifs. We previously introduced a general framework, called Motif Activity Response Analysis (MARA) 421 422 (24), for modeling the gene expression profiles as a linear function of the TFBSs occurring in 423 the promoters and unknown regulatory "activities" of each of the regulators. As detailed in 424 the Methods, we here extended MARA to incorporate information from the PGC-1 α ChIP-Seq data, with the aim of identifying which other TFs are involved in mediating both the 425 direct and indirect regulatory effects of PGC-1a. Specifically, for all "direct target" promoters 426 that were associated with a PGC-1 α binding peak, we modeled the expression of the 427 428 promoter in terms of the predicted TFBSs in the neighborhood of the binding peak, while for 429 "indirect target" promoters we modeled the promoter's expression in terms of the predicted 430 TFBSs in the proximal promoter region, according to the conventional MARA approach (Fig. 431 2A and 2B).

432 First, further supporting our analysis above, direct target promoters were almost exclusively up-regulated and only in a few exceptional cases reached statistical significance for PGC-1a-433 repressed transcripts (Fig. 2C). Among the direct motif activities, the ESRRA position weight 434 matrix was the top ranking motif with a Z score of 6.04 (Suppl. Fig. 2). The corresponding TF 435 436 estrogen-related receptor α (ERR α), an orphan nuclear receptor, has been extensively studied as a central binding partner for PGC-1 α in the regulation of mitochondrial gene 437 expression (30-32). To stratify the different motifs according to their predicted function, we 438 439 then divided all motifs into groups according to the behavior of both their direct and indirect 440 activity changes. Strikingly, all motifs exhibited one of only four different motif activity 441 patterns. First, 6 TFs (Suppl. Fig. S2) were predicted to positively regulate PGC-1 α target genes only in the presence of PGC-1 α (Fig. 2D). Second, we found 6 motifs (Suppl. Fig. S2) 442

443 with significantly up-regulated direct and indirect motif activities upon PGC-1a over-444 expression (Fig. 2E). To our surprise, ERRα was predicted to regulate PGC-1α target genes in this manner, even though in previous reports gene regulation by ERR α in the context of 445 activated PGC-1 α was suggested to be dependent on PGC-1 α coactivation (30-32). Third, we 446 found 13 motifs (Suppl. Fig. S2) that were predicted to regulate PGC-1 α target genes, 447 448 however only in the absence of PGC-1 α (Fig. 2F). Fourth, there was a group of 28 motifs 449 (Suppl. Fig. S2) that showed a significant decrease of indirect motif activity upon PGC-1 α over-expression, but no significant change of their direct motif activity, including NFkB (Fig. 450 2G), a central regulator of inflammation which is indirectly repressed by PGC-1 α (33). 451 452 Intriguingly however, no motif was found that showed significant direct repression of target 453 genes, reinforcing the hypothesis that PGC-1 α -dependent gene repression is an indirect 454 event.

455

456 Nuclear receptors and activator protein-1-like leucine zipper proteins are the main 457 functional partners of PGC-1α in muscle cells

As a next step, we analyzed the occurrence of TF DNA-binding motifs in the PGC-1 α peaks 458 459 identified by ChIP-Seq. We first performed *de novo* motif prediction on the top 200 peaks, 460 using PhyloGibbs (26). As shown in Figure 3A, the motif that PhyloGibbs identified matches significantly (E-value = 7.7834e-10, as calculated by STAMP (27)) the canonical ESRRA motif. 461 In addition to the *de novo* prediction, we also used the same collection of 190 mammalian 462 463 regulatory motifs used by MARA (19) to check which known TF DNA-binding motifs were significantly over-represented in the PGC-1 α peaks relative to a set of background regions. 464 Many of the most significantly enriched motifs represent variations of nuclear receptor 465 binding sequences that are based on the "AG $^{T}/_{G}TCA$ " core hexamer and occur either alone or 466 467 in direct, inverted or everted repeats with variable spacing (Fig. 3B). Of these, the most significantly enriched motif was ESRRA, which is present in ~20% of all peaks. Moreover, 468 469 among all genes with at least one associated binding peak within 10Kb, ~28% are associated with a peak containing a predicted ERR α site. Interestingly, besides the nuclear receptor 470 motifs, we also found the DNA-binding element of the insulator protein CCCTC-binding 471 472 factor (CTCF), and a set of highly similar DNA elements sharing the FOS-JUN-like recognition 473 sequence "TGA^G/_cTCA" bound by the TFs BACH2, FOS, FOSB, FOSL1, JUN, JUNB, JUND, 474 FOSL2, NFE2, and NFE2L2 among the top 15 motifs enriched in PGC-1 α peaks (Fig. 3B).

475 The identity of the exact nuclear receptor binding partner that is bound at each peak is 476 difficult to deduce from DNA-binding motifs, since considerable promiscuity exists between 477 receptors and DNA-binding elements in different configurations of hexameric repeats (34). Moreover, non-nuclear receptor-like TFs are less well studied in the context of PGC-1a-478 479 controlled gene expression. Thus, to identify which regulatory motifs are most over-480 represented among peaks that do not contain nuclear receptor-like sites, we first manually 481 grouped all of the motifs with a sequence logo very similar to that of ESRRA. Next, we 482 discarded all peaks that had one or more predicted TFBSs for any of the motifs in this set. 483 With the remaining 3856 DNA sequences (51.33% of the peaks), we then again assessed the over-representation of each of the 190 mammalian regulatory motifs. In this analysis, 484 "TGA^G/_CTCA" recognition elements, hence FOS-JUN-like motifs, were the most significantly 485 486 enriched among these peaks (Fig. 3C). This result suggests that PGC-1 α peaks naturally fall into two classes: those containing ESRRA-like sites, and those containing sites for FOS-JUN-487 488 like motifs.

489 We then constructed a matrix N, whose elements N_{pm} contain the number of predicted 490 TFBSs for each motif m in each peak region p. We then performed principal component 491 analysis (PCA) on this site-count matrix to identify linear combinations of regulatory motifs that explain most of the variation in site-counts across the PGC-1 α peaks. The first two 492 493 components (out of 190 in total) clearly proved to be the most relevant ones, accounting for 494 10% and 9.6 % of the total variation in our dataset, respectively (Fig. 3D). Figure 3E shows the projection of all motifs on these first two principal components, with the names of the 495 496 motifs with the largest projection indicated in the figure. Whereas most motifs have projections close to zero along the first component, there is one group of motifs with strong 497 negative projections (ESRRA, NR1H4, NR5A1,2, NR6A1) and one group of motifs with strong 498 499 positive projections (BACH2, FOS_FOS(B,L1)_JUN(B,D), FOSL2, NFE2, NFE2L1, NFE2L2). These two sets of sites correspond precisely to the two classes of motifs identified above, 500 confirming that the most significant variation in TFBSs across PGC-1a peaks is caused by the 501 502 occurrence of either ESRRA-like motifs, or FOS-JUN-like motifs. Most interestingly, these two clusters of motifs reflect structurally distinct classes of TFs; the negatively scoring 503

eigenmotifs are characterized by binding of nuclear receptor-type zinc finger domains, while
the eigenmotifs with a positive score correspond to activator protein-1 (AP-1)-like leucine
zipper domains.

The second principal component corresponds to the strength of the binding signal for these 10 motifs, as confirmed by the robust negative correlation (r=-0.92) between the TFBSs posterior sum per peak and the peak's projection along the second principal component (Fig. 3F).

511

512 Validation of top scoring motifs reveals novel functional partners of PGC-1α

Our analysis identified a number of so-far uncharacterized TFs as potentially functional 513 partners for PGC-1 α -controlled gene expression in skeletal muscle cells. In order to 514 515 experimentally validate some of these candidates, we sorted all TFs by a number of criteria including TFBS over-representation in binding peaks, MARA activity upon PGC-1a over-516 expression, and the expression pattern of the TFs themselves. Table 1 shows the top 15 517 518 ranked TFs according to this selection. As expected, the well-known PGC-1a partner ERRa was identified as the most important factor. For our validation experiments, we chose the 519 next two motifs (FOS FOS(B,L1) JUN(B,D) and ZNF143, which is also known as ZFP143) as 520 well as three motifs from further down the list of the top 15 motifs (GTF2I, NFE2L2 and 521 NFYC). 522

FOS, the most up-regulated TF (log2 fold change = 1.78) among the TFs associated with the motif FOS_FOS(B,L1)_JUN(B,D), is a basic leucine zipper transcription factor known to heterodimerize with other leucine zipper proteins in order to form the AP-1 complex (35). The AP-1 complex furthermore contains JUN as well as ATF proteins. Thus, to dissect the function of the AP-1 protein complex, we also included JUN and ATF3, the most highly expressed isoforms of their respective protein families in muscle cells.

For each of these 7 TFs (ATF3, FOS, GTF2I, JUN, NFE2L2, NFYC and ZFP143), we selected a dozen target genes based on the Chi^2 score of the MARA prediction, presence of a PGC-1 α binding peak with at least one predicted binding site for the factor of interest, and at least a 2-fold induction upon over-expression of PGC-1 α . As summarized in Fig. 4 and Suppl. Fig. S3, siRNA-based knockdown of all TFs resulted in a robust reduction of the target mRNAs from - 534 40% to -75%. With the exception of NFYC and JUN, we found that the large majority of predicted target genes were down-regulated upon knockdown of the factor, confirming our 535 predictions (Fig. 4). The most consistent effects were observed for FOS and ZFP143 (all 536 targets down-regulated), followed by GTF2I (11 out of 12 down-regulated) and NFE2L2 and 537 ATF3 (10 out of 12 down-regulated). Interestingly, distinct target genes of the AP-1 complex 538 539 showed differential responsiveness to knockdown of the three AP-1 complex components 540 FOS, JUN and ATF3 (Fig. 4B, Fig. 4C and Fig. 4D). Similarly, PGC-1 α -mediated induction of a majority of the predicted target genes for NFE2L2 (Fig. 4E), ZFP143 (Fig. 4F) and GTF2I (Fig. 541 4G) was reduced upon knockdown of the respective TF when compared to the expression in 542 543 cells with overexpressed PGC-1 α and a scrambled siRNA control. Surprisingly, only 1 of the 544 11 predicted target genes for NFYC that have been chosen for validation was significantly repressed by siRNA-induced reduction of this TF (Fig. 4H), suggesting that other TFs may be 545 involved in mediating the regulatory effects of the NFYC regulatory motif. 546

547

Functional interaction between PGC-1α and different compositions of the AP-1 protein complex

550 Our targeted validation strategy revealed that PGC-1 α target genes predicted to be 551 regulated by the FOS-JUN-like motif react in distinct manners to siRNA-mediated knockdown of individual components of the AP-1 transcription factor protein complex. For example, 552 some genes only reacted to reduction of FOS (Fig. 5A), while others were responsive to the 553 554 knockdown of two (Fig. 5B) or even all three AP-1 protein partners (Fig. 5C) that we have tested using the siRNA-based approach. To further dissect the responsiveness of PGC-1 α 555 target genes to different AP-1 protein complexes, we performed global gene expression 556 arrays upon knockdown of each of the three TF components of the AP-1 complex. Fig. 5D 557 558 depicts the number of genes that were induced by PGC-1 α and that were, at the same time, down-regulated by the siRNA knockdown of any of the three AP-1 complex members. 559 560 Amongst a total of 477 genes, 89% responded to FOS knockdown, 52% to ATF3 knockdown, and 31% to JUN knockdown. Moreover, while 37% of all targets responded exclusively to 561 FOS, the fraction of targets responding exclusively to either JUN or ATF3 was at most 5%. 562 563 This analysis shows that, whereas different target genes respond differently to the knockdown of distinct AP-1 components, FOS is the dominant factor in determining AP-1function in these conditions.

566 As shown in Fig. 3B, 341 genes were associated to a PGC-1 α binding peak containing a 567 predicted site for the FOS-JUN-like motif bound by the AP-1 complex. Of these genes, the 568 expression of 55 was significantly induced by PGC-1 α over-expression in muscle cells. In our siRNA-based validation experiment, we found that 47 out of these 55 PGC-1 α -induced/AP-1 569 570 predicted targets were significantly down-regulated by knockdown of the AP-1 complex components and we called these genes "direct PGC-1 α /AP-1 targets". The remaining 430 571 572 genes out of 477 (Fig. 5D) were defined accordingly as "indirect PGC-1 α /AP-1 targets" that 573 lack a PGC-1 α peak containing a FOS-JUN-like motif, but still are regulated by PGC-1 α and the 574 AP-1 protein components (Fig. 5E). To reveal whether these gene categories exert distinct functions, GO and KEGG enrichment analyses were performed. Surprisingly, the 47 direct 575 PGC-1 α /AP-1 target genes showed a distinct and significant over-representation of the terms 576 577 "response to hypoxia" (GO ID: 0001666; adjusted p-value: 0.0247542) and "mTOR signaling pathway" (KEGG ID: mmu04150; adjusted p-value: 0.030674) that were absent in the GO 578 579 analysis of the remaining PGC-1 α /AP-1 targets (Fig. 5F). Recruitment of FOS to the same regulatory regions as PGC-1 α in the direct AP-1/PGC-1 α target genes was subsequently 580 581 validated by ChIP (Fig. 5G). These results suggest that AP-1, when interacting with PGC-1 α , 582 drives a synergic effect of response to hypoxia; on the other hand, when AP-1 and PGC-1 α act separately, and furthermore through downstream intermediate TFs, they regulate the 583 584 expression of genes involved in mitochondrial organization and energy metabolism.

585 Intriguingly, several of the predicted AP-1/PGC-1 α target genes are also under the control of PGC-1 α working with other transcription factors. For example, the vascular endothelial 586 growth factor (VEGF) or, based on the gene expression arrays, 8 OXPHOS genes seem 587 likewise to be under the control of AP-1 as well as ERR α in the context of elevated PGC-1 α in 588 skeletal muscle (31, 36). We therefore assessed the predicted and experimental overlap of 589 590 these two transcription factors in the regulation of AP-1/PGC-1 α target genes. Interestingly, when the PCA analysis of the PGC-1 α peaks was stratified in terms of eigenpeaks, we 591 observed two distinct groups of peaks associated with AP-1/PGC-1 α target genes (Fig. 5H). 592 593 First, some of these genes exclusively harbored peaks with FOS-JUN-like TFBSs, whereas the second group exhibited either peaks with both FOS-JUN- and ESRRA-like TFBSs, or a 594

combination of distinct peaks with either of these sites within 10 kb from their promoters (Fig. 5H). Next, we validated this prediction by investigating the change in expression of different AP-1/PGC-1α target genes in the context of reduced ERRα expression and function, elicited by a combination of shRNA-mediated knockdown and pharmacological treatment of muscle cells with the ERRα inverse agonist XCT790 (31). In line with the PCA, two distinct groups of ERRα inhibition-sensitive (Fig. 5I-K) and –insensitive (Fig. 5L-N) AP-1/PGC-1α target genes were found.

Finally, since all of the experiments were performed in differentiated myotubes in culture, 602 603 we assessed whether similar gene expression changes of the direct AP-1/PGC-1 α targets 604 involved in hypoxic gene regulation are also observed in skeletal muscle tissue of different 605 gain- (6) and loss-of-function mouse models (7, 8) in vivo. In skeletal muscle-specific PGC-1 α 606 knockout mice, the expression of several of these genes was reduced significantly (Fig. 6A-F). Surprisingly however, some of the predicted transcripts were not altered in this loss-of-607 608 function model for PGC-1 α , for example Nr0b2 (Fig. 6E). To further clarify the role of PGC-1 α 609 in the regulation of these genes, relative transcript levels were next assessed in musclespecific transgenic mice for PGC-1 α (Fig. 6G-L). In most cases, the genes with a reduction in 610 their transcription in the PGC-1 α muscle-specific knockout animals were inversely elevated 611 612 in the PGC-1 α muscle-specific transgenic mice. Moreover, some of these genes were likewise 613 induced by exercise (Fig. 6G-L) and at least in some cases, for example Twf2 and Nr0b2 (Fig. 6J and K), PGC-1 α overexpression and physical activity synergistically boosted gene 614 615 expression, for Nr0b2even in the absence of any effect of the muscle-specific PGC-1 α 616 transgene per se (Fig. 6K).

618 DISCUSSION

619 Exercise-induced skeletal muscle cell plasticity is a highly complex biological program that 620 involves the remodeling of a number of fundamental cellular properties. Since PGC-1 α 621 function has been strongly linked to the induction of an endurance-trained muscle 622 phenotype, we here dissected the PGC-1 α -controlled transcriptional network in muscle cells. First, our results reveal a broad recruitment of PGC-1 α to many different sites in the mouse 623 genome (7512 peaks), the majority of which were either not located within 10 kb distance 624 625 from a promoter or close to a gene that was not regulated by PGC-1 α over-expression at the 626 time of harvest of the cells, as has analogously been observed in many other ChIP-Seq 627 experiments (for example, see ref. 37). Apart from the fact that PGC-1 α could mediate long-628 range enhancer effects that were excluded in our peak-gene assignment, it is conceivable that PGC-1 α recruitment is transcriptionally silent in some binding peaks because it requires 629 630 the recruitment of additional cofactors for activation, which are not present in the 631 conditions or cell type in which our experiments were performed. In addition, it is possible 632 that a large fraction of PGC-1 α binding peaks may be "neutral" in the sense of not having any 633 direct role in regulating gene expression.

634 Second, while an almost equally strong effect of PGC-1 α on gene induction and repression 635 has been reported (31), our analysis now indicates that direct PGC-1a-mediated gene expression is restricted almost exclusively to positively regulated PGC-1 α target genes, 636 whereas the vast majority of gene repression is indirect, i.e. not associated with PGC-1 α 637 recruitment within a 10 kb distance to their promoters. Thus, the fact that almost 95% of all 638 639 repressed genes were not linked to PGC-1 α recruitment strongly implies that this coregulator primarily acts as a coactivator, and not as a corepressor as suggested by the data 640 of some studies (38-40). Importantly, indirect repression of PGC-1 α target genes was also 641 supported by the MARA prediction. The strong indirect inhibition of genes, many of which 642 are involved in inflammatory processes, is predicted by MARA to be mediated by TFs such as 643 644 NFkB and IRF factors. Such an indirect inhibition of NFkB and pro-inflammatory genes by PGC-1 α in muscle cells has been reported previously (33). 645

646 One of the main functions of PGC-1 α in all cells and organs is to boost mitochondrial gene 647 transcription and oxidative metabolism. Accordingly, we observed that Gene Ontology terms 648 related to these pathways were highly enriched when analyzing positively regulated PGC-1 α 649 target genes in muscle cells. Based on previous studies, the regulation of this core function could have been assigned to the direct interaction of PGC-1a and ERRa binding to regulatory 650 elements of these genes (31, 32). Surprisingly, our data indicate that many of the genes that 651 are involved in oxidative metabolic pathways are indirectly controlled by PGC-1 α and, hence, 652 653 do not require PGC-1 α recruitment to enhancer and promoter elements. Likewise 654 unexpectedly, the MARA analysis implies ERRa action on direct and indirect PGC-1a-induced 655 target genes, i.e. in the presence or absence of PGC-1 α coactivation. Thus, while these observations might obviously reflect a temporally distinct control of different PGC-1a target 656 genes that is not represented in our simultaneous analysis of DNA binding and gene 657 658 expression at one time point, it is conceivable that PGC-1 α acts primarily as an upstream 659 regulator of other factors that are subsequently controlling more downstream PGC-1 α target 660 genes without direct involvement of PGC-1a itself.

In skeletal muscle, PGC-1 α has been reported to interact with ERRs, PPARs and other nuclear 661 662 receptors, as well as myocyte enhancer and nuclear respiratory factors to mediate transcriptional regulation (3). Accordingly, ERRa and other nuclear receptor binding motifs 663 664 were amongst the most highly significant binding elements in our present report. Importantly however, we also predict a number of so-far unknown TFs to functionally 665 666 interact with PGC-1a and thereby contribute to PGC-1a-controlled gene expression in 667 skeletal muscle. Since a complete functional validation of all new putative TF partners is beyond the scope of this manuscript, we combined the high-throughput results with several 668 669 computational analyses (see Table 1) to select and test some of the potentially most 670 important factors together with predicted target genes. Notably, in siRNA-based knockdown experiments, we could show that depletion of FOS and its putative AP-1 multimerization 671 672 partners JUN and ATF3 as well as NFE2L2, ZFP143 and GTF2I in muscle cells reduced the ability of PGC-1 α to positively regulate target genes. Second, we could provide evidence of a 673 co-recruitment of FOS and PGC-1 α to the same regulatory sites in the vicinity of AP-1/PGC-674 675 1α target genes, confirming a functional interaction between these TFs and PGC- 1α . Thus, our results indicate that the coactivation repertoire of PGC-1 α in muscle exceeds the 676 prediction of previous studies by far. For example, even in our list of the top 15 motifs, 677 678 several predicted TFs have not yet been investigated in the context of PGC-1 α -controlled 679 gene expression, including BPTF, FOSL2, REST or RREB1. Future studies will aim at a more 680 detailed dissection of the global functional consequences of PGC-1 α coactivation of these 681 TFs in muscle cells.

682 Curiously, almost all of our analyses, and in particular the principal component analysis, 683 highlighted the relevance of FOS-JUN-like motifs. In fact, the largest amount of variation in 684 TFBS occurrence within PGC-1 α binding peaks results from either ESRRA-like or FOS-JUN-like motifs. The FOS-JUN-like motif, in particular, embodies the main binding elements of the AP-685 686 1 complex, which consists of different configurations of FOS, JUN, ATF and MAF proteins (35, 41). Our data comparing gene expression in cells with reduced FOS, JUN and ATF3 levels 687 688 indicate that PGC-1 α functionally interacts with the AP-1 complex in different configurations 689 in the regulation of specific genes. The differential requirement observed for distinct AP-1 690 components might provide an additional layer of control for specific PGC-1a target gene regulation. 691

AP-1 function itself is regulated by a variety of stimuli, including cytokines, growth factors 692 693 and stress, and subsequently controls a number of cellular processes including apoptosis, 694 cell proliferation and differentiation, stress response and hypoxia (41, 42). Mechanistically, we classified PGC-1 α -induced/AP-1-knocked-down targets in either direct or indirect genes. 695 696 Most interestingly, functional analysis of these two groups of genes revealed that when AP-1 697 and PGC-1 α act disjointedly, they are involved in the regulation of mitochondrial and other metabolic genes while, when coactivated by PGC-1 α , AP-1 distinctly alters the expression of 698 genes that are enriched in the ontology terms "response to hypoxia" and "mTOR signaling" 699 700 (Fig. 5F). Intriguingly, a closer analysis of all 47 direct AP-1/PGC-1 α target genes revealed 24 701 genes that are induced by hypoxia, are effectors of hypoxia or attenuate the detrimental 702 consequences of hypoxia (Fig. 6M). For example, several inhibitors of the mTOR signaling pathways are included in this group of genes and hypoxia has been described as a 703 suppressor of mTORC1 activity (43). Another group of genes contributes to the reduction of 704 705 cellular stress, detrimental metabolites, reactive oxygen species and increase in cellular 706 survival to reduce potential harmful consequences of prolonged hypoxia (44). Furthermore, several genes promote endothelial regeneration, vascular remodeling and vascularization 707 708 (45). In this context, PGC-1 α has previously been shown to promote VEGF-induced 709 angiogenesis in skeletal muscle in a hypoxia-inducible factor 1α (HIF- 1α)-independent, ERR α dependent manner (36). Similarly, PGC-1 α regulates the hypoxic response of brown fat (46), 710

711 neuronal and endothelial cells (47) even though the mechanisms of cellular protection 712 exerted by PGC-1 α in these experimental contexts have not been elucidated. Our findings 713 now indicate that, to ensure adequate oxygen and nutrient supplies for oxidative metabolism in skeletal muscle cells, PGC-1a might coordinate metabolic needs through 714 ERRa-induced Vegf expression with a broad, stress-induced AP-1-dependent hypoxia 715 716 program. Such a functional convergence was found for a subset of the direct AP-1/PGC-1 α 717 target genes that likewise seem to be under the control of ERR α together with PGC-1 α (Fig. 5H and I-K). Inversely, for the complementary subset of these genes, the functional 718 719 interaction between AP-1 and PGC-1 α seems distinct from the ERR α -dependent PGC-1 α 720 target gene regulation. Finally, in vivo evidence supports our muscle cell cultured-based 721 prediction, considering that many of the AP-1/PGC-1 α hypoxia-related target genes exhibit reduced and elevated transcript levels in PGC-1 α muscle-specific knockout and transgenic 722 723 animals, respectively. As previously demonstrated for VEGF and skeletal muscle 724 vascularization (36), many aspects of the phenotypic consequences of exercise-induced 725 muscle hypoxia occur in the muscle-specific transgenic mice even in the absence of physical 726 activity. In extension of these studies, we now however found additional genes involved in this process that show an additional, or in case of Nr0b2, even an exclusive synergistic 727 728 activation by exercise in the PGC-1 α transgenic animals. Thus, combined with previous 729 descriptions of muscle plasticity in these mice post-exercise in regard to insulin sensitivity 730 (29), our present findings reiterate the importance of bona fide exercise even in a genetic 731 model for endurance training such as the PGC-1 α muscle-specific transgenic animals.

732 In summary, our data provide a first insight into the transcriptional network controlled by PGC-1a in muscle cells. While one other study of global DNA recruitment of PGC-1a has 733 734 been performed in the human hepatoma cell line HepG2 (48), our results highlight the importance of combining ChIP-Seq experiment, transcriptional data together with a 735 comprehensive computational modeling approach and experimental validation of predicted 736 737 key regulators, in order to be able to discover mechanistic as well as functional outcomes of such a network. Combined with the knowledge of transcriptional regulation, 738 739 posttranslational modifications, alternative splicing and recruitment of different chromatin 740 remodeling protein complexes, a scenario can thus be conceived in which PGC-1 α is able to control and integrate different signaling pathways using a multitude of different 741 742 transcription factor binding partners (10, 11). A better understanding of such regulatory

- networks will eventually allow the targeting of whole biological programs or specific sub-
- 744 modules in pathological states of disregulation.

746 ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the ChIP-seq and geneexpression array data reported in this paper is GSE51191.

749

750 ACKNOWLEDGMENTS

We would like to thank Dr. Anastasia Kralli, Svenia Schnyder, Gesa Santos, Kristoffer 751 752 Svensson and Markus Beer for reagents, help and input for the preparation of this 753 manuscript. This project was funded by the Swiss National Science Foundation 754 (31003A_135397 to EvN, 310030_132900 to CH), SystemsX.ch (CellPlasticity, StoNets, and BrainstemX research projects to EvN), the Swiss Society for Research on Muscle Diseases 755 756 (SSEM), the Neuromuscular Research Association Basel (NeRAB), the Gebert-Rüf Foundation 757 "Rare Diseases" Program, the University of Basel and the Biozentrum. SS was supported by an IPhD fellowship of the SystemsX.ch Swiss Initiative in Systems Biology. We are thankful of 758 759 the [BC]2 Basel Computational Biology Center for providing computational resources.

760

761 **Disclosure declaration:** The authors have no conflict of interest in regard to this manuscript.

762 **REFERENCES**

 Handschin C, Spiegelman BM. 2008. The role of exercise and PGC1alpha in inflammation and chronic disease. Nature 454:463-469.

Pedersen BK, Febbraio MA. 2012. Muscles, exercise and obesity: skeletal muscle as a
 secretory organ. Nat Rev Endocrinol 8:457-465.

Handschin C. 2010. Regulation of skeletal muscle cell plasticity by the peroxisome
 proliferator-activated receptor gamma coactivator 1alpha. J Recept Signal Transduct
 Res 30:376-384.

Handschin C, Spiegelman BM. 2006. Peroxisome proliferator-activated receptor
 gamma coactivator 1 coactivators, energy homeostasis, and metabolism. Endocr Rev
 27:728-735.

Finck BN, Kelly DP. 2006. PGC-1 coactivators: inducible regulators of energy metabolism in health and disease. J Clin Invest 116:615-622.

- Lin J, Wu H, Tarr PT, Zhang CY, Wu Z, Boss O, Michael LF, Puigserver P, Isotani E,
 Olson EN, Lowell BB, Bassel-Duby R, Spiegelman BM. 2002. Transcriptional co activator PGC-1 alpha drives the formation of slow-twitch muscle fibres. Nature
 418:797-801.
- 779 7. Handschin C, Chin S, Li P, Liu F, Maratos-Flier E, Lebrasseur NK, Yan Z, Spiegelman
 780 BM. 2007. Skeletal muscle fiber-type switching, exercise intolerance, and myopathy
 781 in PGC-1alpha muscle-specific knock-out animals. J Biol Chem 282:30014-30021.
- Handschin C, Choi CS, Chin S, Kim S, Kawamori D, Kurpad AJ, Neubauer N, Hu J,
 Mootha VK, Kim YB, Kulkarni RN, Shulman GI, Spiegelman BM. 2007. Abnormal
 glucose homeostasis in skeletal muscle-specific PGC-1alpha knockout mice reveals
 skeletal muscle-pancreatic beta cell crosstalk. J Clin Invest 117:3463-3474.

786 9. Lonard DM, O'Malley BW. 2006. The expanding cosmos of nuclear receptor
 787 coactivators. Cell 125:411-414.

- Lonard DM, O'Malley B W. 2007. Nuclear receptor coregulators: judges, juries, and
 executioners of cellular regulation. Mol Cell 27:691-700.
- 790 11. Spiegelman BM, Heinrich R. 2004. Biological control through regulated
 791 transcriptional coactivators. Cell 119:157-167.
- Puigserver P, Adelmant G, Wu Z, Fan M, Xu J, O'Malley B, Spiegelman BM. 1999.
 Activation of PPARgamma coactivator-1 through transcription factor docking. Science
 286:1368-1371.
- Wallberg AE, Yamamura S, Malik S, Spiegelman BM, Roeder RG. 2003. Coordination
 of p300-mediated chromatin remodeling and TRAP/mediator function through
 coactivator PGC-1alpha. Mol Cell 12:1137-1149.
- 14. Li S, Liu C, Li N, Hao T, Han T, Hill DE, Vidal M, Lin JD. 2008. Genome-wide
 coactivation analysis of PGC-1alpha identifies BAF60a as a regulator of hepatic lipid
 metabolism. Cell Metab 8:105-117.

Lin J, Handschin C, Spiegelman BM. 2005. Metabolic control through the PGC-1
 family of transcription coactivators. Cell Metab 1:361-370.

Handschin C. 2009. The biology of PGC-1alpha and its therapeutic potential. Trends
Pharmacol Sci 30:322-329.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
 alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

 Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205-217.

- Perez-Schindler J, Summermatter S, Santos G, Zorzato F, Handschin C. 2013. The
 transcriptional coactivator PGC-1alpha is dispensable for chronic overload-induced
 skeletal muscle hypertrophy and metabolic remodeling. Proc Natl Acad Sci U S A
 110:20314-20319.
- Arnold P, Erb I, Pachkov M, Molina N, van Nimwegen E. 2012. MotEvo: integrated
 Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple
 alignments of DNA sequences. Bioinformatics 28:487-494.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier
 L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C,
 Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J.
 2004. Bioconductor: open software development for computational biology and
 bioinformatics. Genome Biol 5:R80.
- 821 22. R Development Core Team. 2012. R: A language and environment for statistical
 822 computing. R Foundation for Statistical Computing.
- Al-Shahrour F, Diaz-Uriarte R, Dopazo J. 2004. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 20:578-580.
- 826 24. Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, Katayama S, Schroder K, Carninci P, Tomaru 827 Y, Kanamori-Katayama M, Kubosaki A, Akalin A, Ando Y, Arner E, Asada M, Asahara 828 H, Bailey T, Bajic VB, Bauer D, Beckhouse AG, Bertin N, Bjorkegren J, Brombacher F, 829 Bulger E, Chalk AM, Chiba J, Cloonan N, Dawe A, Dostie J, Engstrom PG, Essack M, 830 831 Faulkner GJ, Fink JL, Fredman D, Fujimori K, Furuno M, Gojobori T, Gough J, Grimmond SM, Gustafsson M, Hashimoto M, Hashimoto T, Hatakeyama M, Heinzel 832 833 S, Hide W, Hofmann O, Hornquist M, Huminiecki L, Ikeo K, Imamoto N, Inoue S, 834 Inoue Y, Ishihara R, Iwayanagi T, Jacobsen A, Kaur M, Kawaji H, Kerr MC, Kimura R, 835 Kimura S, Kimura Y, Kitano H, Koga H, Kojima T, Kondo S, Konno T, Krogh A, Kruger 836 A, Kumar A, Lenhard B, Lennartsson A, Lindow M, Lizio M, Macpherson C, Maeda N, 837 Maher CA, Maqungo M, Mar J, Matigian NA, Matsuda H, Mattick JS, Meier S, Miyamoto S, Miyamoto-Sato E, Nakabayashi K, Nakachi Y, Nakano M, Nygaard S, 838 Okayama T, Okazaki Y, Okuda-Yabukami H, Orlando V, Otomo J, Pachkov M, 839 Petrovsky N, Plessy C, Quackenbush J, Radovanovic A, Rehli M, Saito R, Sandelin A, 840 841 Schmeier S, Schonbach C, Schwartz AS, Semple CA, Sera M, Severin J, Shirahige K, 842 Simons C, St Laurent G, Suzuki M, Suzuki T, Sweet MJ, Taft RJ, Takeda S, Takenaka Y, Tan K, Taylor MS, Teasdale RD, Tegner J, Teichmann S, Valen E, Wahlestedt C, 843 844 Waki K, Waterhouse A, Wells CA, Winther O, Wu L, Yamaguchi K, Yanagawa H, 845 Yasuda J, Zavolan M, Hume DA, Arakawa T, Fukuda S, Imamura K, Kai C, Kaiho A, 846 Kawashima T, Kawazu C, Kitazume Y, Kojima M, Miura H, Murakami K, Murata M, 847 Ninomiya N, Nishiyori H, Noma S, Ogawa C, Sano T, Simon C, Tagami M, Takahashi Y, Kawai J, Hayashizaki Y. 2009. The transcriptional network that controls growth 848 849 arrest and differentiation in a human myeloid leukemia cell line. Nat Genet 41:553-850 562
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H,
 Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller
 W, Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm,
 and yeast genomes. Genome Res 15:1034-1050.
- 855 26. Siddharthan R, Siggia ED, van Nimwegen E. 2005. PhyloGibbs: a Gibbs sampling
 856 motif finder that incorporates phylogeny. PLoS Comput Biol 1:e67.

- 857 27. Mahony S, Benos PV. 2007. STAMP: a web tool for exploring DNA-binding motif
 858 similarities. Nucleic Acids Res 35:W253-258.
- 859 28. Summermatter S, Santos G, Perez-Schindler J, Handschin C. 2013. Skeletal muscle
 860 PGC-1alpha controls whole-body lactate homeostasis through estrogen-related
 861 receptor alpha-dependent activation of LDH B and repression of LDH A. Proc Natl
 862 Acad Sci U S A 110:8738-8743.
- Summermatter S, Shui G, Maag D, Santos G, Wenk MR, Handschin C. 2013. PGC1alpha improves glucose homeostasis in skeletal muscle in an activity-dependent
 manner. Diabetes 62:85-95.
- Huss JM, Torra IP, Staels B, Giguere V, Kelly DP. 2004. Estrogen-related receptor
 alpha directs peroxisome proliferator-activated receptor alpha signaling in the
 transcriptional control of energy metabolism in cardiac and skeletal muscle. Mol Cell
 Biol 24:9079-9091.
- Mootha VK, Handschin C, Arlow D, Xie X, St Pierre J, Sihag S, Yang W, Altshuler D,
 Puigserver P, Patterson N, Willy PJ, Schulman IG, Heyman RA, Lander ES,
 Spiegelman BM. 2004. Erralpha and Gabpa/b specify PGC-1alpha-dependent
 oxidative phosphorylation gene expression that is altered in diabetic muscle. Proc
 Natl Acad Sci U S A 101:6570-6575.
- Schreiber SN, Emter R, Hock MB, Knutti D, Cardenas J, Podvinec M, Oakeley EJ,
 Kralli A. 2004. The estrogen-related receptor alpha (ERRalpha) functions in
 PPARgamma coactivator 1alpha (PGC-1alpha)-induced mitochondrial biogenesis. Proc
 Natl Acad Sci U S A 101:6472-6477.
- 879 33. Eisele PS, Salatino S, Sobek J, Hottiger MO, Handschin C. 2013. The peroxisome
 880 proliferator-activated receptor gamma coactivator 1alpha/beta (PGC-1) coactivators
 881 repress the transcriptional activity of NF-kappaB in skeletal muscle cells. J Biol Chem
 882 288:2246-2260.
- 883 34. Mangelsdorf DJ, Evans RM. 1995. The RXR heterodimers and orphan receptors. Cell
 884 83:841-850.
- Hai T, Curran T. 1991. Cross-family dimerization of transcription factors Fos/Jun and
 ATF/CREB alters DNA binding specificity. Proc Natl Acad Sci U S A 88:3720-3724.
- Arany Z, Foo SY, Ma Y, Ruas JL, Bommi-Reddy A, Girnun G, Cooper M, Laznik D,
 Chinsomboon J, Rangwala SM, Baek KH, Rosenzweig A, Spiegelman BM. 2008. HIF independent regulation of VEGF and angiogenesis by the transcriptional coactivator
 PGC-1alpha. Nature 451:1008-1012.
- 37. Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. 2011. Sequence-specific
 regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm
 fates. Nat Struct Mol Biol 18:120-127.
- 38. Qian J, Chen S, Huang Y, Shi X, Liu C. 2013. PGC-1alpha regulates hepatic hepcidin
 expression and iron homeostasis in response to inflammation. Mol Endocrinol
 27:683-692.
- 39. Jang WG, Kim EJ, Park KG, Park YB, Choi HS, Kim HJ, Kim YD, Kim KS, Lee KU, Lee IK.
 2007. Glucocorticoid receptor mediated repression of human insulin gene expression is regulated by PGC-1alpha. Biochem Biophys Res Commun 352:716-721.
- Sandri M, Lin J, Handschin C, Yang W, Arany ZP, Lecker SH, Goldberg AL, Spiegelman
 BM. 2006. PGC-1alpha protects skeletal muscle from atrophy by suppressing FoxO3
 action and atrophy-specific gene transcription. Proc Natl Acad Sci U S A 103:16260 16265.

- 904 41. Shaulian E, Karin M. 2002. AP-1 as a regulator of cell life and death. Nat Cell Biol
 905 4:E131-136.
- 906 42. Curran T, Franza BR, Jr. 1988. Fos and Jun: the AP-1 connection. Cell 55:395-397.
- 43. Cam H, Easton JB, High A, Houghton PJ. 2010. mTORC1 signaling under hypoxic conditions is controlled by ATM-dependent phosphorylation of HIF-1alpha. Mol Cell
 40:509-520.
- 910 44. **Majmundar AJ, Wong WJ, Simon MC.** 2010. Hypoxia-inducible factors and the 911 response to hypoxic stress. Mol Cell **40:**294-309.
- 45. Wagner PD. 2001. Skeletal muscle angiogenesis. A possible role for hypoxia. Adv Exp
 Med Biol 502:21-38.
- 914 46. Pino E, Wang H, McDonald ME, Qiang L, Farmer SR. 2012. Roles for peroxisome
 915 proliferator-activated receptor gamma (PPARgamma) and PPARgamma coactivators
 916 1alpha and 1beta in regulating response of white and brown adipocytes to hypoxia. J
 917 Biol Chem 287:18351-18358.
- 918 47. Zhao J, Li L, Pei Z, Li C, Wei H, Zhang B, Peng Y, Wang Y, Tao Y, Huang R. 2012.
 919 Peroxisome proliferator activated receptor (PPAR)-gamma co-activator 1-alpha and
 920 hypoxia induced factor-1alpha mediate neuro- and vascular protection by hypoxic
 921 preconditioning in vitro. Brain Res 1447:1-8.
- 48. Charos AE, Reed BD, Raha D, Szekely AM, Weissman SM, Snyder M. 2012. A highly
 integrated and complex PPARGC1A transcription factor binding network in HepG2
 cells. Genome Res 22:1668-1679.

925

927 FIGURE LEGENDS

928 Figure 1. Genome-wide DNA recruitment of PGC-1α in mouse muscle cells.

929 (A) PGC-1α ChIP-Seq binding peaks (read densities) around the TSS of the genes *Acadm* and
930 *Cycs* obtained from the UCSC Genome Browser.

931 (B) Real-time PCR validation of the ChIP enrichment measured at the promoter of a set of 932 PGC-1 α target genes. Bars represent fold enrichment over that of the *Tbp* intron, error bars 933 represent SEM. *p < 0.05; **p < 0.01; ***p < 0.001.

934 (C) Mapping ChIP-Seq PGC-1 α peaks across the genome. Transcription Start Site (TSS) and 935 Transcription End Site (TES) are relative to mm9 RefSeq transcripts. "Intergenic": \geq 10 kb 936 from the nearest transcript; "Upstream of TSS": -10 to 0 kb from the TSS; "Downstream of 937 TES": 0 to 10 kb from the TES. Numbers between brackets indicate, for each category, the 938 ratio between the percentage of PGC-1 α peaks and the percentage of the same number of 939 randomly distributed peaks.

940 (D) Histogram illustrating the number of direct and indirect genes either up- or down-941 regulated by over-expression of PGC-1 α in muscle cells. Direct genes are those associated to 942 promoters found within ±10 kb relative to the nearest peak.

943 (E) Distribution of the distances of 532 peaks from their associated up-regulated gene944 promoters.

945 (F) Distribution of the distances of 43 peaks from their associated down-regulated gene946 promoters.

947 (G-H) Subset of the top significantly enriched GO Biological Process terms identified for
948 directly and indirectly up-regulated (G) and down-regulated (H) PGC-1α target genes.

949

Figure 2. Four distinct mechanistic modes of action for gene expression regulated by PGC1α and TF partners.

952 (A) Classification of direct and indirect target genes in MARA (see Methods)

953 (B) Distribution of peak distance from the closest promoter and phastCons conservation954 score of the peak.

955 (C) Distribution of log2 expression values for all mouse promoters. Expression values were 956 averaged across the 3 GFP and the 3 PGC-1 α samples. Direct targets are depicted in red, 957 indirect targets in grey.

958 (D-G) Activity plot of the motifs ELF1,2,4 (D), ESRRA (E), REST (F) and NFKB1_REL_RELA (G) as
959 predicted by MARA (Motif Activity Response Analysis). Red: direct targets; green: indirect
960 targets.

961

962 Figure 3. PCA reveals FOS-JUN-like leucine zippers as a new class of putative functional 963 PGC-1α partners.

964 (A) Sequence logo of the top position weight matrix discovered *de novo* by PhyloGibbs in the
965 top 200 scoring peaks and of the corresponding canonical motif of ERRα as predicted by
966 STAMP.

967 (B) Top scoring results of motif search performed on all 7512 PGC-1α peaks with MotEvo.
968 Motifs depicted in red and blue correspond to the clusters identified by PCA in panel D.

969 (C) Top scoring results of motif search performed on the 3656 "non ESRRA-like" peaks with970 MotEvo.

971 (D) Fraction of explained variance of the top 10 PCA components.

972 (E) PCA analysis of the 7512 PGC-1α peaks. Eigenmotif scores across Principal Component 1

973 (PC1) and Principal Component 2 (PC2) are shown. Red and blue ellipses highlight motif

974 clusters, as identified by PC1, of nuclear hormone receptor-like zinc finger and FOS-JUN-like975 leucine zipper proteins, respectively.

976 (F) Correlation between Principal Component 2 scores and binding site posterior sum for

977 each peak relative to the top 10 PCA motifs. "r" refers to the Pearson correlation coefficient.

979 Figure 4. Validation of TFs associated with top scoring motifs reveals novel functional PGC980 1α partners

981 (A) siRNA-mediated knockdown efficiency for FOS. Bars represent fold induction over
982 GFP/siCtrl value, error bars represent SEM. *p < 0.05; **p < 0.01; ***p < 0.001. See also
983 Figure S4.

(B-H) qRT-PCR analysis of PGC-1α target genes whose associated peak contains at least one
binding site for the motif: FOS_FOS(B,L1)_JUN(B,D) (B-D), NFE2L2 (E), ZNF143 (F), GTF2I (G),
NFY(A,B,C) (H). Bars represent % change compared to PGC-1α/siCtrl values. Error bars
represent SEM. *p < 0.05; **p < 0.01; ***p < 0.001.

988

Figure 5. PGC-1α controls the hypoxia gene program via a functional interaction with
 different configurations of the AP-1 protein complex.

991 (A-C) qRT-PCR analysis of Cdk15, Nppb and Slc6a19 mRNA levels in response to PGC-1α over-

992 expression and either siFos, siJun or siAtf3 knockdown. Data are normalized to mRNA levels

993 in GFP infected cells. Error bars represent ± SEM. *p < 0.05; **p < 0.01; ***p < 0.001.

994 (D) Venn diagram illustrating the overlap in number of genes up-regulated by PGC-1 α and 995 down-regulated by either FOS, JUN or ATF3 knockdown.

996 (E) Histogram illustrating the number of direct and indirect PGC-1 α /AP-1 target genes.

(F) Subset of the top significantly enriched Gene Ontology and KEGG terms identified for thetwo gene groups illustrated in panel (E).

999 (G) qRT-PCR validation of the ChIP enrichment of c-Fos measured at the gene of *TGFB*1 1000 (validated) and at the promoters of *Nr0b2*, *Gprc5a* and *Dbt* (predicted) target genes. Bars 1001 represent fold enrichment over PGC-1 α exon2 set as 1. Error bars represent SEM. *p < 0.05; 1002 **p < 0.01; ***p < 0.001.

1003 (H) PCA analysis of the 7512 PGC-1 α peaks. Eigenpeak scores across Principal Component 1 1004 (PC1) and Principal Component 2 (PC2) are shown. Colored dots correspond to peaks 1005 associated to the 47 direct PGC-1 α /AP-1 targets. Blue dots refer to genes associated to

peaks containing only FOS-JUN TFBSs, while red dots refer to genes associated to peaks with 1006 1007 FOS-JUN and ESRRA TFBSs, either located in the same peak or in distinct PGC-1 α peaks. 1008 (I-K) gRT-PCR analysis of PGC-1 α /AP1 targets whose associated peaks contain an ESRRA binding site. The bars represent relative mRNA levels compared to AV-shGFP + AV-GFP + 1009 vehicle, which is set as 1. The error bars represent SEM. p < 0.05, p < 0.01, p < 0.01, p < 0.001. 1010 (L-N) qRT-PCR analysis of PGC- 1α /AP1 targets whose associated peaks (if any) do not contain 1011 1012 an ESRRA binding site. The bars represent relative mRNA levels compared to AV-shGFP + AV-GFP + vehicle, which is set as 1. The error bars represent SEM. *p < 0.05, **p < 0.01, ***p < 1013 1014 0.001.

1015

1016 Figure 6. PGC-1 α controls the hypoxic gene program in muscle in vivo. (A-F) qRT-PCR 1017 analysis of hypoxic genes in sedentary control (ctrl) and muscle-specific knockout mice 1018 (MKO). The control group is set as 1. Error bars represent SEM. *p < 0.05, **p < 0.01, ***p < 1019 0.001.

- 1020 (G-L) qRT-PCR analysis of hypoxic genes in treadmill running mice. Control (ctrl) and muscle-
- 1021 specific transgenic (TG) mice were used under sedentary and exercise conditions. The
- 1022 control group sedentary is set as 1. Error bars represent SEM. *p < 0.05, **p < 0.01, ***p <
- 1023 0.001. (M) Schematic representation depicting the downstream effects of the functional

1024 $\hfill interaction between PGC-1\alpha$ and the AP-1 complex in the context of the hypoxia gene

1025 program. Direct targets of PGC-1 α and AP-1 are indicated in bold.

1027 **TABLES**

1028 **Table 1. Global summary of all analyses performed on PGC-1** peaks. The final score is the

1029 count of all analyses where a certain motif passed the defined cutoffs. The motifs chosen for

1030 validation and their corresponding values which satisfied the cutoffs are shown in bold.

Motif name	PCA ^a	Over-repr.	Over-repr.	MARA	MARA	Log2FC	Abs. expr.	Final
		in all PGC-	in "non	activity Z	activity Z	in expr.	in PGC-1α	ranking
		1α peaks ^b	ESRRA-like"	score	score	array₫	sample ^e	
			peaks ^b	direct ^c	indirect ^c			
ESRRA	Yes	1	182	6.04(14.78)	15.49(37.94)	2.31	1829.45	6
FOS_FOS(B,L1)_JUN(B,D)	Yes	5	2	0.88(2.14)	1.81(-4.34)	1.78	1508.85	5
ZNF143		27	28	2.48(6.05)	4.65(9.68)	0.38	384.36	5
BPTF		21	12	1.38(3.37)	2.56(-6.25)	-0.56	333.34	4
ESR1		17	50	2.33(5.69)	4.53(11.04)	-0.47	232.42	4
FOSL2	Yes	6	3	0.88(2.14)	1.51(3.65)	-0.98	717.09	4
GTF2I		34	13	2.09(5.10)	2.38(-5.80)	-0.55	1207.81	4
NFE2L2	Yes	8	5	0.57(1.38)	1.01(-2.37)	-0.38	3673.63	4
NFY(A,B,C)		96	116	2.37(5.80)	3.56(7.62)	1.07	2409.48	4
NR5A1,2	Yes	3	188	3.53(8.66)	7.73(17.00)	-0.08	80.97	4
REST		12	6	0.48(1.15)	2.41(5.70)	-0.89	328.04	4
RREB1		15	10	1.56(3.82)	2.39(-5.42)	0.05	678.44	4
SP1		24	22	3.99(9.76)	0.61(0.33)	-0.32	751.98	4
STAT2,4,6		29	23	0.35(0.52)	4.81(-9.67)	-2.72	380.12	4
TLX13_NFIC(dimer)		19	17	0.84(-2.05)	4.91(-11.97)	-0.34	2339.33	4

1031

1032 ^a Requirement for PCA: being among the top 10 most contributing motifs to PC1 and PC2.

1033 ^b Requirement for motifs over-representation: being among the top 30 significant motifs; ranking position shown.

1034 ^c Requirement for MARA: have a Z-score \geq 2.0. Numbers between brackets show the difference between the PGC-1 α state and the GFP

1035 state, representing the direction in which the motif activity changes following PGC-1 α over-expression.

1036 d Requirement for the expression array (1): having a log2 fold change value \geq 1.0 (corresponding to 2 folds up-regulation)

1037 ^e Requirement for the expression array (2): having an absolute expression in the PGC-1 α sample \ge 100







Motif name	Z score	TFBSs count	Peaks percentage	Genes in the associated peaks	
ESRRA	96.95	1709	19.76%	461 (28.47%)	
BACH2	81.74	1300	16.01%	286 (17.67%)	
NR5A1,2	80.48	2007	22.58%	522 (32.24%)	
NR6A1	72.63	767	8.24%	200 (12.35%)	
FOS_FOS{B,L1}_JUN{B,D}	69.89	2030	19.18%	341 (21.06%)	
FOSL2	68.06	1360	16.05%	308 (19.02%)	
NFE2	62.78	1006	12.15%	229 (14.14%)	
NFE2L2	56.66	697	8.53%	174 (10.75%)	
HNF4A_NR2F1,2	47.20	766	9.15%	209 (12.91%)	
CTCF	47.10	188	2.46%	73 (4.51%)	
PPARG	40.61	389	4.30%	101 (6.24%)	
REST	40.00	98	0.71%	15 (0.93%)	
NR1H4	38.91	586	7.08%	156 (9.64%)	
RXRA_VDR{dimer}	35.02	364	4.26%	100 (6.18%)	
RREB1	34.23	180	1 78%	42 (2 59%)	

Motif name	Z score	TFBSs count	Peaks percentage
BACH2	59.05	712	16.96%
FOS_FOS{B,L1}_JUN{B,D}	51.92	1125	20.18%
FOSL2	49.84	724	16.55%
NFE2	46.72	548	12.76%
NFE2L2	39.79	372	8.84%
REST	35.42	71	0.93%
CTCF	33.65	97	2.46%
CDX1,2,4	27.37	218	4.59%
MZF1	23.94	228	4.95%
RREB1	22.80	108	2.02%
SPI1	21.90	218	5.50%
BPTF	21.88	278	4.80%
GTF2I	21.17	228	4.95%
PATZ1	20.97	162	3.92%
FOX{I1,J2}	20.47	211	4.33%

В





A





