

1 **Gene transcripts associated with muscle strength: a CHARGE meta-analysis of**
2 **7,781 persons**

3

4 LC Pilling*, R Joehanes*, T Kacprowski*, M Peters*, R Jansen*, D Karasik, DP Kiel, LW
5 Harries, A Teumer, J Powell, D Levy, H Lin, K Lunetta, P Munson, S Bandinelli, W Henley, D
6 Hernandez, A Singleton, T Tanaka, G van Grootheest, A Hofman, AG Uitterlinden, R Biffar,
7 S Gläser, G Homuth, C Malsch, U Völker, B Penninx#, JBJ van Meurs#, L Ferrucci#, T
8 Kocher#, J Murabito#, D Melzer#

9

10 * Equal first authors, # Equal senior authors

11

12

13 **Corresponding author:**

14 Professor David Melzer,
15 Epidemiology and Public Health,
16 University of Exeter Medical School,
17 Medical Research, RILD level 3,
18 RD&E Wonford, Barrack Road, Exeter, EX2 5DW, UK.
19 E-mail: D.Melzer@exeter.ac.uk
20 Phone: +44 (0) 1392 406751

21

22

23 **Keywords:**

24 Gene-expression, muscle, strength, blood, human, leukocyte

25

26

27 **Abstract**

28 **Background:** Lower muscle strength in midlife predicts disability and mortality in later life.
29 Blood-borne factors, including growth differentiation factor 11 (GDF11), have been linked to
30 muscle regeneration in animal models. We aimed to identify gene transcripts associated with
31 muscle strength in adults.

32 **Methods:** Meta-analysis of whole blood gene expression (overall 17,534 unique genes
33 measured by microarray) and hand-grip strength in four independent cohorts (n=7,781, ages:
34 20-104 years, weighted mean=56), adjusted for age, sex, height, weight, and leukocyte
35 subtypes. Separate analyses were performed in subsets (older/younger than 60,
36 male/female).

37 **Results:** Expression levels of 221 genes were associated with strength after adjustment for
38 cofactors and for multiple statistical testing, including *ALAS2* (rate limiting enzyme in heme
39 synthesis), *PRF1* (perforin, a cytotoxic protein associated with inflammation), *IGF1R* and
40 *IGF2BP2* (both insulin like growth factor related). We identified statistical enrichment for
41 hemoglobin biosynthesis, innate immune activation and the stress response. Ten genes
42 were only associated in younger individuals, four in males only and one in females only. For
43 example *PIK3R2* (a negative regulator of PI3K/AKT growth pathway) was negatively
44 associated with muscle strength in younger (<60 years) individuals but not older (>=60
45 years). We also show that 115 genes (52%) have not previously been linked to muscle in
46 NCBI PubMed abstracts

47 **Conclusions:** This first large-scale transcriptome study of muscle strength in human adults
48 confirmed associations with known pathways and provides new evidence for over half of the
49 genes identified. There may be age and sex specific gene expression signatures in blood for
50 muscle strength.

51

52 **Introduction**

53 Muscle strength correlates with health and physical function, and poor muscle strength in
54 midlife is a strong, independent predictor of health status decline and mortality over 25 years
55 (40). Sufficient muscle strength in the hands, arms and legs is needed for everyday
56 functioning; persons with poor strength are at high risk of disability, injury from falls, and
57 other age-related morbidities (15, 25).

58 Hand-grip is a frequently used summary measure of strength because it correlates well with
59 strength of other key muscles and is relatively easy to measure with high precision;
60 Bohannon *et al.* reported strong correlations between grip and knee extension strength
61 (Pearson R=0.77 to 0.8) in a sample aged 18 to 85 years (7), with Samson *et al.* reporting
62 similar estimates (43). Muscle strength (including grip strength) is a more important predictor
63 of mortality risk than muscle mass (9, 32), and grip strength (but not muscle mass) was
64 associated with poor physical functioning in older adults (51). The mechanisms underlying
65 the association between lower strength and mortality are not entirely clear, but a recent
66 large-scale multi-country follow-up study (n=142,861) reported that lower grip strength
67 associated most strongly with cardiovascular mortality (28). Current theories emphasize the
68 role of denervation not compensated by adequate re-innervation, mitochondrial dysfunction,
69 cellular senescence, inflammation, changes in microenvironment, and local skeletal changes,
70 among other factors (4, 44).

71 Studies of heterochronic parabiosis (connecting the blood circulations of young and old mice)
72 found that circulating factors, in particular lower GDF11 (growth and differentiation factor 11)
73 in older mice, explained the lower muscle regenerative capacity in older compared to
74 younger muscle (8, 11, 48). It is well established that circulating factors, such as pro-
75 inflammatory mediators and hormones (including testosterone), are strong correlates which
76 predict the slope of decline of muscle mass and strength in aging humans (26). These blood

77 borne factors may function as systemic regulators influencing muscle and may be different
78 from gene expression patterns in muscle itself.

79 Previous studies of transcriptome associations with muscle strength in humans were
80 conducted predominantly in muscle tissue, and are mostly limited by small sample size (30)
81 or focus on candidate genes (34). These studies can be susceptible to false negatives
82 statistical associations due to lack of power or coverage. A transcriptome-wide study of
83 whole blood transcript associations with grip strength conducted by the InCHIANTI Aging
84 Study (mean age 72 years, 71% ≥ 72 years old) found only one gene, *CEBPB*
85 (CCAAT/enhancer-binding protein beta, required for macrophage-mediated muscle repair in
86 a murine model (42)), to be associated with muscle strength in older humans after
87 adjustment for confounders and multiple testing (18). A follow-up study in humans found that
88 *CEBPB* expression increased following exercise-induced muscle damage (6). However,
89 because of the limited sample size of both studies, important transcriptional signals may
90 have been missed.

91 In the present study we sought to test associations between transcripts expressed in whole
92 blood and hand grip strength in multiple adult human cohorts. The majority of RNA in whole
93 blood samples is derived from immature erythrocytes and platelets (~70% from reticulocytes
94 and ~18% from reticulated platelets); however, these are predominantly globin-related and
95 are not actively transcribed by circulating cells (1). The remaining approximately 12% of
96 RNA is from circulating white blood cells of all types, driving non-globin related gene
97 expression. We have also performed subgroup analyses by age-group and gender, to check
98 for heterogeneity in the results. We used a robust meta-analysis framework within four
99 independent cohorts (n=7,781 participants) from the CHARGE (Cohorts for Heart and Aging
100 Research in Genomic Epidemiology) consortium (37) to identify the genes whose levels of
101 expression assessed by blood transcripts were associated with muscle strength.

102

103 **Methods**

104 **Study sample**

105 Characteristics of the cohorts are presented in **Table 1**. Complete data for the planned meta-
106 analysis were available for 7,781 participants from four cohorts; the Framingham Heart Study
107 (23) (FHS, n=5,576, ages=24-90), the InCHIANTI study (13) (n=667, ages=30-104), the
108 Rotterdam study (20) (RS, n=556, ages=46-89) and the Study of Health in Pomerania (52)
109 (SHIP, n=982, ages=20-81) (total n=7,781). The FHS study included two related generations
110 of participants (accounted for in the statistical methodology); the FHS Generation 2 (n=2,421,
111 ages=40-90) and Generation 3 (n=3,155, ages=24-78) cohorts were included. The
112 Netherlands Study of Depression and Anxiety (35) (NESDA, n=1,989, ages=18-65) is also
113 reported, but was not included in the discovery meta-analysis due to unavailable data on
114 white cell proportions necessary for the meta-analysis protocol. Overall the cohorts were
115 quite similar with respect to sex-distribution and sampling methods, differing only by age
116 distribution and lower mean hand-grip strength in the RS. Detailed study design and cohort
117 information has been previously published (21).

118 Four additional subset analyses were performed. The sub-samples available were 1) older
119 participants ≥ 60 years (n=2,402), 2) younger participants < 60 years (n=5,379), 3) male
120 participants (n=3,557), 4) female participants (n=4,224).

121

122 **Phenotype**

123 The primary phenotype was hand-grip strength in kg (a normally distributed phenotype). In
124 the FHS, hand grip strength was measured with a Jamar dynamometer with three trials
125 performed in each hand, and the maximum of the six trials for each participant was used in

126 the analysis. In InCHIANTI each participant recorded their maximum grip strength three
127 times in each hand, and the maximum recorded value of the six trials was used. In the
128 Rotterdam Study grip strength of the non-dominant hand was measured three times for each
129 participant, and the maximum recorded value was used. In the SHIP cohort participants were
130 asked to press the hand dynamometer firmly for several seconds, once per hand (left and
131 right), and the maximum value was used. Each participant in NESDA was measured twice
132 with a Jamar dynamometer in their dominant hand, with the maximum recorded value used.

133

134 **Peripheral gene expression data**

135 Blood samples were drawn from participants and RNA was isolated, reverse-transcribed to
136 cDNA, which was then amplified and hybridized to a microarray individually for each cohort;
137 methods described in detail (21). Briefly, the FHS used Affymetrix Human Exon 1.0 ST
138 GeneChips, characterizing the expression of 16,798 unique genes (after exclusion of
139 probesets with relative log expression mean values <3). The InCHIANTI and SHIP studies
140 used the Illumina HumanHT-12 v3 Expression BeadChip Kit, and the RS used the Illumina
141 HumanHT-12 v4 Expression BeadChip Kit, with 37,348 probes measured on both Illumina
142 platforms (22,911 unique genes; after exclusions of probes expressed above background in
143 <5% of participants this becomes 15,639 unique genes). These four studies in the primary
144 analysis all used PAXgene tubes to isolate and stabilize the RNA, thereby limiting the
145 technical variability between studies. Finally the NESDA cohort utilized the Affymetrix Human
146 Genome U219 Array, with expression information available on 18,212 unique gene
147 identifiers. Quantile normalization and log₂ transformation was performed on the gene
148 expression data in each cohort, and both probes and samples were z-transformed. Raw data
149 from gene expression profiling are available online (FHS [*NCBI dbGAP*: phs000363.v7.p8],
150 InCHIANTI [*GEO*: GSE48152], NESDA [*NCBI dbGAP*: phs000486.v1.p1], RS [*GEO*:
151 GSE33828] and SHIP-TREND [*GEO*: GSE36382]).

152 Systematically mapping pairs of probes to RefSeq transcripts (one Affymetrix Exon ST and
153 one Illumina HumanHT-12 probe) found 26,746 probe-pairs corresponding to 17,534 unique
154 RefSeq gene symbols. The assignment of a probe to one or more transcripts was performed
155 as described previously (45). For the Illumina arrays, the transcript sequences derived from
156 the 48,803 probe sequences provided in the Illumina annotation file (HumanHT-
157 12_V3_0_R3_11283641_A, version 3.0, 7/1/2010) were mapped against all available mRNA
158 sequences provided in the UCSC genome annotation database (version 06/30/2013) using
159 string matching. Altogether 29,818 probes were successfully mapped to one or more
160 validated mRNAs. Probes that could neither be mapped to a unique mRNA nor to a single
161 annotated RefSeq gene using the UCSC database were flagged accordingly in the
162 annotation file. In total, 27,171 probes (55.7%) were unambiguously associated with a single
163 mRNA or gene. The same method and version of the UCSC database was used for mapping
164 the probes of the Affymetrix GeneChip Human Exon 1.0 ST microarray. For this array, probe
165 sequences were obtained from the annotation file version HuEx-1_0-st-v2.r2 restricting the
166 probes to the main probe types of the core dataset with unique cross hybridization type, and
167 combining them at the level of transcript cluster. For this array system, 196,515 probes
168 (86.0%) of 17,876 transcript clusters were unambiguously associated with a single mRNA or
169 gene. Finally, the probes of both array systems were combined based on the same
170 transcripts obtained from the mapping against the UCSC database.

171 The Human Genome Nomenclature Committee (22) list 19,060 protein-coding genes (Sept
172 15, 2014), less than the total “unique identifiers” mapped by the two arrays used in the
173 overall meta-analysis; this discrepancy is due to probes on the array mapping to non-protein-
174 coding transcripts, which we have included under the term “unique genes” or “transcripts” in
175 this manuscript.

176

177

178

179 **Statistical analysis**

180 Using the R statistical software (39) and package “lme4” (3) each cohort performed a linear
181 mixed effects model for each probe in their microarray data, using the probe as the outcome,
182 muscle strength as an independent variable, and with the following covariates included as
183 fixed effects; age, sex, height (cm), weight (kg), cell count estimates (neutrophils, monocytes,
184 basophils and eosinophils), and fasting state (where applicable). By including these factors
185 as covariates in the models our results are independent of inter-individual variation in, for
186 example, lymphocyte cell counts. The following covariates were included as random effects;
187 batch (e.g. amplification and/or hybridization), study site (in InCHIANTI), family structure (in
188 FHS), and RNA quality (e.g. RNA Integrity Number - RIN - where available). Empirical cell
189 counts were only available in half of the FHS cohort; the rest of the cohort was imputed using
190 partial least square regression methods (see Online Methods for more details).

191

192 **Meta-analysis**

193 A sample-size weighted meta-analysis method was used, where an overall p-value and Z-
194 score for each probe are calculated which together describe the significance of the effect,
195 and the direction and magnitude, respectively; this method was chosen over the effect
196 size/standard error method because of the multiple array technologies and technical
197 considerations that differed between the cohorts. The analysis was done using the Meta-
198 Analysis Tool for Genome Wide-Association Scans (METAL) (54) which took the effect size,
199 sample size, and p-values from the individual cohort results as input (we set the “minor
200 allele”, “major allele”, “minor allele frequency”, and “strand” to the same fixed value for all

201 cohorts and probes, as this package was developed for GWAS and these options are not
202 relevant for gene expression data).

203 For each analysis (the primary analysis including all individuals, and the four subset
204 analyses) the Illumina-based cohorts (InCHIANTI, RS, SHIP) were meta-analyzed together,
205 as these technologies are very similar, then a secondary meta-analysis was performed which
206 used the FHS results and the Illumina results as the input; these are the final meta-analysis
207 results reported. This reduced the heterogeneity in the meta-analysis due to array
208 differences between the cohorts.

209 Before interpretation of the results, probes were excluded if they were expressed in <5% of
210 the sample or if the heterogeneity p-value calculated by METAL was <0.05. The Benjamini-
211 Hochberg (BH) (5) false-discovery rate (FDR) correction was applied to determine the
212 statistically significant probes for each analysis. Validation was defined as a gene with
213 $p < 0.05$ in the NESDA cohort.

214

215 **Ontology enrichment and network analysis**

216 The WEB-based GENE SeT ANALYSIS Toolkit (WebGestalt) online resource is a method for
217 determining pathway enrichment (56). We conducted a “Gene Ontology” analysis (database
218 version: Nov 11, 2012) and a “Human Phenotype Ontology” analysis (database version: May
219 20, 2014), which uses a systematic approach to phenotype abnormalities to link them into
220 ontologies (53). Default analysis options were selected, including BH multiple testing
221 adjustment, and the list of 17,534 genes included in the meta-analysis we used as the
222 “background”.

223 A co-expression analysis was performed in the FHS (as the study with the largest sample
224 size) where Spearman correlations were determined between each gene significantly
225 associated with muscle strength, and all other genes (after adjusting the data for the
226 covariates mentioned in the Statistical Analysis methods section). Genes correlated with ρ
227 ≥ 0.5 were selected for visualization in Cytoscape (v3.2.1). Ontology enrichment of networks
228 was performed in Cytoscape using the BiNGO plugin (v2.44).

229

230 ***A priori* genes associated with muscle function**

231 We selected sets of genes known to influence muscle function for *a priori* analysis to
232 highlight whether the pathways in muscle tissue are also important in whole blood. Kelch
233 proteins, including KLHL19, KLHL31, KLHL39m, and KLHDC1, are involved in skeletal
234 muscle function and development [8]; canonical and non-canonical Wnt signaling play crucial
235 roles in maintenance and development of skeletal muscle [7]; insulin-like growth factors
236 (IGF's) are also known to play roles in muscle growth and homeostasis [9], and finally TGF- β
237 family members, including myostatin (GDF8) (27) and GDF11. Supplementation of GDF11 in
238 mice was reported to ameliorate the sarcopenia-like phenotype [15]. In their 2007 study,
239 Melov *et al* identified 586 unique genes expressed in muscle that were associated with
240 endurance exercise training and differed between older and younger men (30), which we
241 also checked for associations with muscle strength in this analysis.

242

243 **Systematic literature search for genes**

244 For each significant gene in the analysis a systematic search of literature was performed by
245 accessing the "publications" list from GeneCards (www.genecards.org) (41), which has the

246 advantage of including publications where the gene ID may have changed over time. From
247 this list the title and abstract were downloaded from NCBI PubMed
248 (www.ncbi.nlm.nih.gov/pubmed). Searches were then made within each publication for the
249 text string “muscle”, and results counted.

250

251 **Results**

252 **Meta-analysis: genes associated with muscle strength**

253 Overall, 26,746 probe-pairs (corresponding probes on the Affymetrix and Illumina platforms),
254 mapping to 17,534 unique gene identifiers, were available for the meta-analysis. Including
255 data from all 7,781 participants, 208 unique genes (246 probe-pairs) were associated with
256 muscle strength (FDR<0.05; **Figure 1**; see **Supplementary Table 1** for all significant results)
257 after correction for multiple confounders, and excluding results with significant heterogeneity
258 (het $p < 0.05$) between cohorts; 133 were negatively associated with grip strength, 75 were
259 positively associated.

260 Of the 208 unique genes associated with muscle strength in the meta-analysis all were
261 significant in FHS alone (nominal $p < 0.05$), and 79 (38%) were also “independently”
262 associated with muscle strength in the Illumina meta-analysis (nominally significant; $p < 0.05$).
263 Details of the statistically most significant ‘top’ 20 transcripts are shown in **Table 2**. The
264 proportion “independently replicated” was greater for the top 30 most significant genes
265 identified in the meta-analysis (21 of 30 = 70%).

266

267 **Meta-analysis in subsets of the participants**

268 The analyses in subsets of the participants identified 13 genes associated with muscle
269 strength at transcriptome-wide significance in the meta-analysis that were not identified in the
270 analysis of all participants together (**Table 3**; see **Supplementary Table 2-4** for full results
271 for each subset). We also investigated whether the 208 genes identified in the analysis of all
272 individuals were nominally significant ($p < 0.05$) in the subsets. 153 of the 208 genes were
273 nominally significant in the older participants, 198 in the younger, 200 in the males and 121

274 in the females (**Supplementary Table 1**). **Supplementary Table 5** includes additional
275 information for all genes included in **Tables 2 & 3**.

276

277 **Significant genes only available on one array**

278 Due to differences between the array technologies and relative abundance of transcripts, not
279 all the genes were eligible for the meta-analysis. In the analysis on all individuals, 1,123
280 probes (898 unique identifiers) were present on the Affymetrix Exon array that did not have a
281 corresponding probe on the Illumina array; 21 of these probes were significantly associated
282 with hand-grip strength after BH adjustment for multiple testing (see **Table 4** for top 20
283 probes in the “all individuals” analysis and **Supplementary Tables 6-8** for list of significant
284 probes in each of the FHS analyses with significant results). In the Illumina array, 7,768
285 probes (6,119 unique gene identifiers) were available that did not map to a gene/transcript in
286 the Affymetrix Exon array (after excluding lowly expressed probes). None of the probes were
287 significantly associated with muscle strength after BH multiple testing correction.

288

289 **Ontology enrichment of strength-associated genes**

290 Two analyses were performed to identify pathways using the WebGestalt web resource
291 based on the 208 genes associated with muscle strength in the meta-analysis of all
292 participants;

293 **1. Gene Ontology** analysis found that 10 biological processes were significantly enriched
294 (FDR<0.05) (**Table 5**) including “hemoglobin metabolic process” and related processes,
295 “innate immune response” (18 genes) and the stress response (55 genes); 10 molecular
296 functions were enriched (including “protein binding genes”), and 10 cellular component

297 pathways were enriched (including “intracellular membrane-bound organelle”) (See
298 **Supplementary Table 9**).

299 **2. Human Phenotype Ontology** analysis found 10 phenotypes significantly enriched in the
300 genes, including “Anemia due to reduced life span of red cells”, “Hemolytic Anemia”, and
301 “Abnormality of erythrocytes” (See **Supplementary Table 10**).

302 Additionally, network analysis in the FHS of all genes correlated ($\rho \geq 0.5$) with the strength-
303 associated genes ($n=425$) revealed four clusters with at least 10 genes, all of which had a
304 number of significantly (FDR $p < 0.05$) enriched pathways (Supplementary Table 11): 1) the
305 largest cluster ($n=333$ of 425 genes) included “protein ubiquitination”, “erythrocyte
306 homeostasis”, and “cellular metabolic process”. 2) The second cluster ($n=32$ genes) was
307 enriched for genes in “regulation of cell communication”, “actin cytoskeleton” and “ATP
308 metabolic process”. 3) The third cluster ($n=18$) had two enriched ontologies only: “cell
309 surface receptor-linked signaling pathway” and “cytolysis”. 4) The final cluster ($n=10$) was
310 enriched for terms including “negative regulator or immune system process” and “negative
311 regulation of complement activation”.

312

313 ***A priori* genes associated with muscle function**

314 Of 20 *IGF*-related genes tested in this meta-analysis two were significantly associated with
315 muscle strength: *IGF1R* (positively associated) and *IGF2BP2* (negatively associated; meta-
316 analysis $FDR=3.2 \times 10^{-2}$ and $FDR=1.2 \times 10^{-3}$, respectively). Expression of myostatin (*MSTN*),
317 follistatin (*FST*) and *GDF11* were not associated with muscle strength ($FDR > 0.05$). 40 unique
318 Kelch genes were tested in the meta-analysis; none were associated with muscle strength in
319 whole blood ($FDR > 0.05$). 18 unique Wnt genes (from *WNT1* to *WNT9B*) were tested in the
320 meta-analysis; none were associated with muscle strength in whole blood ($FDR > 0.05$). All 10

321 Frizzled genes (*FZD1-10*, receptors for the Wnt pathway) were also available to test; none
322 were associated with muscle strength. Similarly all three Dishevelled genes were available to
323 test (*DVL1-3*, acts directly downstream of the Frizzled receptors) and none were associated
324 with muscle strength. Of 586 genes identified by Melov *et al* that were differentially
325 expressed in muscle tissue between old and young men following endurance training (30)
326 four were associated with muscle strength in this analysis: *ANP32B*, *CIRBP*, *MCM7* and
327 *MGST1*.

328

329

330 **Most associated genes are not previously linked to muscle in the literature**

331 For each of the 221 genes associated with muscle strength we searched in the published
332 literature cataloged on GeneCards and NCBI Pubmed titles and abstracts, using the search
333 term “muscle”: for 115 of these 221 genes (52%) there were no mentions of muscle (as of
334 Nov 12, 2014; **Supplementary Table 12**).

335

336 **Few genes replicate in the NESDA cohort**

337 NESDA was not included in the meta-analysis due to data limitations: the lack of empirically
338 determined or reliably imputed white cell count data, the use of a different microarray
339 technology (a predecessor to the Exon array used by FHS, much more dissimilar than the
340 v3/v4 Illumina arrays are to one another), and a younger population than the other cohorts
341 included in the meta-analysis (max age=65, see **Table 1**).

342 As noted above, in total 221 unique genes were associated with muscle strength across all
343 the meta-analyses performed. Of 208 genes significantly associated with muscle strength in
344 analysis 1 (all participants) it was possible to test 144 in the NESDA cohort; 7 genes were
345 also associated with muscle strength ($p < 0.05$) in the NESDA cohort (*ACSL6*, *ALDH5A1*,
346 *CARHSP1*, *FGL1*, *NRG1*, *PIGB*, *SIGLEC7*).

347

348 **Associations with knee strength**

349 Maximum knee and grip strength (both in Kg) were measured in 619 participants in the
350 InCHIANTI study and were highly correlated (Pearson $R = 0.751$), and were significantly
351 associated after adjustment for age, sex, height and weight (coefficient = 0.193, $p = 8.2 \times 10^{-15}$)
352 in linear regression models with knee strength as the dependent variable.

353

354

355 **Discussion**

356 In this discovery study we set out to determine whether specific transcript levels in blood are
357 associated with muscle strength in multiple human cohorts including mostly middle-aged
358 volunteers. Previous cross-sectional (and longitudinal) studies have shown that the degree
359 and *rate* of loss of strength (and muscle mass) is greater in older participants (31). We
360 therefore performed stratified analyses by age and gender to determine whether transcripts
361 or pathways associated with muscle strength in whole blood differ between these groups.
362 208 unique genes were associated with muscle strength in the analysis of all participants
363 (**Table 2** for 20 most robust associations). Thirteen additional unique genes were identified
364 that were only associated when participants were separated into older/younger or
365 male/female groups (**Table 3**). In total 221 unique genes were associated with muscle
366 strength in at least one analysis, 52% of which were not previously linked to the term
367 “muscle” in the published literature cataloged on GeneCards and PubMed (as of Nov 12,
368 2014).

369 We observe significant associations between muscle strength and expression of *IGF1R* and
370 *IGF2BP2* (positive and negative directions of association with muscle, respectively), growth
371 factors involved in skeletal muscle growth (33, 46); the former is known to enhance cell
372 survival by mediating IGF1 signaling, and the latter modulates *IGF2* translation and has
373 genetic variants associated with type-2 diabetes (19). Of 586 genes that differ in expression
374 in muscle between old and young men after endurance training (30), four were associated
375 with grip strength in this study: *MGST1* (negative direction), an immune mediator which may
376 protect against oxidative stress (49); *MCM7* (positive direction), which regulates DNA
377 replication during proliferation (12); *CIRBP* (positive direction), which promotes inflammation
378 in response to shock and sepsis (38); and *ANP32B* (negative direction), a cell-cycle
379 progression and anti-apoptosis factor (41). These latter results suggest that most blood
380 based gene expression associated with strength is different from that seen in muscle itself,

381 which is not unexpected given the respective systemic regulatory versus myofibril
382 maintenance functions involved. Further work should explore whether transcripts that alter in
383 response to exercise show overlaps between circulating cells and muscle. Interestingly, no
384 genes from the Wnt or Kelch pathways (both known to be important for muscle function (14,
385 29)) were associated with strength in this analysis; nor was *GDF11*, a protein that can
386 reverse age-related muscle dysfunction in mice (48), although as noted we observe
387 associations between strength and expression of two IGF-related genes.

388 Other genes of note include *CCR6* and *PRF1* (both positively associated with muscle
389 strength and age (17)): *CCR6* is implicated in B-cell maturation and recruitment (36); perforin
390 (*PRF1*) is a protein secreted by cytotoxic T-cells which creates pores in membranes to permit
391 apoptosis-inducing granzyme into the target cell (50). These findings are consistent with the
392 notion that inflammation may be associated with muscle repair, maintenance and turnover, at
393 least in part by interfering with the production and biological activity of IGF-1 (2).

394 *NANOG* expression was measured in the FHS analysis only and is positively associated with
395 strength in the FHS analysis (**Table 4**); *NANOG* can reverse aging of some stem cells (16)
396 and, in combination with three other genes (*OCT4*, *SOX2*, and *LIN28*, not significant in this
397 analysis), can induce pluripotency of somatic cells (55). This may suggest that differentiation
398 (of whole blood cells) is inversely correlated with muscle strength, but the mechanisms are
399 unclear.

400

401 **Genes identified in subset analyses**

402 Thirteen genes were associated with muscle strength at transcriptome-wide significance in
403 the subset analysis only (**Table 3**). These were predominantly in the younger (<60 years)
404 group and included *PIK3R2*, a negative regulator of the PI3K/AKT growth pathway; the

405 negative expression association with strength suggests that there is increased PI3K activity
406 (due to reduced expression of *PIK3R2*) with increasing muscle strength, in whole blood. This
407 association is observed in the younger subset ($p=6 \times 10^{-5}$) but not in the older subset, even
408 nominally ($p=0.92$), suggesting differences in growth pathway expression in blood with
409 respect to muscle strength as individuals age. Similarly expression of *PDK4* (inhibits
410 pyruvate dehydrogenase in mitochondria, thereby reducing the conversion of pyruvate into
411 acetyl-CoA) is negatively associated with strength, suggesting increased pyruvate
412 dehydrogenase activity with increased strength in younger individuals only. *PKN2*,
413 (associated with height (19) and cell-cycle progression), is positively associated with strength
414 in the younger individuals, underlining the difference in growth pathways in whole blood
415 between younger and older individuals. See **Supplementary Table 5** for more details on the
416 other results.

417 Defensin, Alpha 4, Corticostatin (*DEFA4*, negative strength association in the analysis of
418 females only) is a cytotoxic peptide that has antimicrobial activity against Gram-negative
419 bacteria (predominantly) (36). In males, expression of *RAC1* (membrane-associated GTPase
420 involved in signal transduction, including growth signals) and *NDUFS1* (member of
421 mitochondrial complex 1, may form part of the active site) were positively associated with
422 muscle strength; these associations may suggest that on average males have specific
423 energy and growth-related gene expression relating to strength.

424 No genes were associated (transcriptome-wide) in the older participants only; although the
425 sample size was still reasonably high (2,402 participants), variability in the strength
426 phenotype as individuals age and development of various co-morbidities plus chronic
427 inflammation may reduce the power to detect associations. Subset-specific gene-expression
428 associations with strength reported here need to be replicated and added to, as we may lack
429 statistical power in this study to detect smaller-effect associations, and the microarrays do
430 not quantify all transcripts or isoforms present.

431

432 **Enrichment analysis**

433 WebGestalt analyses identified statistically significant enrichment for genes in the biological
434 process “Hemoglobin Metabolic Process” and the phenotypic abnormality “Hemolytic
435 Anemia”, amongst others. Anemia is a cross-sectional correlate of muscle strength and
436 predicts accelerated muscle strength decline with aging (10), while hemoglobin levels are
437 positively associated with muscle strength and density (10). Circulating reticulocytes
438 (erythrocyte precursors with some residual RNA present) were not adjusted for in this
439 analysis and are likely the source of the associations with genes such as *ALAS2* (strongest
440 meta-analysis association, negative direction - a rate-limiting step in heme biosynthesis (24)).

441 “Innate Immune Response” - which includes macrophages - genes were also enriched in the
442 results. *CEBPB*, the gene implicated in the macrophage wound-healing response (42) and
443 significantly associated with muscle strength in the 2012 study by InCHIANTI (18), did not
444 replicate in the other cohorts. This could be due to methodological differences between the
445 previous study and this meta-analysis, as well as differences in age distribution (81% of the
446 InCHIANTI cohort is aged ≥ 60 years, compared to 31% in this analysis, which includes the
447 InCHIANTI cohort; **Table 1**). The implications are unclear given the mouse model evidence
448 of plausible biological mechanism (42) and evidence in humans that exercise-induced
449 muscle damage is associated with *CEBPB* expression changes in whole blood (6). Further
450 work is required in older and frail groups.

451 Co-expression analysis of all genes correlated with those identified in the meta-analysis to be
452 significantly associated with strength revealed four clusters. Ontology enrichment analysis of
453 these revealed very similar results to those identified using WebGestalt only on the genes
454 significantly associated with strength, emphasizing the association of immune activation and

455 cell signaling pathways to muscle strength in whole blood, in addition to hemoglobin
456 pathways.

457

458 **Limitations**

459 There are several potential limitations of this study including its cross-sectional design; it is
460 not possible to determine a causal direction in this study for the associations reported, but
461 the robustly identified markers emerging provide a sound foundation for follow-up studies to
462 address causation. Grip strength is strongly correlated with strength in other key muscle
463 systems (see introduction), but further work will be needed to confirm more specific gene
464 expression associations with strength in other muscle groups. Grip strength can be
465 influenced by non-muscle strength factors, including functional anomalies in the hands, for
466 example caused by Rheumatoid Arthritis (47), and work is needed to clarify whether any of
467 our findings reflect these alternative influences.

468 Another potential limitation is the mixed cell subtype composition of “whole blood”: our
469 analysis approach based on overall expression should have greater power to detect net
470 expression changes in common immune cell types or large changes in expression of highly
471 specific genes, but will have less power to detect smaller expression changes within less
472 numerous cell subtypes. The cell subtype origins of the top transcripts reported here now
473 need to be identified. Additionally, the microarray technology used across the participating
474 cohorts was not the same. However, 21 (70%) of the top 30 meta-analysis results were
475 independently replicated between the platforms, which suggests that the top (most strongly
476 associated) results are very robust to cohort and array differences. Also, the current analysis
477 has identified expressed genes statistically associated with muscle strength, but future work
478 will be needed to identify the mechanisms underlying these associations, and whether these
479 act on muscle directly or through indirect pathways, perhaps with effects on central

480 command, cerebellar coordination or neural transmission. Finally, work is needed on whether
481 the identified strength associated gene expression transcripts are predictive of subsequent
482 changes in strength or functional decline.

483 **Conclusions**

484 In this first large-scale transcriptome wide study in human blood, we have identified robust
485 associations between the expression of 221 genes and muscle strength in adults. Several
486 known pathways were confirmed, including growth factor-related genes, the innate immune
487 response and hemoglobin metabolism. For 115 genes this analysis appears to provide the
488 first published link to muscle. The analysis also suggests that parts of the expression
489 signatures may be specific to subgroups, notably with 10 genes associated with muscle
490 strength only in younger people.

491 Further work is needed to establish which of the identified genes predict future changes in
492 strength. The findings of genes via expression microarrays may help identify key changes in
493 cell subtypes in blood contributing to strength, through studies of the cellular origins of gene
494 expression signals. Future research should also include longitudinal data to assess whether
495 expression of the identified genes predicts poor muscle strength or functional outcomes.

496

497

498

499 **List of Abbreviations**

500

BH	Benjamini-Hochberg
cDNA	Complementary DNA (usually from PCR, can be originally DNA or RNA)
CHARGE	Cohorts for Heart and Aging Research in Genomic Epidemiology
DNA	Deoxyribonucleic Acid
FDR	False-Discovery Rate
FHS	Framingham Heart Study
GWAS	Genome-Wide Association Study
IL6	Interleukin 6
InCHIANTI	Invecchiare Nel Chianti – aging in the Chianti region of Italy
METAL	Meta-Analysis Tool for Genome Wide-Association Scans
NESDA	Netherlands Study of Depression and Anxiety
QC	Quality Control
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RS	Rotterdam Study
SHIP	Study of Health in Pomerania
WebGestalt	WEB-based GEne SeT AnaLysis Toolkit

501

502

503 **Disclosures**

504 The authors declare that they have no competing interests.

505

506 **Authors' Contributions**

507 LCP participated in the design of analysis, performed analysis of the InCHIANTI data,
508 performed the meta-analysis, interpreted the results and co-wrote the manuscript. RJ, TK,
509 MJP and RJ2 participated in design of analysis, performed analysis in their respective
510 cohorts, provided interpretation of results and contributed to draft manuscripts. DK, DPK and
511 LWH had substantial contributions to study design and results interpretation. AT performed
512 microarray re-annotation by sequence mapping. JEP provided methods input and comments
513 to manuscript. DL, HL, KL, PJM, SB, WEH, DGH, ABS, TT, GG, AH, AGU, RB, SG, GH, CM
514 and UV were involved in data acquisition and methods development, including designing this
515 analysis. BWJHP, JBJM, LF, TK2 and JMM contributed to the analysis plan and methods
516 discussion, provided interpretation of results and contributed to draft manuscripts, and
517 oversaw analyses performed within their cohorts. DM participated in the design of the
518 analysis, interpreted the results and co-wrote the manuscript. All authors read, provided
519 comments, and approved the final manuscript.

520

521 **Funding and Acknowledgements**

522 FHS gene expression profiling was funded through the Division of Intramural Research
523 (Principal Investigator, Daniel Levy), National Heart, Lung, and Blood Institute, National
524 Institutes of Health, Bethesda, MD. Dr. Murabito is supported by NIH grant R01AG029451.
525 Dr. Kiel is supported by NIH R01 AR41398. The Framingham Heart Study is supported by
526 National Heart, Lung, and Blood Institute contract N01-HC-25195.

527 The InCHIANTI study was supported in part by the Intramural Research Program, National
528 Institute on Aging, NIH, Baltimore MD USA. D.M. and L.W.H. were generously supported by
529 a Wellcome Trust Institutional Strategic Support Award (WT097835MF). W.E.H. was funded
530 by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied
531 Health Research and Care (CLAHRC) for the South West Peninsula. The views expressed in
532 this publication are those of the authors and not necessarily those of the NHS, the NIHR or
533 the Department of Health in England.

534 The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht
535 program of the Netherlands Organisation for Health Research and Development (Zon-Mw,
536 grant number 10-000-1002) and is supported by participating universities and mental health
537 care organizations (VU University Medical Center, GGZ inGeest, Arkin, Leiden University
538 Medical Center, GGZ Rivierduinen, University Medical Center Groningen, Lentis, GGZ
539 Friesland, GGZ Drenthe, Scientific Institute for Quality of Healthcare (IQ healthcare),
540 Netherlands Institute for Health Services Research (NIVEL) and Netherlands Institute of
541 Mental Health and Addiction (Trimbos Institute).

542 The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University,
543 Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw),
544 the Netherlands Organisation of Scientific Research NWO Investments (nr.
545 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-

546 015; RIDE2), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare
547 and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The
548 authors are grateful to the study participants, the staff from the Rotterdam Study and the
549 participating general practitioners and pharmacists. The generation and management of
550 RNA-expression array data for the Rotterdam Study was executed and funded by the Human
551 Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine,
552 Erasmus MC, the Netherlands. We thank Marjolein Peters, MSc, Ms. Mila Jhamai, Ms.
553 Jeannette M. Vergeer-Drop, Ms. Bernadette van Ast-Copier, Mr. Marijn Verkerk and Jeroen
554 van Rooij, BSc for their help in creating the RNA array expression database.

555 SHIP is part of the Community Medicine Research net of the University of Greifswald,
556 Germany, which is funded by the Federal Ministry of Education and Research (grants no.
557 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social
558 Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald
559 Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of
560 Education and Research (grant 03IS2061A). The University of Greifswald is a member of the
561 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus
562 program of the InterSystems GmbH.

563

564 **References**

- 565 1. **Affymetrix**. Technical Note: Factors Affecting Blood Gene Expression [Online]. 2011.
566 http://www.panomics.com/downloads/QG_BloodGene_TN_RevB_110719.pdf.
- 567 2. **Barbieri M, Ferrucci L, Ragno E, Corsi A, Bandinelli S, Bonafè M, Olivieri F,**
568 **Giovagnetti S, Franceschi C, Guralnik JM, Paolisso G**. Chronic inflammation and
569 the effect of IGF-I on muscle strength and power in older persons. *Am J Physiol*
570 *Endocrinol Metab* 284: E481–E487, 2003.
- 571 3. **Bates D, Mächler M, Bolker B**. Fitting linear mixed-effects models using lme4. *J Stat*
572 *Softw* ... submitted: 51, 2012.
- 573 4. **Baylis D, Ntani G, Edwards MH, Syddall HE, Bartlett DB, Dennison EM, Martin-**
574 **Ruiz C, von Zglinicki T, Kuh D, Lord JM, Aihie Sayer A, Cooper C**. Inflammation,
575 Telomere Length, and Grip Strength: A 10-year Longitudinal Study. *Calcif Tissue Int*
576 95: 54–63, 2014.
- 577 5. **Benjamini Y, Hochberg Y**. Controlling the false discovery rate: a practical and
578 powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.* .
- 579 6. **Blackwell J, Harries LW, Pilling LC, Ferrucci L, Jones A, Melzer D**. Changes in
580 CEBPB expression in circulating leukocytes following eccentric elbow-flexion exercise.
581 *J. Physiol. Sci.* (November 13, 2014). doi: 10.1007/s12576-014-0350-7.
- 582 7. **Bohannon RW, Magasi SR, Bubela DJ, Wang YC, Gershon RC**. Grip and Knee
583 extension muscle strength reflect a common construct among adults. *Muscle and*
584 *Nerve* 46: 555–558, 2012.
- 585 8. **Carlson B, Faulkner J**. Muscle transplantation between young and old rats: age of
586 host determines recovery. *Am J Physiol Physiol* 256: C1262, 1989.
- 587 9. **Cesari M, Pahor M, Lauretani F, Zamboni V, Bandinelli S, Bernabei R, Guralnik**
588 **JM, Ferrucci L**. Skeletal muscle and mortality results from the InCHIANTI Study. *J*
589 *Gerontol A Biol Sci Med Sci* 64: 377–84, 2009.
- 590 10. **Cesari M, Penninx BWJH, Lauretani F, Russo CR, Carter C, Bandinelli S,**
591 **Atkinson H, Onder G, Pahor M, Ferrucci L**. Hemoglobin levels and skeletal muscle:
592 results from the InCHIANTI study. *J Gerontol A Biol Sci Med Sci* 59: 249–54, 2004.
- 593 11. **Conboy IM, Conboy MJ, Wagers AJ, Girma ER, Weissman IL, Rando TA**.
594 Rejuvenation of aged progenitor cells by exposure to a young systemic environment.
595 *Nature* 433: 760–4, 2005.
- 596 12. **Cortez D, Glick G, Elledge SJ**. Minichromosome maintenance proteins are direct
597 targets of the ATM and ATR checkpoint kinases. *Proc Natl Acad Sci U S A* 101:
598 10078–83, 2004.
- 599 13. **Ferrucci L, Bandinelli S, Benvenuti E, Di Iorio A, Macchi C, Harris TB, Guralnik**
600 **JM**. Subsystems contributing to the decline in ability to walk: bridging the gap between
601 epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc* 48:
602 1618–25, 2000.

- 603 14. **Gupta VA, Beggs AH.** Kelch proteins: emerging roles in skeletal muscle development
604 and diseases. *Skelet Muscle* 4: 11, 2014.
- 605 15. **Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB.** Lower-extremity
606 function in persons over the age of 70 years as a predictor of subsequent disability. *N*
607 *Engl J Med* 332: 556–61, 1995.
- 608 16. **Han J, Mistriotis P, Lei P, Wang D, Liu S, Andreadis ST.** Nanog reverses the
609 effects of organismal aging on mesenchymal stem cell proliferation and myogenic
610 differentiation potential. *Stem Cells* 30: 2746–59, 2012.
- 611 17. **Harries LW, Hernandez D, Henley W, Wood AR, Holly AC, Bradley-Smith RM,**
612 **Yaghoobkar H, Dutta A, Murray A, Frayling TM, Guralnik JM, Bandinelli S,**
613 **Singleton A, Ferrucci L, Melzer D.** Human aging is characterized by focused
614 changes in gene expression and deregulation of alternative splicing. *Aging Cell* 10:
615 868–78, 2011.
- 616 18. **Harries LW, Pilling LC, Hernandez LDG, Bradley-Smith R, Henley W, Singleton**
617 **AB, Guralnik JM, Bandinelli S, Ferrucci L, Melzer D.** CCAAT-enhancer-binding
618 protein-beta expression in vivo is associated with muscle strength. *Aging Cell* 11: 262–
619 268, 2012.
- 620 19. **Hindorff L, MacArthur J, Wise A, Junkins H, Hall P, Klemm A, Manolio T.** A
621 Catalog of Published Genome-Wide Association Studies [Online]. [date unknown].
622 www.genome.gov/gwastudies.
- 623 20. **Hofman A, van Duijn CM, Franco OH, Ikram MA, Janssen HLA, Klaver CCW,**
624 **Kuipers EJ, Nijsten TEC, Stricker BHC, Tiemeier H, Uitterlinden AG, Vernooij**
625 **MW, Witteman JCM.** The Rotterdam Study: 2012 objectives and design update. *Eur J*
626 *Epidemiol* 26: 657–86, 2011.
- 627 21. **Huan T, Esko T, Peters MJ, Pilling LC, Schramm K, Schurmann C, Chen BH, Liu**
628 **C, Joehanes R, Johnson AD, Yao C, Ying S, Courchesne P, Milani L,**
629 **Raghavachari N, Wang R, Liu P, Reinmaa E, Dehghan A, Hofman A, Uitterlinden**
630 **AG, Hernandez DG, Bandinelli S, Singleton A, Melzer D, Metspalu A, Carstensen**
631 **M, Grallert H, Herder C, Meitinger T, Peters A, Roden M, Waldenberger M, Dörr**
632 **M, Felix SB, Zeller T, Vasan R, O'Donnell CJ, Munson PJ, Yang X, Prokisch H,**
633 **Völker U, van Meurs JBJ, Ferrucci L, Levy D.** A Meta-analysis of Gene Expression
634 Signatures of Blood Pressure and Hypertension. *PLOS Genet* 11: e1005035, 2015.
- 635 22. **HUGO Gene Nomenclature Committee.** HGNC [Online]. [date unknown].
636 <http://www.genenames.org/>.
- 637 23. **Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP.** An investigation
638 of coronary heart disease in families. The Framingham offspring study. *Am J*
639 *Epidemiol* 110: 281–90, 1979.
- 640 24. **Khan AA, Quigley JG.** Control of intracellular heme levels: heme transporters and
641 heme oxygenases. *Biochim Biophys Acta* 1813: 668–82, 2011.
- 642 25. **Lang T, Streeper T, Cawthon P, Baldwin K, Taaffe DR, Harris TB.** Sarcopenia:
643 etiology, clinical consequences, intervention, and assessment. *Osteoporos Int* 21:
644 543–59, 2010.

- 645 26. **Lee CE, McArdle A, Griffiths RD.** The role of hormones, cytokines and heat shock
646 proteins during age-related muscle loss. *Clin Nutr* 26: 524–34, 2007.
- 647 27. **Lee S-J, Lee Y-S, Zimmers TA, Soleimani A, Matzuk MM, Tsuchida K, Cohn RD,**
648 **Barton ER.** Regulation of Muscle Mass by Follistatin and Activins. *Mol Endocrinol* 24:
649 1998–2008, 2010.
- 650 28. **Leong DP, Teo KK, Rangarajan S, Lopez-Jaramillo P, Avezum A, Orlandini A,**
651 **Seron P, Ahmed SH, Rosengren A, Kelishadi R, Rahman O, Swaminathan S,**
652 **Iqbal R, Gupta R, Lear SA, Oguz A, Yusoff K, Zatonska K, Chifamba J, Igumbor**
653 **E, Mohan V, Anjana RM, Gu H, Li W, Yusuf S, Prospective Urban Rural**
654 **Epidemiology (PURE) Study investigators.** Prognostic value of grip strength:
655 findings from the Prospective Urban Rural Epidemiology (PURE) study. *Lancet* (2015).
656 doi: 10.1016/S0140-6736(14)62000-6.
- 657 29. **Von Maltzahn J, Chang NC, Bentzinger CF, Rudnicki MA.** Wnt signaling in
658 myogenesis. *Trends Cell Biol* 22: 602–9, 2012.
- 659 30. **Melov S, Tarnopolsky M a, Beckman K, Felkey K, Hubbard A.** Resistance exercise
660 reverses aging in human skeletal muscle. *PLoS One* 2: e465, 2007.
- 661 31. **Mitchell WK, Williams J, Atherton P, Larvin M, Lund J, Narici M.** Sarcopenia,
662 dynapenia, and the impact of advancing age on human skeletal muscle size and
663 strength; a quantitative review. *Front Physiol* 3: 260, 2012.
- 664 32. **Newman AB, Kupelian V, Visser M, Simonsick EM, Goodpaster BH, Kritchevsky**
665 **SB, Tylavsky FA, Rubin SM, Harris TB.** Strength, but not muscle mass, is
666 associated with mortality in the health, aging and body composition study cohort. *J*
667 *Gerontol A Biol Sci Med Sci* 61: 72–77, 2006.
- 668 33. **Paoni NF, Peale F, Wang F, Errett-Baroncini C, Steinmetz H, Toy K, Bai W,**
669 **Williams PM, Bunting S, Gerritsen ME, Powell-Braxton L.** Time course of skeletal
670 muscle repair and gene expression following acute hind limb ischemia in mice. *Physiol*
671 *Genomics* 11: 263–272, 2002.
- 672 34. **Patel HP, Al-Shanti N, Davies LC, Barton SJ, Grounds MD, Tellam RL, Stewart**
673 **CE, Cooper C, Sayer AA.** Lean Mass, Muscle Strength and Gene Expression in
674 Community Dwelling Older Men: Findings from the Hertfordshire Sarcopenia Study
675 (HSS). *Calcif. Tissue Int.* (July 24, 2014). doi: 10.1007/s00223-014-9894-z.
- 676 35. **Penninx BWJH, Beekman ATF, Smit JH, Zitman FG, Nolen WA, Spinhoven P,**
677 **Cuijpers P, De Jong PJ, Van Marwijk HWJ, Assendelft WJJ, Van Der Meer K,**
678 **Verhaak P, Wensing M, De Graaf R, Hoogendijk WJ, Ormel J, Van Dyck R.** The
679 Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and
680 methods. *Int J Methods Psychiatr Res* 17: 121–40, 2008.
- 681 36. **Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O,**
682 **Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O’Leary NA, Pujar S,**
683 **Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin**
684 **C, Webb D, Weber J, Wu W, DiCuccio M, Kitts P, Maglott DR, Murphy TD, Ostell**
685 **JM.** RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* 42:
686 D756–63, 2014.

- 687 37. **Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI,**
688 **Uitterlinden AG, Harris TB, Witteman JCM, Boerwinkle E.** Cohorts for Heart and
689 Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of
690 prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ*
691 *Cardiovasc Genet* 2: 73–80, 2009.
- 692 38. **Qiang X, Yang W-L, Wu R, Zhou M, Jacob A, Dong W, Kuncewitch M, Ji Y, Yang**
693 **H, Wang H, Fujita J, Nicastro J, Coppa GF, Tracey KJ, Wang P.** Cold-inducible
694 RNA-binding protein (CIRP) triggers inflammatory responses in hemorrhagic shock
695 and sepsis. *Nat Med* 19: 1489–95, 2013.
- 696 39. **R Core Team.** R: A language and environment for statistical computing. R Foundation
697 for Statistical Computing: 2014.
- 698 40. **Rantanen T.** Muscle strength, disability and mortality. *Scand J Med Sci Sports* 13: 3–
699 8, 2003.
- 700 41. **Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D.** GeneCards: integrating
701 information about genes, proteins and diseases. *Trends Genet* 13: 163, 1997.
- 702 42. **Ruffell D, Mourkioti F, Gambardella A, Kirstetter P, Lopez RG, Rosenthal N,**
703 **Nerlov C.** A CREB-C/EBPbeta cascade induces M2 macrophage-specific gene
704 expression and promotes muscle injury repair. *Proc Natl Acad Sci U S A* 106: 17475–
705 80, 2009.
- 706 43. **Samson MM, Meeuwssen IB, Crowe A, Dessens JA, Duursma SA, Verhaar HJ.**
707 Relationships between physical performance measures, age, height and body weight
708 in healthy adults. *Age Ageing* 29: 235–242, 2000.
- 709 44. **Schaap LA, Pluijm SMF, Deeg DJH, Visser M.** Inflammatory markers and loss of
710 muscle mass (sarcopenia) and strength. *Am J Med* 119: 526.e9–17, 2006.
- 711 45. **Schurmann C, Heim K, Schillert A, Blankenberg S, Carstensen M, Dörr M,**
712 **Endlich K, Felix SB, Gieger C, Grallert H, Herder C, Hoffmann W, Homuth G, Illig**
713 **T, Kruppa J, Meitinger T, Müller C, Nauck M, Peters A, Rettig R, Roden M,**
714 **Strauch K, Völker U, Völzke H, Wahl S, Wallaschofski H, Wild PS, Zeller T,**
715 **Teumer A, Prokisch H, Ziegler A.** Analyzing illumina gene expression microarray
716 data from different tissues: methodological aspects of data analysis in the metaxpress
717 consortium. *PLoS One* 7: e50938, 2012.
- 718 46. **Sharples AP, Al-Shanti N, Hughes DC, Lewis MP, Stewart CE.** The role of insulin-
719 like-growth factor binding protein 2 (IGFBP2) and phosphatase and tensin homologue
720 (PTEN) in the regulation of myoblast differentiation and hypertrophy. *Growth Horm*
721 *IGF Res* 23: 53–61, 2013.
- 722 47. **Shiratori AP, Iop R da R, Júnior NGB, Domenech SC, Gevaerd M da S.** Evaluation
723 protocols of hand grip strength in individuals with rheumatoid arthritis: A systematic
724 review. *Rev. Bras. Reumatol.* 54: 140–147, 2014.
- 725 48. **Sinha M, Jang YC, Oh J, Khong D, Wu EY, Manohar R, Miller C, Regalado SG,**
726 **Loffredo FS, Pancoast JR, Hirshman MF, Lebowitz J, Shadrach JL, Cerletti M,**
727 **Kim M-J, Serwold T, Goodyear LJ, Rosner B, Lee RT, Wagers AJ.** Restoring
728 systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle.
729 *Science* 344: 649–52, 2014.

- 730 49. **Siritantikorn A, Johansson K, Ahlen K, Rinaldi R, Suthiphongchai T, Wilairat P,**
731 **Morgenstern R.** Protection of cells from oxidative stress by microsomal glutathione
732 transferase 1. *Biochem Biophys Res Commun* 355: 592–6, 2007.
- 733 50. **Thiery J, Keefe D, Boulant S, Boucrot E, Walch M, Martinvalet D, Goping IS,**
734 **Bleackley RC, Kirchhausen T, Lieberman J.** Perforin pores in the endosomal
735 membrane trigger the release of endocytosed granzyme B into the cytosol of target
736 cells. *Nat Immunol* 12: 770–7, 2011.
- 737 51. **Visser M, Deeg DJ, Lips P, Harris TB, Bouter LM.** Skeletal muscle mass and
738 muscle strength in relation to lower-extremity performance in older men and women. *J*
739 *Am Geriatr Soc* 48: 381–386, 2000.
- 740 52. **Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, Aumann N, Lau**
741 **K, Piontek M, Born G, Havemann C, Ittermann T, Schipf S, Haring R, Baumeister**
742 **SE, Wallaschofski H, Nauck M, Frick S, Arnold A, Jünger M, Mayerle J, Kraft M,**
743 **Lerch MM, Dörr M, Reffelmann T, Empen K, Felix SB, Obst A, Koch B, Gläser S,**
744 **Ewert R, Fietze I, Penzel T, Dören M, Rathmann W, Haerting J, Hannemann M,**
745 **Röpcke J, Schminke U, Jürgens C, Tost F, Rettig R, Kors JA, Ungerer S,**
746 **Hegenscheid K, Kühn J-PJ, Kühn J-PJ, Hosten N, Puls R, Henke J, Gloger O,**
747 **Teumer A, Homuth G, Völker U, Schwahn C, Holtfreter B, Polzer I, Kohlmann T,**
748 **Grabe HJ, Roszkopf D, Kroemer HK, Kocher T, Biffar R, John U, Hoffmann W.**
749 Cohort profile: the study of health in Pomerania. *Int J Epidemiol* 40: 294–307, 2011.
- 750 53. **Wang J, Duncan D, Shi Z, Zhang B.** WEB-based GENE SeT AnaLysis Toolkit
751 (WebGestalt): update 2013. *Nucleic Acids Res* 41: W77–83, 2013.
- 752 54. **Willer CJ, Li Y, Abecasis GR.** METAL: fast and efficient meta-analysis of
753 genomewide association scans. *Bioinformatics* 26: 2190–1, 2010.
- 754 55. **Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie**
755 **J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA.** Induced pluripotent
756 stem cell lines derived from human somatic cells. *Science* 318: 1917–20, 2007.
- 757 56. **Zhang B, Kirov S, Snoddy J.** WebGestalt: an integrated system for exploring gene
758 sets in various biological contexts. *Nucleic Acids Res* 33: W741–8, 2005.
- 759

760 **Table 1** | Characteristics of the study cohorts

761

Variables	Meta-analysis cohorts					Additional cohort ∞
	FHS Gen 2 *	FHS Gen 3 *	InCHIANTI ~	RS3 ~	SHIP ~	NESDA
N	2,421	3,155	667	556	982	1,989
N >= 60 years (%)	1,319 (54%)	48 (1%)	547 (81%)	264 (45%)	268 (27%)	158 (8%)
Sex (male), n (%)	1,095 (45%)	1,470 (47%)	311 (46%)	272 (46%)	435 (44%)	1,328 (67%)
Age (yrs), mean±SD	66 ± 8.9	46 ± 8.8	72 ± 15	60 ± 7.9	50 ± 14	42 ± 13
Age (yrs), min:max	40 : 90	24 : 78	30 : 104	46 : 89	20 : 81	18 : 65
WBC-counts	yes †	yes	yes	yes	yes	no
<i>Hand-grip strength</i>						
Mean ± SD (Kg)	31 ± 12	38 ± 12	29 ± 12	25 ± 9.4	38 ± 12	38 ± 12
Min : max (Kg)	1 : 76	5 : 84	3 : 76	2 : 55	11 : 73	10 : 90
Microarray platform	Affymetrix GeneChip Human Exon 1.0 ST	Affymetrix GeneChip Human Exon 1.0 ST	Illumina HumanHT- 12 v3 BeadChip	Illumina HumanHT- 12 v4 BeadChip	Illumina HumanHT- 12 v3 BeadChip	Affymetrix Human Genome U219 Array

FHS=Framingham Heart Study; RS3=Rotterdam Study 3; SHIP=Study of Health in Pomerania; NESDA=Netherlands Study of Depression and Anxiety; SD=standard deviation; WBC=white blood cell

* FHS cohorts analyzed together prior to overall meta-analysis

~ Illumina-based cohorts analyzed together prior to overall meta-analysis

∞ cohort not included in meta-analysis due to data missing from analysis protocol

† cell counts imputed in this dataset; see methods

762 **Table 2** | Top 20 unique genes associated with muscle strength in the meta-analysis of all participants, with robust replication in Illumina

763

Zscore	<i>Meta-analysis</i>		<i>p-values</i>		Gene	EntrezID	Name
	p-value	BH p-value	FHS	Illumina			
-5.67	1.5x10 ⁻⁸	2.0x10 ⁻⁴	3.9x10 ⁻⁷	9.9x10 ⁻³	<i>ALAS2</i>	212	aminolevulinate, delta-, synthase 2
5.37	7.8x10 ⁻⁸	3.1x10 ⁻⁴	4.3x10 ⁻⁷	4.0x10 ⁻²	<i>HEATR5A</i>	25938	HEAT repeat containing 5A
-5.28	1.3x10 ⁻⁷	3.9x10 ⁻⁴	2.7x10 ⁻⁵	1.2x10 ⁻³	<i>PNP</i>	4860	purine nucleoside phosphorylase
-5.17	2.4x10 ⁻⁷	4.7x10 ⁻⁴	1.1x10 ⁻⁶	5.0x10 ⁻²	<i>STOM</i>	2040	stomatin
5.14	2.8x10 ⁻⁷	4.7x10 ⁻⁴	7.9x10 ⁻⁶	1.1x10 ⁻²	<i>RPS6KA5</i>	9252	ribosomal protein S6 kinase, 90kDa, polypeptide 5
-5.09	3.7x10 ⁻⁷	5.8x10 ⁻⁴	5.4x10 ⁻⁵	1.8x10 ⁻³	<i>MBNL3</i>	55796	Muscleblind-Like Splicing Regulator 3
-5.07	3.9x10 ⁻⁷	5.8x10 ⁻⁴	2.3x10 ⁻⁶	4.4x10 ⁻²	<i>RAD23A</i>	5886	RAD23 homolog A (<i>S. cerevisiae</i>)
5.02	5.1x10 ⁻⁷	6.8x10 ⁻⁴	7.7x10 ⁻⁵	1.7x10 ⁻³	<i>HNRNPA0</i>	10949	heterogeneous nuclear ribonucleoprotein A0
-4.90	9.5x10 ⁻⁷	1.1x10 ⁻³	2.0x10 ⁻⁵	1.5x10 ⁻²	<i>GID4</i>	79018	GID Complex Subunit 4
-4.88	1.1x10 ⁻⁶	1.1x10 ⁻³	9.3x10 ⁻⁶	3.4x10 ⁻²	<i>TFDP1</i>	7027	transcription factor Dp-1
4.80	1.6x10 ⁻⁶	1.4x10 ⁻³	9.9x10 ⁻⁶	4.7x10 ⁻²	<i>ARRDC3</i>	57561	arrestin domain containing 3
-4.77	1.7x10 ⁻⁶	1.5x10 ⁻³	1.8x10 ⁻⁵	3.3x10 ⁻²	<i>RIOK3</i>	8780	RIO kinase 3
-4.71	2.5x10 ⁻⁶	1.8x10 ⁻³	1.9x10 ⁻⁵	4.1x10 ⁻²	<i>NTAN1</i>	123803	N-terminal asparagine amidase
4.66	3.1x10 ⁻⁶	2.2x10 ⁻³	8.1x10 ⁻⁴	6.0x10 ⁻⁴	<i>PDE4D</i>	5144	phosphodiesterase 4D, cAMP-specific
-4.62	3.8x10 ⁻⁶	2.4x10 ⁻³	1.1x10 ⁻⁴	1.2x10 ⁻²	<i>RGCC</i>	28984	Regulator Of Cell Cycle
4.61	4.1x10 ⁻⁶	2.5x10 ⁻³	9.6x10 ⁻⁵	1.4x10 ⁻²	<i>POLR2B</i>	5431	polymerase (RNA) II (DNA directed) polypeptide B
-4.59	4.4x10 ⁻⁶	2.5x10 ⁻³	6.1x10 ⁻⁵	2.4x10 ⁻²	<i>EIF1B</i>	10289	eukaryotic translation initiation factor 1B
4.57	4.8x10 ⁻⁶	2.7x10 ⁻³	5.4x10 ⁻⁴	2.0x10 ⁻³	<i>CIRBP</i>	1153	cold inducible RNA binding protein
-4.57	4.8x10 ⁻⁶	2.7x10 ⁻³	6.2x10 ⁻⁵	2.7x10 ⁻²	<i>ASGR2</i>	433	asialoglycoprotein receptor 2
-4.55	5.3x10 ⁻⁶	2.9x10 ⁻³	3.8x10 ⁻⁵	4.5x10 ⁻²	<i>CA1</i>	759	carbonic anhydrase I

BH=Benjamini-Hochberg. Ordered by meta-analysis p-value. Showing top 20 results with nominal 'replication' (p<0.05) in Illumina. Duplicate gene entries excluded. Full table in Supplementary Table 1.

764
 765 **Table 3 |** Thirteen genes associated with muscle strength in age or gender specific subset analyses

766

<i>Subset</i>		<i>Meta-analysis of subset</i>			<i>p-values</i>		<i>Gene</i>	<i>Name</i>
<i>Group</i>	<i>N</i>	<i>Zscore</i>	<i>P-value</i>	<i>BH P-value</i>	<i>FHS</i>	<i>Illumina</i>		
Younger & Male	5379	-4.879	1.1 x10 ⁻⁶	9.5 x10 ⁻³	2.4 x10 ⁻⁶	1.3 x10 ⁻¹	<i>ASAP1</i>	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1
Younger	5379	4.495	6.9 x10 ⁻⁶	2.5 x10 ⁻²	7.5 x10 ⁻⁵	3.3 x10 ⁻¹	<i>PKN2</i>	protein kinase N2
Younger	5379	-4.426	9.6 x10 ⁻⁶	2.5 x10 ⁻²	2.9 x10 ⁻⁵	6.1 x10 ⁻¹	<i>RNF175</i>	ring finger protein 175
Younger	5379	-4.407	1.1 x10 ⁻⁵	2.5 x10 ⁻²	3.3 x10 ⁻⁵	2.4 x10 ⁻¹	<i>TMED5</i>	transmembrane emp24 protein transport domain containing 5
Younger	5379	4.235	2.3 x10 ⁻⁵	3.5 x10 ⁻²	4.2 x10 ⁻⁵	4.7 x10 ⁻¹	<i>ZFYVE27</i>	zinc finger, FYVE domain containing 27
Younger & Male	5379	-4.217	2.5 x10 ⁻⁵	3.5 x10 ⁻²	3.4 x10 ⁻³	2.5 x10 ⁻³	<i>GID8</i>	GID complex subunit 8
Younger	5379	-4	6.3 x10 ⁻⁵	4.2 x10 ⁻²	1.7 x10 ⁻⁴	1.2 x10 ⁻²	<i>PIK3R2</i>	phosphoinositide-3-kinase, regulatory subunit 2
Younger	5379	-3.961	7.5 x10 ⁻⁵	4.2 x10 ⁻²	7.3 x10 ⁻⁶	3.3 x10 ⁻¹	<i>PK4</i>	pyruvate dehydrogenase kinase, isozyme 4
Younger	5379	-3.961	7.5 x10 ⁻⁵	4.2 x10 ⁻²	2.0 x10 ⁻⁶	6.7 x10 ⁻¹	<i>CTNNA1</i>	catenin (cadherin-associated protein), alpha-like 1
Younger	5379	-3.937	8.3 x10 ⁻⁵	4.4 x10 ⁻²	7.6 x10 ⁻⁵	1.2 x10 ⁻¹	<i>COQ9</i>	coenzyme Q9
Male	3557	4.565	5.0 x10 ⁻⁶	1.9 x10 ⁻²	1.2 x10 ⁻⁵	1.1 x10 ⁻¹	<i>RAC1</i>	ras-related C3 botulinum toxin substrate 1
Male	3557	4.041	5.3 x10 ⁻⁵	4.3 x10 ⁻²	8.0 x10 ⁻⁵	1.9 x10 ⁻¹	<i>NDUFS1</i>	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa ...
Female	4224	-4.907	9.3 x10 ⁻⁷	2.5 x10 ⁻²	8.0 x10 ⁻⁴	1.1 x10 ⁻⁴	<i>DEFA4</i>	defensin, alpha 4, corticostatin

BH=Benjamini-Hochberg. Some genes were identified in in multiple groups; for these genes the statistics for the larger group are given; in all cases direction of association is the same. Ordered by subset, then P-value.

767

768

769 **Table 4** | Top 20 probes in the FHS analysis that did not map to a corresponding Illumina probe, ordered by p-value

770

Estimate	P-value	BH P-value	Chr	Start	Gene	Name
-0.0037	1.76x10 ⁻⁶	1.49x10 ⁻³	1	144989319		
-0.0028	4.89x10 ⁻⁶	2.88x10 ⁻³	9	37800563	<i>DCAF10</i>	DDB1 and CUL4 associated factor 10
0.0031	8.57x10 ⁻⁶	4.12x10 ⁻³	1	150522766	<i>ADAMTSL4</i>	ADAMTS(a disintegrin and metalloproteinase with thrombospondin motifs)-like 4
0.0038	9.29x10 ⁻⁶	4.15x10 ⁻³	16	89980135		
0.0040	1.95x10 ⁻⁵	5.97x10 ⁻³	2	162412847	<i>SLC4A10</i>	solute carrier family 4, sodium bicarbonate transporter, member 10
0.0029	3.24x10 ⁻⁵	7.74x10 ⁻³	16	9186734		
0.0031	3.93x10 ⁻⁵	8.51x10 ⁻³	1	44440179	<i>ATP6V0B</i>	ATPase, H ⁺ transporting, lysosomal 21kDa, V0 subunit b
-0.0027	4.10x10 ⁻⁵	8.64x10 ⁻³	3	63819562	<i>THOC7</i>	THO complex 7 homolog (Drosophila)
0.0032	5.38x10 ⁻⁵	9.81x10 ⁻³	20	3898284		
0.0019	1.49x10 ⁻⁴	1.70x10 ⁻²	20	1316212		
0.0018	1.89x10 ⁻⁴	1.91x10 ⁻²	3	128628719	<i>ACAD9</i> <i>KIAA1257</i>	acyl-CoA dehydrogenase family, member 9 KIAA1257
-0.0017	2.21x10 ⁻⁴	2.08x10 ⁻²	1	8021733	<i>PARK7</i>	parkinson protein 7
0.0036	3.18x10 ⁻⁴	2.69x10 ⁻²	12	57809458		
0.0010	3.20x10 ⁻⁴	2.70x10 ⁻²	1	153901987	<i>DENND4B</i>	DENN/MADD domain containing 4B
-0.0019	3.56x10 ⁻⁴	2.87x10 ⁻²	3	10157370	<i>BRK1</i>	BRICK1, SCAR/WAVE actin-nucleating complex subunit
-0.0028	3.84x10 ⁻⁴	2.97x10 ⁻²	2	95517671	<i>TEKT4</i>	tektin 4
-0.0014	3.95x10 ⁻⁴	3.03x10 ⁻²	1	247937996	<i>OR9H1P</i>	olfactory receptor, family 9, subfamily H, member 1 pseudogene
0.0018	6.01x10 ⁻⁴	3.89x10 ⁻²	9	124042152	<i>GSN-AS1</i>	GSN(gelsolin) antisense RNA 1
0.0028	6.41x10 ⁻⁴	4.03x10 ⁻²	6	20451305		
-0.0021	6.90x10 ⁻⁴	4.19x10 ⁻²	12	8024178	<i>NANOGP1</i> <i>NANOG</i>	Nanog homeobox pseudogene 1 Nanog homeobox

BH=Benjamini-Hochberg. Blank gene symbols were not annotated to a specific gene. Table continued in Supplementary Table 9. Ordered by P-value. Not all probes map to gene ID's.

771

772

773 **Table 5** | Ten biological processes were enriched in the genes associated with muscle strength in the analysis of all participants

774

Biological Process	C	O	rawP	adjP	Top Genes
hemoglobin metabolic process (GO:0020027)	14	5	3.2x10 ⁻⁷	0.0004	<i>ALAS2, AHSP, FECH, EIF2AK2, EPB42</i>
hemoglobin biosynthetic process (GO:0042541)	9	3	1.0x10 ⁻⁴	0.0139	<i>ALAS2, FECH, EIF2AK2</i>
innate immune response (GO:0045087)	539	18	3.8x10 ⁻⁵	0.0139	<i>RPS6KA5, IFI27, TLR5, DUSP3, FCGR1B</i>
negative regulation of protein metabolic process (GO:0051248)	460	16	6.5x10 ⁻⁵	0.0139	<i>IGF2BP2, CIRBP, BANP, GCLC, DUSP3</i>
posttranscriptional regulation of gene expression (GO:0010608)	371	14	8.0x10 ⁻⁵	0.0139	<i>HNRNPA0, IGF2BP2, EIF1B, CIRBP, ASGR2</i>
response to arsenic-containing substance (GO:0046685)	20	4	6.5x10 ⁻⁵	0.0139	<i>GCLC, GSTO1, FECH, DDX3X</i>
regulation of translational initiation by eIF2a ... (GO:0010998)	2	2	1.0x10 ⁻⁴	0.0139	<i>FECH, EIF2AK2</i>
regulation of eIF2 alpha phosphorylation by heme (GO:0010999)	2	2	1.0x10 ⁻⁴	0.0139	<i>FECH, EIF2AK2</i>
response to stress (GO:0006950)	2952	55	4.7x10 ⁻⁵	0.0139	<i>ALAS2, RPS6KA5, RAD23A, HNRNPA0, UBQLN1</i>
protoporphyrinogen IX metabolic process (GO:0046501)	11	3	2.0x10 ⁻⁴	0.0167	<i>FECH, ALAS2, EIF2AK2</i>

775

776 C=number of reference genes in the category; O=number of genes in the gene set and also in the category; rawP=p value from hypergeometric

777 test; adjP=p value adjusted by the multiple test adjustment (Benjamini-Hochberg); Top Genes=top five genes from pathway based on meta-

778 analysis P-value

779

780 **Figure Legend**

781

782 **Figure 1.** Gene transcripts associated with muscle strength in all participants

783

784 **A)** Compares the individual meta-analyses performed in the Illumina-cohorts and the FHS
785 separately. The dark grey points represent gene transcripts significantly associated with
786 muscle strength (FDR<0.05). The light grey points were not significant in this analysis. The
787 unfilled grey points were excluded due to significant heterogeneity (Cochran's Q-test $p < 0.05$
788 (54)). The solid grey line shows the trend across all the genes.

789 **B)** Shows the meta-analysis results by Manhattan plot. The dashed line indicates those
790 probes significantly associated with grip strength after Benjamini-Hochberg correction, the
791 solid line shows those significant after Bonferroni correction, for comparison.

792

793 **Additional Files**

794 Table.Supplementary.1.Meta-analysis-results.all-participants.xlsx

- 795 • Microsoft Excel file (.xlsx)
- 796 • Significant meta-analysis results in analysis of all participants

797

798 Table.Supplementary.2.Meta-analysis-results.younger-participants.xlsx

- 799 • Microsoft Excel file (.xlsx)
- 800 • Significant meta-analysis results in analysis of younger participants only

801

802 Table.Supplementary.3.Meta-analysis-results.male-participants.xlsx

- 803 • Microsoft Excel file (.xlsx)
- 804 • Significant meta-analysis results in analysis of male participants only

805

806 Table.Supplementary.4.Meta-analysis-results.female-participants.xlsx

- 807 • Microsoft Excel file (.xlsx)
- 808 • Significant meta-analysis results in analysis of female participants only

809

810 Table.Supplementary.5.Additional-information-for-genes.xlsx

- 811 • Microsoft Excel file (.xlsx)
- 812 • Additional information for prominent genes
- 813 • Tables 2 and 3 contain genes of particular robustness or interest that are here
- 814 described in further detail

815

816 Table.Supplementary.6.FHS-only-analysis-results.all-participants.xlsx

- 817 • Microsoft Excel file (.xlsx)
- 818 • Significant results for genes available in the FHS data only (all participants)

819

820 Table.Supplementary.7.FHS-only-analysis-results.younger-participants.xlsx

- 821 • Microsoft Excel file (.xlsx)
- 822 • Significant results for genes available in the FHS data only (younger participants)

823

824 Table.Supplementary.8.FHS-only-analysis-results.male-participants.xlsx

- 825 • Microsoft Excel file (.xlsx)
- 826 • Significant results for genes available in the FHS data only (male participants)
- 827
- 828 Table.Supplementary.9.WebGestalt-Gene-Ontology-pathway-analysis.xlsx
- 829 • Microsoft Excel file (.xlsx)
- 830 • Significantly enriched Gene Ontology terms in the 208 genes
- 831 • WebGestalt-determined Gene Ontology terms that are significantly enriched in the
- 832 208 genes found to be associated with muscle strength in the meta-analysis of all
- 833 participants
- 834
- 835 Table.Supplementary.10.WebGestalt-Phenotype-Ontology-pathway-analysis.xlsx
- 836 • Microsoft Excel file (.xlsx)
- 837 • Significantly enriched Human Phenotype Ontology terms in the 208 genes
- 838 • WebGestalt-determined Human Phenotype Ontology terms that are significantly
- 839 enriched in the 208 genes found to be associated with muscle strength in the meta-
- 840 analysis of all participants
- 841
- 842 Table.Supplementary.11.network_GO-enrichment_results.xlsx
- 843 • Microsoft Excel file (.xlsx)
- 844 • Ontology enrichment results for the four clusters described in the network-analysis
- 845 results section
- 846 • For each network of co-expressed genes with ≥ 10 genes in the network (n=4) we
- 847 present the results of a Gene Ontology enrichment analysis.
- 848
- 849 Table.Supplementary.12.Literature-search.xlsx
- 850 • Microsoft Excel file (.xlsx)
- 851 • Systematic search of the literature for 221 genes associated with muscle
- 852 • For each of the 221 genes found in this analysis to be associated with muscle
- 853 strength, GeneCards and PubMed were used to identify published research articles
- 854 that contain both the gene ID (including part synonyms) and the word “muscle”
- 855

