The complex genetics of gait speed

Citation for published version:

Digital Object Identifier (DOI):
10.18632/aging.101151

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Aging

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
The complex genetics of gait speed: genome-wide meta-analysis approach

ABSTRACT

Emerging evidence suggests that the basis for variation in late-life mobility is attributable, in part, to genetic factors, which may become increasingly important with age. Our objective was to systematically assess the contribution of genetic variation to gait speed in older individuals. We conducted a meta-analysis of gait speed GWASs in 31,478 older adults from 17 cohorts of the CHARGE consortium, and validated our results in 2,588 older adults from 4 independent studies. We followed our initial discoveries with network and eQTL analysis of candidate signals in tissues. The meta-analysis resulted in a list of 536 suggestive genome wide significant SNPs in or near 69 genes. Further interrogation with Pathway Analysis placed gait speed as a polygenic complex trait in five major networks. Subsequent eQTL analysis revealed several SNPs significantly associated with the expression...
SIGNIFICANCE

Despite promising results from candidate gene studies, a systematic and comprehensive examination of genetic determinants of gait speed in a large sample of older adults has been lacking. Furthermore, previous study samples have been too small to detect the expected modest genetic effects especially in such complex and polygenic encoded traits. To address these limitations, we conducted a meta-analysis of GWAS of gait speed in 31,478 older adults and validate our candidate signal in a cohort of 2588 older adults. Close to 600 candidate genetic variants have been linked to gait speed. Such efforts have provided us with an increased knowledge of the biological systems which impact on gait speed; this may contribute to improved treatment strategies and drug development to promote aging with grace.

INTRODUCTION

Gait speed has been described as the “sixth vital sign” because it is a core indicator of health and function in aging and disease [1]. Decline in gait speed is ubiquitous with aging in both men and women [2]. Gait speed is used to establish thresholds in community based activities, such as crossing a street [3, 4] or ambulating [5-7]. Slow gait speed is a consistent risk factor for disability, cognitive impairment, institutionalization, falls, hospitalization and mortality [8-10]. Improvement in gait speed is associated with better function and survival.

Many genetic and non-genetic factors (environment and disease) are likely to affect quantitative complex traits such as gait speed. There are individual differences in rates of decline in physical function, and genetic epidemiological studies provide a method for decomposing that variance into genetic and environmental sources. Twin studies suggest that genetic factors account for 15-51% of the variance of gait speed in older adults [11, 12]. Moreover, the contribution of genetic factors may increase with age [2, 11, 13-15]. Offspring of parents with exceptional longevity have better physical function and gait speed in age-specific comparisons to other individuals of comparable age and other characteristics [16, 17]. Effective gait requires the integration of many physiological systems, including the central and peripheral nervous system that create and execute the motor program, the musculoskeletal system that moves and supports the body, and the cardio-pulmonary function that provides perfusion of adequate nutrients and oxygen to all of the integrated parts. All these physiological systems can be affected by genetic variation. Given the many pathways that may contribute to gait impairment, effect sizes of individual genetic variants are expected to be limited.

Previous candidate gene studies have implicated several loci as relevant to gait speed. Single nucleotide polymorphisms (SNP) in the Angiotensin-Converting Enzyme (ACE) gene have been linked to better mobility response to exercise. The R577X polymorphism in the alpha-actinin-3 encoding gene (ACTN3) was associated with elite athletic performance, and muscle strength and power in the general population, especially in women [18]. There is evidence that APOE I/D and ACTN3 R577X polymorphisms, individually or in combination, have a significant influence on mobility and gait speed phenotypes in older women [19, 20]. Catechol-O-methyltransferase (COMT) polymorphisms have been associated with cognitive functions and gait speed [21]. The Met (158) Val polymorphism in COMT was linked to faster gait speed in older adults [21]. In addition, apolipoprotein E (APOE) genetic variation has been shown to influence the risk of gait speed decline [22-24]. Despite these promising results from candidate gene studies, a systematic and comprehensive examination of genetic determinants of gait speed in a large sample of older adults has been lacking. Furthermore, previous study samples have been too small to detect the expected modest genetic effects [25] especially in such complex and polygenic encoded traits [26].

To address these limitations, we conducted a meta-analysis of GWAS studies of gait speed in 31,478 older adults from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. We then tested our findings in a validation cohort of 2588 older adults participating in four independent studies.

RESULTS

Gait speed is considered a marker of health and fitness in aging. Slow gait in older adults is associated with increased risk of multiple adverse events including loss of independence, increased risk of disability, falls [27,
progression of age-related disease including dementia [29] and death [9]. Slowing of gait is multifactorial with major contributions from potentially modifiable risk factors such as physical inactivity, cognitive impairment, muscle weakness, pain, poor vision, falls and obesity [30]. Gait speed was timed over fixed distance, and reported in m/sec units.

In a meta-analysis of 31,478 subjects from 17 cohorts (Table 1, Supplementary Text) with ~2.5M imputed SNPs (Supplementary Table 1) 536 SNPs (202 were independent (LD, r² < 0.8) based on the HaploReg tool [31]) with p< 1×10⁻⁴ of which 88 (48 were independent signals) had a p-value less than 1×10⁻⁵ and one SNP attained a p-value of less than p< 1×10⁻⁶ (Table 2, Supplementary Table 2). The Q-Q plot (Supplementary Figure 1) did not provide evidence of inflation of test statistics. The Manhattan plot (Figure 1), highlighted 2 regions on chromosome 6 with high LD and suggestive association with gait speed (Regional plots [32] are displayed in Figure 2). These suggestive regions were further interrogated. Although none of the analyzed SNPs were genome wide significant (p< 5×10⁻⁸), one was present in the top ten (POM121L2), and 7 other genes (CEP112, PHACTR1, CNTN5, PTPRT, FHOD3, ADAMTS18, PRIM2) were highlighted based on the presence of SNPs with suggestive significant associations (p<0.0001) as well as low recombination rate and linkage disequilibrium r² >0.8 which may indicate significant signals in the segment (Figure 2, Supplementary Table 2, Supplementary Figure 2). The 536 suggestive SNPs (p< 1×10⁻⁴ in the screening group) were tested for validation in four additional cohorts, GENOA, LLS, MrOSGBG and MrOSMalmo (2588 subjects). Among the top 10 SNPs (six independent) only three exceeded nominal significance which slightly improved the combined meta-analysis significance for HLA-DPB1 SNPs (rs9501255, rs7763822 & rs3749985), however genome-wide levels of significance were not attained (Table 2).

### Candidate gene approach

None of the imputed variants previously reported as gait speed candidate genes such as ACE, ACTN3, COMT and APOE reached a nominally significant (p<0.05) threshold (Supplementary Table 3).

| Table 1. Demography of the screening and validation cohorts |
|----------------|----------------|----------------|
| **Cohort** | **Age, y** | **%Female** | **N with gait and GWAS** | **Gait protocol** |
| Screening | | | | |
| AGES | >65 | 58.9 | 3,166 | 6 meter walk |
| ARIC | >60 | 59.5 | 445 | 7.6 meter walk |
| BLSA | >60 | 49.5 | 334 | 6 meter walk |
| CHS | ≥65 | 60.9 | 3,184 | 6 meter walk |
| FHS | >65 | 56.1 | 2,384 | 4 meter walk |
| HABC | >70 | 47.1 | 1,482 | 6 meter walk |
| HRS | >65 | 56.4 | 5,073 | 2.5 meter walk |
| InCHIANTI | >60 | 55.8 | 898 | 4 meter walk |
| LBC1921 | 77-80 | 58.4 | 510 | 6 meter walk |
| LBC1936 | 67-71 | 49.5 | 1,001 | 6 meter walk |
| MrOS | ≥65 | None | 4,643 | 6 meter walk |
| ROSMAP | >60 | 69.2 | 1,646 | 2.5 meter walk |
| RS-I | >55 | 53 | 706 | 6 meter walk |
| RS-II | >55 | 51.8 | 813 | 6 meter walk |
| RS-III | >45 | 56.0 | 1,392 | 6 meter walk |
| SOF | ≥65 | 100 | 3,441 | 6 meter walk |
| TACOG | >60 | 42 | 360 | 6 meter walk |
| Total Screening | | | 31,478 | |
| Validation | | | | |
| GENOA | >60 | 55 | 471 | 7.6 meter walk |
| LLS | >60 | 47.2 | 235 | 4 meter walk |
| MrOSGBG | >69 | None | 960 | 6 meter walk |
| MrOSMalmo | >69 | None | 922 | 6 meter walk |
| Total Validation | | | 2,588 | |
Figure 1. Manhattan plot of meta-analysis of genome wide association studies of gait speed for ~2.5 million genotype and imputed SNPs. The blue line indicates the threshold used to select the 536 suggestive genome wide significant SNPs.

Table 2. Top 10 association meta-analysis results for gait speed

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.:Position</th>
<th>E/NE Allele</th>
<th>F E Allele</th>
<th>Closest Gene</th>
<th>Δ(kb)/gene location</th>
<th>Screening Set (n=31,478)</th>
<th>Validation Set (n=2,588)</th>
<th>Screening + Validation Set (n=34,066)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17527406</td>
<td>6:33709545</td>
<td>C/G</td>
<td>0.016</td>
<td>UQCC2(MNF3)</td>
<td>intron</td>
<td>0.040(0.007) 5.22E-7</td>
<td>0.014(0.032) 0.65</td>
<td>0.014(0.032) 0.65 6.883e-7</td>
</tr>
<tr>
<td>rs9501255*</td>
<td>6:33087321</td>
<td>T/C</td>
<td>0.038</td>
<td>HLA-DPB1</td>
<td>3' UTR</td>
<td>0.023(0.005) 1.53e-6</td>
<td>0.048(0.023) 0.04</td>
<td>0.048(0.023) 0.04 3.326e-7</td>
</tr>
<tr>
<td>rs7763822*</td>
<td>6:33092651</td>
<td>T/C</td>
<td>0.038</td>
<td>HLA-DPB1</td>
<td>3</td>
<td>0.023(0.005) 1.54e-6</td>
<td>0.047(0.023) 0.04</td>
<td>0.047(0.023) 0.04 3.440e-7</td>
</tr>
<tr>
<td>rs7746199*</td>
<td>6:33086656</td>
<td>C/G</td>
<td>0.038</td>
<td>HLA-DPB1</td>
<td>3' UTR</td>
<td>0.023(0.005) 1.55e-6</td>
<td>0.048(0.023) 0.04</td>
<td>0.048(0.023) 0.04 3.385e-7</td>
</tr>
<tr>
<td>rs12155739</td>
<td>8:102084750</td>
<td>C/T</td>
<td>0.166</td>
<td>NCALD</td>
<td>intron</td>
<td>-0.041(0.008) 2.04E-6</td>
<td>-0.024(0.032) 0.45</td>
<td>-0.024(0.032) 0.45 1.858e-6</td>
</tr>
<tr>
<td>rs3800318*</td>
<td>6:27295862</td>
<td>A/T</td>
<td>0.830</td>
<td>POM121L2</td>
<td>15</td>
<td>0.011(0.002) 1.58E-6</td>
<td>0.011(0.008) 0.19</td>
<td>0.011(0.008) 0.19 7.125e-7</td>
</tr>
<tr>
<td>rs13211166</td>
<td>6:27298161</td>
<td>A/T</td>
<td>0.190</td>
<td>POM121L2</td>
<td>11</td>
<td>0.011(0.002) 2.12E-6</td>
<td>0.010(0.009) 0.24</td>
<td>0.010(0.009) 0.24 1.136e-6</td>
</tr>
<tr>
<td>rs9403969</td>
<td>6:148622038</td>
<td>T/G</td>
<td>0.737</td>
<td>SASH1</td>
<td>70</td>
<td>0.009(0.002) 2.34e-6</td>
<td>0.005(0.007) 0.50</td>
<td>0.005(0.007) 0.50 2.351e-6</td>
</tr>
<tr>
<td>rs16897515*</td>
<td>6:27310241</td>
<td>A/C</td>
<td>0.161</td>
<td>POM121L2</td>
<td>missense</td>
<td>0.011(0.002) 2.41E-6</td>
<td>0.007(0.009) 0.42</td>
<td>0.007(0.009) 0.42 2.080e-6</td>
</tr>
</tbody>
</table>

*First gene segment, #second gene segment. E/NE-Effect, Non-Effect allele; F E-Frequency of Effect Allele; Δ-distance to proximal gene; HetPVal- Heterogeneity P Value.
Figure 2a

Figure 2b
Figure 2c

Figure 2d
Pathway analysis

We used the 536 suggestive SNPs to generate the network analysis, in which 283 SNPs representing 68 genes (Supplementary Table 4) were located in both the IPA dataset and the SeattleSeqAnnotation141 for SNP annotation (the remaining 253 SNPs did not map to a gene). Among the genes having the highest number of defining SNPs, were CEP112 (38 SNPs), PHACTR1 (23 SNPs), CNTN5 (19 SNPs), PTPRT (18 SNPs), FHOD3 (17 SNPs), ADAMTS18 (12 SNPs) and PRIM2 (11 SNPs). The vast majority of these genes’ products are located in the cytoplasm and plasma membrane while the rest are in the nucleus, extracellular space and other cellular spaces. Ten types of protein actions (enzyme, transporter, phosphatase, transcription regulator, kinase, ion channel, transmembrane receptor, translation regulator, ligand-dependent nuclear receptor and peptidase) are enumerated in Supplementary Table 5. Five of them serve as a biomarker for diagnosis, disease progression, prognosis, and unspecified application and five of them were targets for drug development including PRIM2, GABRA1, LYN, PRKCE and SCN11A. Five major putative disease and function networks were established using the candidate genes (based on the IPA software analysis significance classification) and were classified accordingly to cancer, gastrointestinal disease, organismal injury and abnormalities, neurological disease, cell and tissue morphology, cellular function, development and maintenance, amino acid metabolism, small molecule biochemistry, gene expression, cell-to-cell signaling and interaction, nervous system development and function, cellular assembly and organization. Seven genes were not mapped to any network (see Figure 3, Supplementary Table 6).

Figure 2. LocusZoom plots for the suggested top 10 SNPs (5 genes) associated with gait speed of the combined analysis. (A) POM121L2; (B) HLA-DPB1, (C) UQCC2 (MNF1), (D) SASH1, (E) NCALD. In each plot, the −log10 of p values are on the left y-axis; the SNP genomic position (HG19) on the x-axis; the estimated recombination rate from 1000 genomes Nov. 2014 EUR are on the right y-axis and plotted in blue. The most significant SNP is in purple diamond and plotted using the p value attained from the meta-analysis. SNPs are colored to reflect linkage disequilibrium (LD) with the most significant SNP in red (pairwise r2 from 1000 genomes Nov. 2014 EUR). Gene annotations are from the SeattleSeqAnnotation141.
By querying a large collection of eQTL results (listed in Supplementary Text), we obtained a long list of possible SNP relationships with gene expression (Supplementary Table 7). We also identified the strongest eQTL SNP for each particular transcript in each study. Those SNPs with low p-values (for association with gene expression, $p<10^{-8}$) and high LD ($D'>0.9$) with the functional variant, were picked as candidates of signal concordance between the eQTL signals and gait speed signal. Following this analysis, several transcripts including PRSS16 and WDSUB1 were highlighted (Supplementary Table 7). We also observed a relationship between a SNP and PTPRT expression (in liver tissue), which in addition to the meta-analysis and pathway analysis emphasized its potential functional link through its synaptic function and neuronal development, both of which may contribute to [33] gait speed. By emphasizing a strong relationship of the best eQTL with our queried SNPs, we likely underreport SNP-expression relationships due to missing LD information and the inability to project LD relationships for trans-eQTLs in the region.

Applying HaploReg v4.1 analysis to the 536 variants resulted in 9 categories (Supplementary Table 8): miscRNA (1 variant); snoRNA (2 variants); microRNA (4 variants); snRNA (9 variants); pseudogenes (14 variants); sequencing in progress (43 variants); LINC RNA (86 variants); and 372 variants within protein coding genes. In addition, some variants annotate to the same gene resulting in a total of 139 genes (protein-coding or non-coding). Of those genes, 6 are exceptionally long, containing over a million base-pairs, the longest of which is PTPRT coded by 1117219bp. The shortest genes are the ones coding for micro (MIR3143) or small nuclear (U7) RNAs at 63bp each. There is only partial information regarding the chromatin state of each variant. However, from the information gathered in the analysis we observed 14 transcription start sites and 245 enhancers (Supplementary Table 8).
individuals in four validation cohorts, we did not discover any genome-wide significant association with gait speed nor did we confirm gait speed associations with previously reported candidate genes (i.e. ACE, ACTN3, COMT and APOE) (Supplementary Table 3). However, our analyses revealed some potentially relevant SNPs that could be targeted for further analyses regarding their associations with gait speed.

Our results shed light on several candidate genetic polymorphisms that did not achieve genome-wide significance but which had multiple signals on the gene segment, an observation that supported the association with the trait of interest. In addition, these SNPs map to genes that were either linked to physiologic functions expected to influence gait speed (such as neuromuscular function, cardiac function and muscle health or brain function) ADAMTS18, a gene associated with bone mineral density, could be associated with gait speed if individuals with variants in this gene had suffered from fracture leading to slowing of gait [34]. In functional studies ADAMTS18 levels were significantly lower in subjects with non-healing skeletal fractures compared to normal subjects [35]. POM121L2 - an ion transport gene [36] - was listed in the top ten meta-analysis genes with four variants, making it a potential candidate for our study. This gene has been linked to schizophrenia, [37], suggesting a potential brain-related association with gait speed. One of the top candidates in our analysis was UQCC2 (also known as M19 or MNF1), a mitochondrial membrane protein that regulates skeletal muscle differentiation and insulin secretion [38]. Although UQCC2 function has a clear link to gait speed, the fact that in this study only one SNP found within UQCC2 demonstrated suggestive significance, which provides less confidence of a true association. NCALD, a calcium-binding protein, has been associated with diabetic nephropathy [39]. The region that was highlighted next to SASH1, a tumor suppressor gene, has multiple signals associated with gait speed. However, there is a high recombination rate between this region and the candidate gene (Figure 2), suggesting a higher dissociation between the gene and the signaled region. The last candidate from the top 10 SNP association list is HLA-DPB1, an immune response gene that has been linked to rheumatoid and inflammatory myopathies [40, 41]. Interestingly, one of its variants (rs7763822) was indicated in systemic sclerosis susceptibility in Korean subjects [42] suggesting a pleiotropic effect.

Similar to CEP112 variants, PHACTR1 regulates cardiac α-actin isoform ratio [44] and actomyosin assembly [45]; CNTN5 is associated with neuron function [46]; PTPRT regulates synaptic function and neuronal development [33] and serves as a genuine susceptibility locus for rheumatoid arthritis[33]; PHOD3, is a key regulator in the cardiac muscle [47] and sarcomere organization in striated muscle cells [48]; and PRIM2 is involved in DNA replication and transcription and is crucial for normal growth and development [49]. This list of genes repeatedly implicates associated signals that are important for neuromuscular function, cardiac function and muscle health, which could reasonably contribute to the complex trait of gait speed.

A second tier of locus with repetitive signals established among the 536 suggestive SNPs included PDZ3, which is implicated in muscle function and regeneration [50-52], CACNG3, a voltage-dependent calcium channel subunit [53] that was previously linked to ataxic phenotype in mice [54], ASTN2 that functions in neuronal migration [55] and that was associated with hip osteoarthritis susceptibility [56], SIM1 involved in coordinating muscle activity and generating rhythmic activity [57] and also associated with obesity [58], and MDGA2, which is required for proper development of cranial motoneuron subtypes [59].

The eQTL analysis (various tissues and cell types, listed in Supplementary Text) of the 536 suggestive SNPs reported a couple of candidate genes such as PRSS16, a gene encoding serine protease expressed exclusively in the thymus. PRSS16 was associated with exercise [60] and was linked to COMT (a candidate gene for gait speed [20]). Both are regulated by ZNF804a [61]. This link between the two genes (PRSS16 and COMT) may support our gait speed association results. Another candidate gene from our eQTL analysis was WDSUB1 a U-box ubiquitin ligases encoded protein which was associated with sudden cardiac death [62]. A link with cardiovascular diseases may indicate a potential cardiovascular effect on gait speed. The last candidate in this analysis is PTPRT, a gene that regulates synaptic function and neuronal development. It is possible that its link to gait speed (operates through its role in diabetes [63]). The fact that it was present in all three sets of analysis results may suggest a stronger candidate for further analysis.

The lead motif of the network analysis in all 5 disease networks was “cellular function”, however, the candidate SNPs from the multiple analysis strategies strongly suggested links to bone, skeleton, muscle and brain, incorporating development, structure and function. While our SNP associations did not achieve
genome wide significance, we believe that we demonstrated a potential link to gait speed. To exclude false positive signals, these associations should be pursued further in controlled experiments as well as animal models, which will increase our understanding of the biology of gait speed deterioration with aging. Such efforts would provide us with an increased knowledge of the biological systems which impact on gait speed; this may contribute to improved treatment strategies and drug development to promote aging with grace.

This study did not provide conclusive evidence for the genetics contributing to gait speed. While the large sample is a strength (and we have the power to detect smaller effects), the observed associations suggest that an even larger sample is required to establish genetic contributions to the gait speed phenotype. The individual effects of common SNPs for complex traits such as gait speed are expected to be very small. From studies of other polygenic complex traits, it has been observed that the number of discovered variants is strongly correlated with experimental sample size [64]. Another potential explanation why we did not observe genome wide significant associations is that there are many potential pathways that contribute to gait speed, including nervous system function (neuromuscular, central nervous system), musculoskeletal conditions such as sarcopenia and osteoarthritis, cardiovascular disease, visual function, psychological factors and other contributors. This complexity of phenotype may make it difficult to discover associations. Phenotype refinement may be a future approach to explore.

In summary, the lack of genome-wide significant signals from this moderately large sample of older adults suggests that larger samples (or study to sub-classify the gait speed phenotype) will be needed to identify SNP-based associations. Also, it may suggest that downstream mechanisms are more likely to make more important contributions to gait speed. Gait speed is a complex phenotype with many potential contributors; it is not unsurprising that it should be governed by multiple genes. However, we were able to use network analyses to define some potential networks of genes that might be of relevance for this phenotype. Future studies may be best positioned to focus on one network in more detail and to examine gene-environment or gene-behavior-environment interactions.

**METHODS**

**Subjects**

The Aging and Longevity Working Group of the CHARGE Consortium [65, 66], was formed to facilitate genome-wide association study meta-analyses of age associated diseases and phenotypes among multiple large and well-phenotyped cohorts of older individuals who underwent genotyping.

**Screening cohorts**

A combined cohort of 31,478 subjects age 60 years and older with timed walks constituted our discovery sample (Table 1). Timed walk at usual pace was converted to gait speed (m/s) to harmonize the phenotype across cohorts. Participants of the following 17 European descendent cohorts were included (Supplementary Material):

The Age, Gene/Environment Susceptibility-Reykjavik (AGES), The Atherosclerosis Risk in Communities (ARIC), Baltimore Longitudinal study on Aging (BLSA), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), Health, Aging, and Body Composition Study (HABC), Health and Retirement Study (HRS), Invecchiare in Chianti (InCHIANTI), Lothian Birth Cohorts 1921 (LBC1921) and 1936 (LBC1936), Osteoporotic Fractures in Men Study (MrOS), The Religious Orders Study and Rush Memory and Aging Project (ROSMAP), Rotterdam Study (RS-I, -II, -III), Study of Osteoporotic Fractures (SOF), Tasmanian Study of Cognition and Gait (TASCOG) (Table 1, Supplementary Text). All participants with gait speed assessments including participants who were able to walk with assistance of a cane were included in this analysis. Exclusion criteria included missing gait assessments and inability to walk (Supplement Text).

**Validation cohorts**

The validation cohort consisted of 2,588 subjects (>60 years) from the Genetic Epidemiology Network of Arteriopathy (GENOA), Leiden Longevity Study (LLS), Osteoporotic Fractures in Men Study (MrOS), Sweden, Malmö[MrOSMalmö] and Gothenburg [MrOSGBG] studies (Table 1, Supplement text). Together these cohorts reach the minimum number of subjects required for sufficient statistical power (Our power calculation shows that given a fixed sample size (n=2500) our analysis will have >80% power to detect MAF=0.01, alpha<0.0001) to validate significant signal(s) from the screening cohort using the same harmonized gait speed phenotype. Results from the screening and validation cohorts were meta-analyzed.

**Phenotype definition**

The different methods of assessing gait speed in individual cohorts are described in Table 1 and
Variability in the methods of assessing gait speed in the participating cohorts included differences in distance walked (8 to 25 feet) and measurement techniques (instrumented walkway versus stopwatch). Previous reports including 4 cohorts from this report (CHS, HABC, InCHIANTI and SOF) have suggested that there is a high correlation (r^2>0.9) between the different methods of measuring gait speed [9, 24, 67]. The mean overall gait speed was 1.13±0.25 m/sec, and varied from 0.66 ± 0.16 m/sec to 1.66 ± 0.41 m/sec in the individual cohorts (Supplementary Table 9, Supplementary Figure 3).

Genotyping

A structured, pre-specified analytical plan was applied to each of the 17 cohorts included in the screening sample. Genome-wide analysis of imputed genotypes, summarized in Supplementary Text, were conducted in each cohort. Imputation (using either BimBam or MACH) resulted in approximately 2.5 million HapMap SNPs being available for analysis. Imputation details, QC and SNP count per cohort can be found in Supplementary Text and Supplementary Table 1.

Exclusion criteria for SNP in each of the 21 cohorts (screening and validation) included: 1) minor allele frequency (MAF) < 0.005); 2) imputation quality (R^2 or oevar_imp < 0.3); and for the meta-analysis, SNPs with average MAF ≤ 0.01 and total N < 15,000.

Cohort-specific analyses

Multiple linear regression of imputed SNP dosages on gait speed was performed using an additive model, i.e. as a count of the number of variant alleles present (1 degree of freedom). Sex-combined analysis was performed. Adjustment for age (at time of exam), sex, study site (for cohorts with multiple sites), principal components to control for population stratification, height, and presence of osteoarthritis (yes/no) if available were included. For cohorts with osteoarthritis data available, the analysis was done excluding participants with osteoarthritis (Supplementary Text and Supplementary Table 10).

Meta-analysis

Inverse variance weighted meta-analysis was performed on summary statistics of the cohort-level association analyses. Meta-analysis of gait speed (Screening and validation cohorts were analyzed separately as well as together (joint meta-analysis)) was performed using METAL [68] with a fixed effects model of beta estimates and standard errors from each cohort. In addition, we applied heterogeneity test between studies (on both screening and validation cohorts) using METAL. A p-value threshold (Bonferroni-adjusted) of p<5x10^-8 was used to indicate genome-wide statistical significance.

Pathway analysis

We assembled a list of 536 meta-analyzed SNPs (representing 69 genes) that were highly suggestively associated (p < 1 x 10^-4) with gait speed. This list resulted in 67 candidate genes (Annotated by Ingenuity Pathway Analysis (IPA) and SeattleSeqAnnotation) being identified which were used in the IPA analysis (www.ingenuity.com). The resulting classification of networks, pathways, biological processes and molecular functions are represented in tables and graphic format (Figure 3, Supplementary Table 4, 5 and 6).

Expression quantitative trait loci (eQTL) analysis

We examined existing eQTL resources for the candidate suggestive list of 536 SNPs (p<10^-4) to further explore their biological and functional relevance to gait speed (Supplementary Text). We queried these SNPs against an eQTL database (listed in Supplementary Text) containing eQTL results from over 100 studies across a wide range of tissues. A general overview of a subset of >50 eQTL studies has been published [69], with specific citations for the included datasets included in the Supplementary Material.

Further we applied the HaploReg v4.1 annotation tool for TF analysis of 536 SNPs suggestively associated with gait speed.

CONFLICTS OF INTEREST

The authors of this manuscript have no conflict of interests to declare.

FUNDING

Acknowledgments and funding for the participating studies are listed below.

AGES

This study has been funded by NIH contracts N01-AG-1-2100 and 271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.
ARIC

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. This project was also partially supported from NIH grant R01-HL093029.

BLSA

The BLSA was supported by the Intramural Research Program of the NIH, National Institute on Aging.

CHS

The CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN26820080007C, HHSN268200960009C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://chs-nhlbi.org/. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

FHS

The Framingham Heart Study phenotype-genotype analyses were supported by NIA R01AG29451 (JMM, KLL). The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. The investigators thank the participants and the staff for their support of the Framingham Heart Study. Dr Kiel's effort was supported by NIAMS grant RO1 AR41398.

HABC

This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant R01 AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

HRS

HRS is supported by the National Institute on Aging (NIA U01 AG009740). Genotyping was funded separately by NIA (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.

InCHIANTI

The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).

LBC1921 and LBC1936

We thank the cohort participants and team members who contributed to these studies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype
collection in the Lothian Birth Cohort 1936 was supported by Research Into Ageing (continues as part of Age UK The Disconnected Mind project). Genotyping of the cohorts was funded by the BBSRC. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and Medical Research Council (MRC) is gratefully acknowledged.

MrOS

The Osteoporotic Fractures in Men (MrOS) Study is supported by National Institutes of Health funding. The following institutes provide support: the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Institute on Aging (NIA), the National Center for Research Resources (NCRR), and NIH Roadmap for Medical Research under the following grant numbers: U01 AR45580, U01 AR45614, U01 AR45632, U01 AR45647, U01 AR45654, U01 AR45583, U01 AG18197, U01-AG027810, and UL1 RR024140. The National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) provides funding for the MrOS ancillary study ‘Replication of candidate gene associations and bone strength phenotype in MrOS’ under the grant number R01-AR051124. The National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) provides funding for the MrOS ancillary study ‘GWAS in MrOS and SOF’ under the grant number RC2AR058973.

ROSMAP

The data from the Rush Alzheimer’s Disease Center used in these analyses was supported by National Institute on Aging grants P30AG10161, R01AG17917, R01AG15819, R01AG30146, the Illinois Department of Public Health, R01NS078009 and the Translational Genomics Research Institute.

RS-I, -II, -III

The generation and management of GWAS genotype data for the Rotterdam Study are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. This research is supported by the Dutch Technology Foundation STW, which is part of the NWO, and which is partly funded by the Ministry of Economic Affairs. MAI is supported by ZonMW grant number 916.13.054.

SOF

The Study of Osteoporotic Fractures (SOF) is supported by National Institutes of Health funding. The National Institute on Aging (NIA) provides support under the following grant numbers: R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, and R01 AG027576.

TASCOG

All participants and research staff who took part in TASCOG are duly acknowledged, in addition to funding bodies including the National Health and Medical Research Council (NHMRC) of Australia, and the National Institutes of Health (NIH) USA.

GENOA

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054457, HL087660, HL119443) and the National Institute of Neurological Disorders and Stroke (NS041558) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (S.T.T, Mariza de Andrade, Julie Cunningham) and was made possible by the University of Texas Health Sciences Center (Eric Boerwinkle, Megan Grove-Gaona). We would also like to thank the families that participated in the GENOA study.

LLS

The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2011) under grant agreement number 259679. This study was financially supported by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), by Unilever Colworth and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007).
MrOS Sweden

This study was supported by the Swedish Research Council and by grants from the Swedish government (under the Avtal om Läkarutbildning och Medicinsk Forskning [Agreement for Medical Education and Research]), the Lundberg Foundation, the Torsten Söderberg Foundation, the Novo Nordisk Foundation and the Knut and Alice Wallenberg Foundation.

REFERENCES


19. Lima RM, Leite TK, Pereira RW, Rabelo HT, Roth SM,


doi: 10.1371/journal.pone.0031815


doi: 10.1016/j.neuron.2013.10.060


Cohort description

The Age, Gene/Environment Susceptibility-Reykjavik (AGES)

The Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907–1935 and living in Reykjavik in 1967 [1]. A total of 19381 attended, resulting in 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow-up and was examined in all stages. One group was designated a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGES-Reykjavik study re-examined 5764 survivors of the original cohort who had participated before in the Reykjavik Study. Of those, 3,219 have genomic genotypes and only 3,166 went through gait assessment that included 6 meter walk in usual pace.

The Atherosclerosis Risk in Communities (ARIC)

The ARIC study is a population-based cohort study of atherosclerosis and clinical atherosclerotic diseases [2]. At its inception (1987-1989), 15,792 men and women, including 11,478 white and 4,266 black participants were recruited from four U.S. communities: Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi. In the first 3 communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing. Between 2004 and 2006, the ARIC study re-examined 5764 survivors of the original cohort who had participated before in the Reykjavik Study. Of those, 3,219 have genomic genotypes and only 3,166 went through gait assessment that included 6 meter walk in usual pace.

The Atherosclerosis Risk in Communities (ARIC)

The ARIC study is a population-based cohort study of atherosclerosis and clinical atherosclerotic diseases [2]. At its inception (1987-1989), 15,792 men and women, including 11,478 white and 4,266 black participants were recruited from four U.S. communities: Suburban Minneapolis, Minnesota; Washington County, Maryland; Forsyth County, North Carolina; and Jackson, Mississippi. In the first 3 communities, the sample reflects the demographic composition of the community. In Jackson, only black residents were enrolled. Participants were between age 45 and 64 years at their baseline examination in 1987-1989 when blood was drawn for DNA extraction and participants consented to genetic testing. Between 2004 and 2006, participants who had undergone magnetic resonance scanning at the third ARIC visit were invited to participate in the ARIC MRI study [3]. Gait assessment was performed on 1134 ARIC participants. Time to walk 25 feet (7.62 m) at the participants’ usual pace was recorded in an unobstructed corridor with a stop watch. Four hundred and forty five participants of European ancestry with genome-wide genotype data and a gait speed measurement were enrolled to this study.

Baltimore Longitudinal study on Aging (BLSA)

The Baltimore longitudinal study on Aging (BLSA) study is a population-based study aimed to evaluate contributors of healthy aging in the older population residing predominantly in the Baltimore-Washington DC area [4]. Starting in 1958, participants are examined every one to four years depending on their age. Currently there are approximately 1100 active participants enrolled in the study. Blood samples were collected for DNA extraction, and genome-wide genotyping was completed for 1231 subjects using Illumina 550K. This analysis focused on a subset of the participants (N=334) with European ancestry with data on walking speed (6 meter walk in normal pace). The BLSA has continuing approval from the Institutional Review Board (IRB) of Medstar Research Institute.

Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for CHD and stroke in adults ≥65 years conducted across four field centers [5].The original predominantly Caucasian cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons were enrolled for a total sample of 5,888. Only 3980 CHS participants who were free of CVD at baseline, consented to genetic testing, and had DNA available for genotyping were GWASed. Finally, to maintain race homogeneity we picked 3184 Caucasian with gait speed (4.6 meter walk normal pace) and genome wide assessments to participate in the current study.

Framingham Heart Study (FHS)

The FHS is a longitudinal community-based multi-generational study funded by the National Heart Lung and Blood Institute [6]. The Original cohort (Gen1) has undergone 32 biennial examinations since 1948; the Offspring cohort (Gen2) has participated in 9 exams from 1971 onwards, and the Omni group 1 cohort in 4 examinations from 1994 onwards. The Gen3 and Omni group 2 cohorts completed 2 examinations since 2002 and are currently starting the third examination cycle (April 2016). All participants undergo extensive research examinations and surviving Original cohort, Offspring and Gen 3 participants had genome-wide genotyping with the Affymetrix 500K Array Set and 50K Human Gene Focused Panel available at the start of this study [7]. At Offspring exam 8 (2005-2008) and Original cohort exam 26 (1999-2001), participants were asked to walk a 4 meter course at a normal pace while being timed with a stop watch by trained technicians. The usual pace walk was repeated and the faster of the two walks was used for analysis. Participants were excluded if under age 60. The final sample included 2384 participants (56.1% women), mean age 72.4 (SD 8.5) years (range 60 to 98) with gait speed and genomic
genotyping assessed. Informed consent was obtained at each attended exam and the Boston University Medical Center Institutional Review Board approved the protocol for all examinations.

Health, Aging, and Body Composition Study (HABC)

The Health Aging and Body Composition (HABC) Study is a NIA-sponsored cohort study of the factors that contribute to incident disability and the decline in function of healthier older persons, with a particular emphasis on changes in body composition in old age. Between March 1997 and July 1998, 3075 70-79 year old community-dwelling adults (41% African-American) were recruited to participate in the Health ABC Study; characteristics of the cohort have been described elsewhere [8]. Medicare beneficiary listings were used to recruit in metropolitan areas surrounding Pittsburgh, Pennsylvania, and Memphis, Tennessee. Eligibility criteria included having no difficulty walking one-quarter of a mile, climbing 10 steps, or performing activities of daily living (transferring, bathing, dressing, and eating); no history of active treatment for cancer in the prior 3 years; and no plans to move from the area within 3 years. Genotyping was successful for 2,802 unrelated individuals (1663 Caucasians and 1139 African Americans). To reduce race bias we include only Caucasians of which 1482 have their gait speed assessed in normal pace (6 meter walk) have enrolled to the study.

Health and Retirement Study (HRS)

The Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50 [9]. The current sample is over 26,000 persons in 17,000 households. Respondents are interviewed every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006 and 2008. Gait speed was measured only on respondents ≥ 65 years of age. Respondents were removed if they had gait velocities <0.05 or gait velocities > 5sd from the mean. A total of 5,073 subjects who have both a measure of gait speed (2.5 meter walk at a normal pace) and high quality imputed genomic genotypes were included in the analysis.

Invecchiare in Chianti (InCHIANTI)

The InCHIANTI study is a population-based epidemiological study aimed at evaluating the factors that influence mobility in the older population living in the Chianti region in Tuscany, Italy [10]. The details of the study have been previously reported. Briefly, 1616 residents were selected from the population registry of Greve in Chianti (a rural area; 11,709 residents with 19.3% of the population greater than 65 years of age), and Bagno a Ripoli (Antella village near Florence; 4,704 inhabitants, with 20.3% greater than 65 years of age). The participation rate was 90% (n=1453), and the subjects ranged between 21-102 years of age. Overnight fasted blood samples were for genomic DNA extraction. Illumina Infinium HumanHap 550K SNP arrays were used for genotyping. Data from 898 subjects were used for this analysis with genetic and walking speed (4 meter walk in normal pace) data. The study protocol was approved by the Italian National Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore, MD).

Lothian Birth Cohorts 1921 (LBC1921) and 1936 (LBC1936)

The Lothian Birth Cohorts include surviving participants from the Scottish Mental Surveys of 1932 or 1947 (SMS1932 and SMS1947), having been born, respectively in 1921 (LBC1921) and 1936 (LBC1936) [11-13]. The LBC1921 cohort consists of 550 relatively healthy individuals, 316 females and 234 males, assessed on cognitive and medical traits at about 79 years of age. When tested, the sample had a mean age of 79.1 years (SD = 0.6). The LBC1936 consists of 1091 relatively healthy individuals assessed on cognitive and medical traits at about 70 years of age. At baseline the sample of 548 men and 543 women had a mean age 69.6 years (SD = 0.8). They were all Caucasian and almost all lived independently in the Lothian region (Edinburgh city and surrounding area) of Scotland. Genotyping was performed at the Wellcome Trust Clinical Research Facility, Edinburgh. Among participants with genome-wide data and gait speed assessment (6 meter walk in normal pace), 510 (LBC1921) and 1001 (LBC1936) individuals were available for the present analysis.

Osteoporotic Fractures in Men Study (MrOS)

The Osteoporotic Fractures in Men Study (MrOS) is a multi-center prospective, longitudinal, observational study of risk factors for vertebral and all non-vertebral fractures in older men, and of the sequelae of fractures in men [14, 15]. MrOS study population consists of 5,994 community dwelling, ambulatory men aged 65 years or older from six communities in the United States (Birmingham, AL; Minneapolis, MN; Palo Alto, CA; Monongahela Valley near Pittsburgh, PA; Portland, OR; and San Diego, CA). Inclusion criteria were designed to provide a study cohort that is representative
of the broad population of older men. Genomic DNA from participants in the Osteoporotic Fractures in Men (MrOS) Study was extracted from whole blood samples collected at the baseline visit using the Flexigene protocol (Qiagen, Valencia, CA, USA) at the University of Pittsburgh. Among the 5994 MrOS participants enrolled at the baseline visit, 5130 samples with whole genome genotyping data that passed QC. Of which, only 4,643 with gait speed assessed in normal pace (6 meter walk) were enrolled to the study.

**The Religious Orders Study and Rush Memory and Aging Project (ROSMAP)**

Data came from 2 community based cohort studies of aging and dementia, the Religious Orders Study and Rush Memory and Aging Project (ROSMAP). Details about the study design have been described previously [16, 17]. Both studies were approved by the institutional review board of Rush University Medical Center. Participants were free of known dementia at enrollment and agreed to annual clinical evaluation and brain donation at the time of death. An informed consent and an Anatomic Gift Act form were obtained from each participant. The follow-up rate among the survivors exceeds 90%. The two studies are conducted by the same team of investigators and share a large common core of test batteries, which allows combined analysis of the data. Gait speed was derived by timing with a stop watch how long it took a participant to walk 8 feet (2.5m) at their usual pace [18]. DNA was extracted from whole blood, lymphocytes, or frozen postmortem brain tissue. Genotyping was performed at the Broad Institute’s Center for Genotyping and the Translational Genomics Research Institute [19]. Among participants with genome-wide data and gait speed assessment 1,646 individuals were available for the present study.

**Rotterdam Study (RSI, -II, -III)**

The Rotterdam Study is a population-based study in Rotterdam that currently investigates 14,926 inhabitants from a suburb of the city aged 45 years or over. Participants were enrolled during three recruitment phases – in 1990 (cohort 1), 2000 (cohort 2), and 2006 (cohort 3) [20, 21]. Visits to the research center are planned every 3-4 years for various medical examinations. Genotyping was successfully performed on 11,496 participants. Gait assessment was introduced in the study protocol in 2009. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands, implementing the Wet Bevolkingsonderzoek: ERGO (Population Studies Act: Rotterdam Study). All participants provided written informed consent to participate in the study and to obtain information from their treating physicians. Gait assessment of 3651 subjects included 5.79-m long pressure-activated walkway (GAITRite Platinum; CIR systems, Sparta, NJ: 4.88-m active area; 120-Hz sampling rate) [22, 23]. Follow thorough exclusion the reminder 2911 subjects were genomic genotyped and imputed to the HapMap 2 reference panel.

**Study of Osteoporotic Fractures (SOF)**

The Study of Osteoporotic Fractures (SOF) is a prospective multicenter study of risk factors for vertebral and non-vertebral fractures [24]. The cohort is comprised of 9704 community dwelling women 65 years old or older recruited from populations-based listings in four U.S. areas: Baltimore, Maryland; Minneapolis, Minnesota; Portland, Oregon; and the Monongahela Valley, Pennsylvania. Women enrolled in the study were 99% Caucasian with African American women initially excluded from the study due to their low incidence of hip fractures. The SOF participants were followed up every four months by postcard or telephone to ascertain the occurrence of falls, fractures and changes in address. To date, follow-up rates have exceeded 95% for vital status and fractures, a review of pre-operative radiographs. The SOF study recruited only women. Among the 9704 SOF participants enrolled at the baseline visit, 3625 samples with whole genome genotyping data that passed QC. Of which, only 3,441 with gait speed assessed in normal pace (6 meter walk) were enrolled to the study.

**Tasmanian Study of Cognition and Gait (TASCOG)**

TASCOG is a study of cerebrovascular mechanisms underlying gait, balance and cognition in a population-based sample of Tasmanian people aged at least 60 years [25]. Individuals aged 60–86 years (n = 395) living in Southern Tasmania, Australia, were randomly selected from the electoral roll between 2006 and 2008 to participate in the study. Individuals were excluded if they lived in a nursing home, had a contraindication for magnetic resonance scanning (MRI) or were unable to walk without a gait aid. The response rate was 55%, and genotyping was performed at the Diamantina Institute, University of Queensland. The study was approved by the Human Health and Medical Research Ethics Committee, University of Tasmania. Genomic data and gait speed assessment (GAITRite) were available for 360 subjects that are part of this study.

**Genetic Epidemiology Network of Arteriopathy (GENOA)**

The Genetic Epidemiology Network of Arteriopathy (GENOA) study consists of hypertensive sibships
recruited for linkage and association studies in order to identify genes that influence blood pressure and its target organ damage [26]. In the initial phase of the GENOA study (Phase I: 1996-2001), all members of sibships containing ≥ 2 individuals with essential hypertension clinically diagnosed before age 60 were invited to participate, including both hypertensive and normotensive siblings. In the second phase of the GENOA study (Phase II: 2000-2004), 1239 European American participants were successfully re-recruited to measure potential target organ damage due to hypertension. From 2001-2006, Phase II GENOA participants that had a sibling willing and eligible to participate underwent a neurocognitive testing battery to assess several domains of cognitive and neurological functioning, including the assessment of gait speed (N=967). Participants were excluded from this analysis if they were less than 60 years of age or gait velocities >1.9m/s. The sample includes 471 European Americans (55.0% female) with imputed genotypes and a measure of gait speed on a 25 foot (7.6 meter) walking course.

Leiden Longevity Study (LLS)

The LLS has been designed to investigate biomarkers of healthy ageing and longevity [27] and has been described in detail previously [28]. It is a family-based study consisting of 1,671 offspring of 421 nonagenarian sibling pairs of Dutch descent, and their 744 partners. DNA from the LLS was extracted from white blood cells at baseline using conventional methods and genotyping was performed with Illumina Human660W-Quad and OmniExpress BeadChips (Illumina, San Diego, CA, USA). Imputation was performed with IMPUTE using the HapMap 2 reference panel [29]. Walking speed at usual pace was determined over 4 meters. Among participants with genome-wide data and gait speed assessment 235 individuals were available for the present study.

Osteoporotic Fractures in Men Study (MrOS) Sweden (Malmö[MrOSMalmo] and Gothenburg [MrOSGBG])

The Osteoporotic Fractures in Men (MrOS) study is a multicenter, prospective study including older men in Sweden, Hong Kong and the United States. The MrOS Sweden study (n=3014) [30] consists of three subcohorts from three different Swedish cities (n=1005 in Malmö, n=1010 in Gothenburg, and n=999 in Uppsala). Study subjects (men aged 69 to 81 years) were randomly identified using national population registers. A total of 62% of the MrOS Sweden subjects who have both GAIT information and high quality imputed genomic genotypes participated in the study (n=922 in Malmö, n=960 in Gothenburg). To be eligible for the study, the subjects had to be able to walk without assistance, provide self-reported data, and sign an informed consent. The study was approved by the ethics committees at the Universities of Gothenburg, Lund, and Uppsala. Informed consent was obtained from all study participants. Genome-wide genotyping was performed in the MrOS Gothenburg and MrOS Malmö sub cohorts. Walking speed at usual pace was determined over 6 meters. Both duration of the walk and the number of steps were measured.

Expression quantitative trait loci (eQTL) analysis

A general overview of a subset of >50 eQTL studies has been published [31], with specific citations for >100 datasets included in the current query following here. Blood cell related eQTL studies included fresh lymphocytes [32], fresh leukocytes [33], leukocyte samples in individuals with Celiac disease [34], whole blood samples [35-54], lymphoblastoid cell lines (LCL) derived from asthmatic children [55, 56], HapMap LCL from 3 populations [57], a separate study on HapMap CEU LCL [58], additional LCL population samples [59-65], neutrophils [66, 67], CD19+ B cells [68], primary PHA-stimulated T cells [62, 65], CD4+ T cells (2083654), peripheral blood monocytes [59, 68-72], long non-coding RNAs in monocytes [73] and CD14+ monocytes before and after stimulation with LPS or interferon-gamma [74], CD11+ dendritic cells before and after Mycobacterium tuberculosis infection [75] and a separate study of dendritic cells before or after stimulation with LPS, influenza or interferon-beta [76]. Micro-RNA QTLs [77, 78], DNase-I QTLs [79], histone acetylation QTLs [80], and ribosomal occupancy QTLs [81] were also queried for LCL. Splicing QTLs [82] and micro-RNA QTLs [83] were queried in whole blood.

Non-blood cell tissue eQTLs searched included omental and subcutaneous adipose [37, 48, 54, 63, 84], visceral fat [37] stomach [84], endometrial carcinomas [85], ER+ and ER- breast cancer tumor cells [86], liver [37, 84, 87-90], osteoblasts [91], intestine [92] and normal and cancerous colon [93, 94], skeletal muscle [37, 95], breast tissue (normal and cancer) [96, 97], lung [48, 98-101], skin [48, 59, 63, 102], primary fibroblasts [62, 65, 103], sputum [104], pancreatic islet cells [105], prostate [106], rectal mucosa [107], arterial wall [37] and heart tissue from left ventricles [48, 108] and left and right atria [109]. Micro-RNA QTLs were also queried for gluteal and abdominal adipose [110] and liver [111]. Methylation QTLs were queried in pancreatic islet cells [112]. Further mRNA and micro-RNA QTLs were queried from ER+ invasive breast cancer samples, colon-, kidney renal clear-, lung- and prostate-adenocarcinoma samples [113].
Brain eQTL studies included brain cortex [36, 72, 114-116], cerebellar cortex [117], cerebellum [115, 118-121], frontal cortex [117, 119, 121], gliomas [122], hippocampus [117, 119], inferior olivary nucleus (from medulla) [117], intralobular white matter [117], occipital cortex [117], parietal lobe [120], pons [121], pre-frontal cortex [118, 119, 123, 124], putamen (at the level of anterior commissure) [117], substantia nigra [117], temporal cortex [115, 117, 119, 121], thalamus [119] and visual cortex [118].

Additional eQTL data was integrated from online sources including ScanDB, the Broad Institute GTEx Portal, and the Pritchard Lab (eqtl.uchicago.edu). Cerebellum, parietal lobe and liver eQTL data was downloaded from ScanDB and cis-eQTLs were limited to those with P<1.0E-6 and trans-eQTLs with P<5.0E-8. Results for GTEx Analysis V4 for 13 tissues were downloaded from the GTEx Portal and then additionally filtered as described below www.gtexportal.org: thyroid, leg skin (sun exposed), tibial nerve, aortic artery, tibial artery, skeletal muscle, esophagus mucosa, esophagus muscularis, lung, heart (left ventricle), stomach, whole blood, and subcutaneous adipose [48]. Splicing QTL (sQTL) results generated with sQTLseeker with false discovery rate P≤0.05 were retained. For all gene-level eQTLs, if at least 1 SNP passed the tissue-specific empirical threshold in GTEx, the best SNP for that eQTL was always retained. All gene-level eQTL SNPs with P<1.67E-11 were also retained, reflecting a global threshold correction of P=0.05/(30,000 genes X 1,000,000 tests).

### Analysis Plan

1) Analysis Plan:

a. Imputation: all cohorts have imputed to HapMap, using either BimBam or MACH.

b. Cohort-specific analyses

i. Multiple linear regression of imputed SNPs on gait speed (m/s)

ii. All analyses will be sex-combined

iii. SNPs will be coded as additive model as a count of the number of variant alleles present (1 degree of freedom).

iv. Covariate adjustment:

1. age (at time of exam)
2. gender
3. study site (for cohorts with multiple sites)
4. principal components that control for population stratification (in some cohorts)
5. height
6. Osteoarthritis*

   c. Meta-analysis: Inverse variance weighted meta-analysis to be performed on summary statistics of imputed data. Meta-analysis of gait speed outcome will be performed using a fixed effects model of beta estimates and standard errors from each cohort.

   i. Significance threshold: A threshold of p-value 5x10-8 will be used to determine genome-wide statistical significance.

*Cohorts with Osteoarthritis measurement will add a variable yes/no converted to 1/0 for any sort of osteoarthritis and will provide two analyses one with and one without this variable.

- Cohorts without Osteoarthritis measurement will stick to the original analytic plan.
- Cohorts with Osteoarthritis measurement will provide:

  I. Analysis which includes everyone with or without osteoarthritis.

  II. Analysis only for the one without this variable.

  d. Cohorts with mixed ethnicity will be stratified by race. Meta analysis will test both possibilities with the additional race as a separate cohort and without.

2) Data format: The data delivery format for the meta-analysis will be according to the CHARGE protocol for file sharing.

a. ShareSpaces, a secure web-based file-sharing system implemented by the University of Washington's Catalyst computing group, will be used.

b. The following variables should be included when sharing imputed results for meta-analysis (Table 11).
**SUPPLEMENTARY FIGURES**

**Supplementary Figure 1.** Q-Q plot of expected (red line) vs. observed (black dot line) $-\log_{10}$ $p$-values for meta-analysis of genome-wide association studies of gait speed.
Supplementary Figure 2. LocusZoom plots for the genes (7 genes) with most suggestive variants (not listed in the top tens) associated with gait speed of the combined analysis (A) CEP112; (B) PHACTR1, (C) CNTN5, (D) FHOD3, (E) PRIM2, (F) PTPRT, (G) ADAMTS18). In each plot, the −log10 of p values are on the left y-axis; the SNP genomic position (HG19) on the x-axis; the estimated recombination rate from 1000 genomes Nov. 2014 EUR are on the right y-axis and plotted in blue. The most significant SNP is in purple diamond and plotted using the p value attained from the meta-analysis. SNPs are colored to reflect linkage disequilibrium (LD) with the most significant SNP in red (pairwise r2 from 1000 genomes Nov. 2014 EUR). Gene annotations are from the SeattleSeqAnnotation141.
Supplementary Figure 3. Gait speed values distribution within cohorts.
SUPPLEMENTARY TABLES

Please browse links in Full Text version to see Supplementary Tables S1-S11.

SUPPLEMENTARY REFERENCES


Dissecting the genetics of the human transcriptome identifies novel trait-related eQTLs and corroborates the regulatory relevance of non-protein coding loci. Hum Mol Genet. 2015; 24:4746–63. doi: 10.1093/hmg/ddv194


