

OPEN  
ANALYSIS

# Integrative GWAS and co-localisation analysis suggests novel genes associated with age-related multimorbidity

Clare E. West<sup>1,2</sup>✉, Mohd Karim<sup>2,3</sup>, Maria J. Falaguera<sup>2,3</sup>, Leo Speidel<sup>4,5</sup>, Charlotte J. Green<sup>6</sup>, Lisa Logie<sup>6,7</sup>, Jeremy Schwartztruber<sup>3,8</sup>, David Ochoa<sup>2,8</sup>, Janet M. Lord<sup>9,10</sup>, Michael A. J. Ferguson<sup>6</sup>, Chas Bountra<sup>1</sup>, Graeme F. Wilkinson<sup>7</sup>, Beverley Vaughan<sup>1</sup>, Andrew R. Leach<sup>2,3</sup>, Ian Dunham<sup>2,3,8</sup> & Brian D. Marsden<sup>1,11</sup>

Advancing age is the greatest risk factor for developing multiple age-related diseases. Therapeutic approaches targeting the underlying pathways of ageing, rather than individual diseases, may be an effective way to treat and prevent age-related morbidity while reducing the burden of polypharmacy. We harness the Open Targets Genetics Portal to perform a systematic analysis of nearly 1,400 genome-wide association studies (GWAS) mapped to 34 age-related diseases and traits, identifying genetic signals that are shared between two or more of these traits. Using locus-to-gene (L2G) mapping, we identify 995 targets with shared genetic links to age-related diseases and traits, which are enriched in mechanisms of ageing and include known ageing and longevity-related genes. Of these 995 genes, 128 are the target of an approved or investigational drug, 526 have experimental evidence of binding pockets or are predicted to be tractable, and 341 have no existing tractability evidence, representing underexplored genes which may reveal novel biological insights and therapeutic opportunities. We present these candidate targets for exploration and prioritisation in a web application.

## Introduction

Life expectancy is increasing globally, bringing with it the need for new approaches and therapeutics to promote healthy ageing<sup>1</sup>. For instance, the UK Office for National Statistics reported that the proportion of the population aged over 85 is projected to increase from 2.5% in 2020 to 4.3% by 2050<sup>2</sup>. However, the number of years in good health – the “healthspan” – is not increasing at the same pace<sup>3</sup> and people are experiencing an increasing number of years of ill-health in later life. More than half of people over 65 in the UK suffer from multiple long-term health conditions<sup>4</sup> known as multimorbidity, for which it is common for people to be taking five or more medications, known as polypharmacy<sup>5</sup>. The traditional model of tackling diseases independently during research, drug development, clinical trials, and treatment may not be best serving these patients, who represent an increasingly large proportion of the general population. Advancing age is the greatest risk factor for developing multimorbidity, suggesting that biological ageing processes play a pathogenic role. A Geroscience approach, one that aims at tackling combinations of diseases via their common underlying ageing pathways, has

<sup>1</sup>Centre for Medicines Discovery, University of Oxford, Oxford, UK. <sup>2</sup>Open Targets, Wellcome Genome Campus, Hinxton, UK. <sup>3</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK. <sup>4</sup>Francis Crick Institute, London, UK. <sup>5</sup>Genetics Institute, University College London, London, UK. <sup>6</sup>Drug Discovery Unit, University of Dundee, Dundee, UK. <sup>7</sup>Medicines Discovery Catapult, 35 Mereside Alderley Park, Macclesfield, Cheshire, UK. <sup>8</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. <sup>9</sup>MRC-Versus Arthritis Centre for Musculoskeletal Ageing Research, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK. <sup>10</sup>NiHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK. <sup>11</sup>Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK. ✉e-mail: [clarewestc@gmail.com](mailto:clarewestc@gmail.com)

been proposed as a way to treat and prevent age-related morbidity more effectively whilst potentially reducing the burden of polypharmacy<sup>6–8</sup>.

Ageing is associated with time-dependent functional decline, initially manifesting as subclinical physiological changes in the body, progressing to more systemic changes, and ultimately defined by disease conditions and multimorbidity<sup>1</sup>. The rate of decline relative to chronological age varies between individuals, and there is evidence from animal models and humans that this can be modified through interventions<sup>4,7</sup>. Specific biological processes are thought to contribute to ageing and the pathology of age-related diseases, termed the Hallmarks of Ageing<sup>9</sup> and comprise: Telomere shortening, epigenetic modifications, genomic instability, reduced proteostasis, altered nutrient sensing, mitochondrial dysfunction, cell senescence, stem cell exhaustion and altered intercellular communication, with increased systemic inflammation a key component of the latter.

In support of generic mechanisms driving multimorbidity, previous genetic studies have found that age-related diseases are genetically correlated, and that genes with multi-trait associations often relate to biological processes that are implicated in ageing. Belloy *et al.*<sup>10</sup> found significant genetic overlap between 8 age-related diseases and parental lifespan, based on 961 genome-wide significant variant-trait associations in the GWAS catalog. The authors identified 12 multi-trait loci, including variants relating to inflammation, obesity, blood lipid levels, DNA repair mechanisms, and telomere maintenance. For instance, Apolipoprotein E (APOE) is involved in lipid transport and neural health, and is strongly linked to multiple age-related traits including Alzheimer's disease, cardiovascular disease, stroke, and longevity<sup>10</sup>.

Dönertaş *et al.*<sup>11</sup> identified common genetic associations between age-related diseases using a data-driven approach. Using self-reported disease data for UK Biobank (UKBB) participants aged up to 70 years, diseases were clustered based on age-of-onset profiles. One cluster of diseases showed a rapid increase in incidence after age 40, and contained cardiovascular diseases, diabetes, osteoporosis and cataracts. A second cluster shared a slower age-related rate of increase from age 20, including musculoskeletal and gastrointestinal diseases, anaemia, deep vein thrombosis, thyroid problems and depression. Two final clusters contained diseases with childhood-onset (primarily inflammatory diseases) and uniformly distributed age-of-onset (respiratory and infectious diseases). GWAS performed on each disease found a high level of genetic similarity within clusters of diseases with similar age-of-onset profiles, even after controlling for shared disease categories and co-occurrences. Furthermore, the diseases in clusters with age-dependent onset profiles were associated with 581 genes, which were significantly enriched in known ageing-, longevity-, and senescence-related genes.

Pun *et al.*<sup>12</sup> used deep learning models including omics, literature, and key opinion leader scores from multiple data sources to predict a list of 484 genes implicated in a set of 14 age-associated diseases. Well-known age-related genes were found, along with a number of novel genes. A final set of 9 genes was prioritised with strong links to inflammation and epigenetic programming.

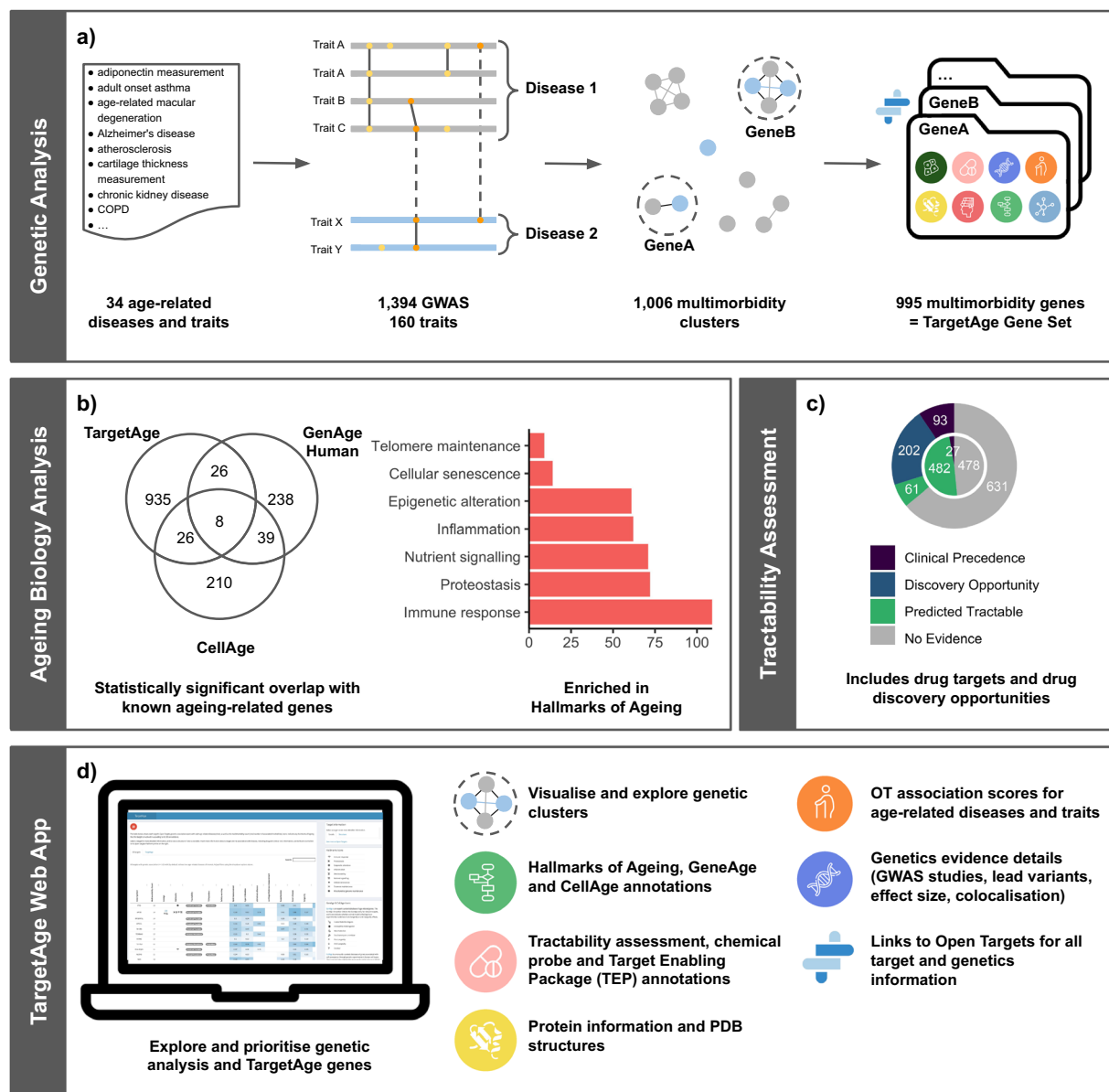
Whilst this biological relationship between ageing and multimorbidity suggests a potential strategy for therapeutic intervention via shared pathways, translating this biological understanding to clinical benefits in patients presents a challenge<sup>6</sup>. Although a number of promising novel drug targets have been explored for the treatment of multimorbidity, as well as repurposing existing drugs such as Metformin<sup>13,14</sup>, there has so far been limited success in clinical trials<sup>7</sup>.

Here we harness publicly available resources for systematic drug target identification and prioritisation for individual diseases, as well as high-quality manually curated databases of genes and drugs related to ageing and longevity, to produce a set of potentially tractable targets for ageing-related comorbidities and multimorbidity (Fig. 1a–d). As human genetic evidence has been shown to improve the odds of successful drug discovery<sup>15</sup>, we focus on a genetics approach to identifying potential drug targets implicated in multiple ageing-related diseases. We use the Open Targets Platform<sup>16</sup> and Open Targets Genetics Portal<sup>17</sup> to perform a systematic analysis of 1,394 genome-wide association studies (GWAS) mapped to a curated list of age-related diseases and traits to identify genetic signals that appear to be shared between two or more of these traits (Fig. 1a). We map these signals to the most likely candidate causal genes using locus-to-gene (L2G), a machine learning-based model, and assess these genes in terms of existing links to ageing and actionability as possible drug targets. These include genes with demonstrated longevity-altering effects in model organisms or links to ageing-related pathways, as well as a large number of novel genes, many of which are predicted to be tractable drug targets (Fig. 1b,c). Finally, we integrate these data in TargetAge, a web application to enable the identification and prioritisation of possible novel drug targets for age-related multimorbidity (<https://targetage.shinyapps.io/TargetAge/>) (Fig. 1d).

## Results

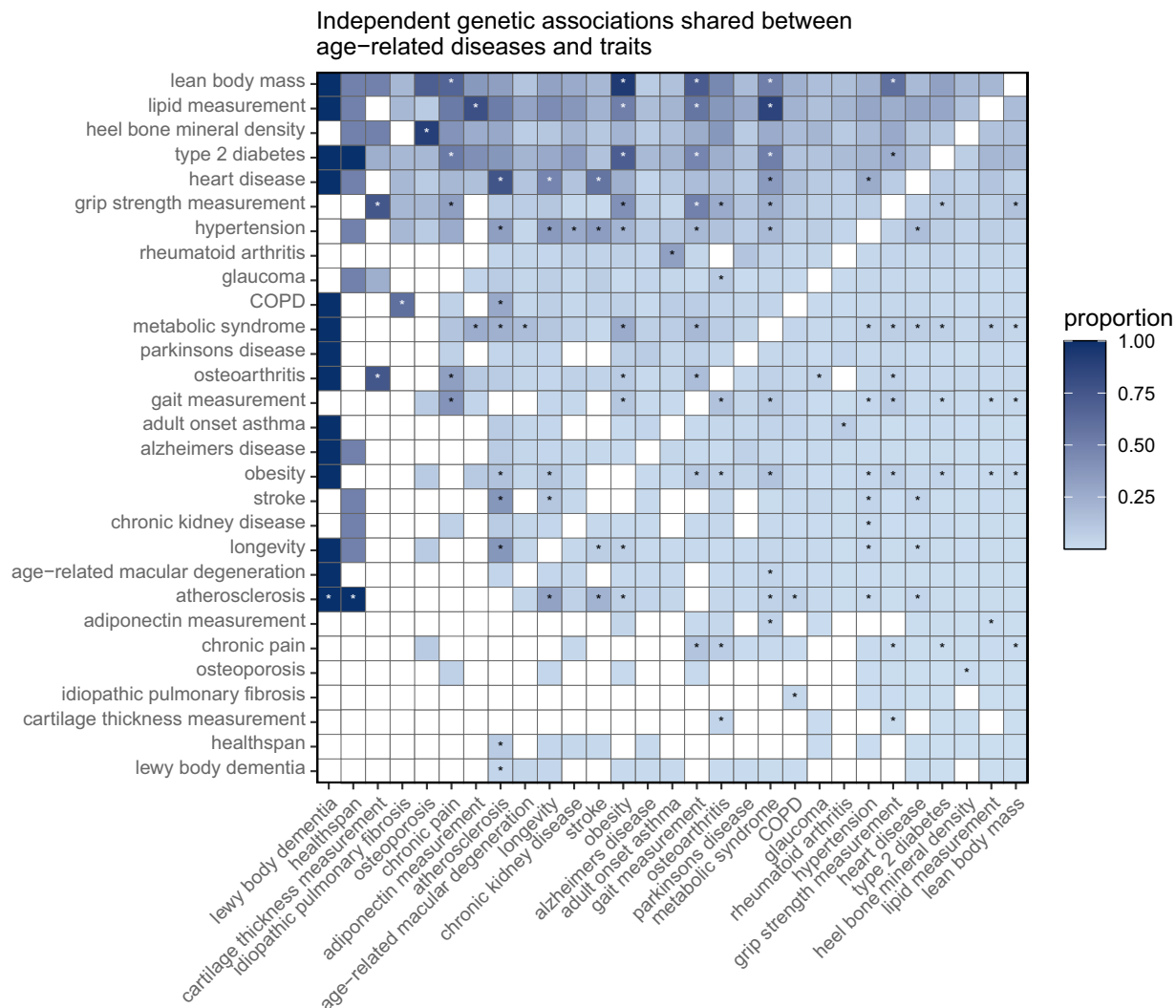
**GWAS studies for age-related diseases and biomarkers.** An overview of the genetic analysis is illustrated in Fig. 1a. We manually curated a list of 34 chronic age-related diseases and biomarkers that correlate with ageing and multimorbidity (such as grip strength measurement and gait speed) based on a review by Melzer *et al.*<sup>18</sup> (SI Table 1). We defined these using experimental factor ontology (EFO) terms<sup>19</sup>, and included relevant descendant terms for these diseases (e.g. in addition to “osteoarthritis” we also included the child term “osteoarthritis, hip”). Terms categorised as genetic, familial or congenital disease (i.e. inherited from parents typically manifesting at, before, or just after birth), as well as pregnancy or perinatal diseases were removed as these are less likely to be ageing-related traits. We additionally excluded cell proliferation disorders; whilst some cancers are prevalent amongst older people, cancers have distinct genetic and environmental disease pathways that are not in the scope of the present study. A full list of the resulting 414 traits and corresponding EFO codes can be found at Figshare (All Traits Details)<sup>20</sup>.

We retrieved all GWAS studies included in Open Targets Genetics (<https://genetics.opentargets.org/>) with a sample size (or number of cases, for case-control studies) of at least 2,000 for which the reported traits mapped to these age-related traits of interest, or their descendants in the ontology. This resulted in 1,394 GWAS studies



**Fig. 1** Overview of the TargetAge analysis. **(a)** Genetic Analysis: GWAS data from Open Targets Genetics is retrieved for 34 age-related diseases and traits and descendant terms in the Experimental Factor Ontology (EFO). Overlapping genetic hits that may share a causal variant are clustered to identify independent genetic signals. Clusters containing multiple age-related diseases are mapped to candidate causal genes using the L2G method from Open Targets Genetics. Target annotations are retrieved from Open Targets Platform for the 995 resulting protein coding genes, referred to as the TargetAge gene set. **(b)** Ageing Biology Analysis: The TargetAge set has statistically significant overlaps with three sets of age-related genes: GenAge, CellAge, and genes annotated with GO terms mapped to the Hallmarks of Ageing. **(c)** Tractability Assessment: The TargetAge set were assessed for clinical precedence and tractability as drug targets using a therapeutic antibody (outer ring) or small molecule (inner) approach, as well as for the availability of chemical probes and Target Enabling Packages (TEPs). **(d)** TargetAge Web App: The TargetAge gene set are presented in a freely available web application including the genetic evidence, target annotations, links to ageing, and tractability assessment.

representing 160 different traits mapping to 30 of the 34 curated age-related traits (SI Table 1). The total number of studies available for each trait ranged from 1 (idiopathic pulmonary fibrosis, chronic pain, and cartilage thickness measurement) to 649 (lipid measurement) with a median of 18.5. Most studies were carried out in populations with European ( $n = 828$ , 59.4%), majority European ( $n = 268$ , 19.2%) or East Asian ( $n = 107$ , 7.7%) ancestry. For these studies, we retrieved the lead variants from each locus with a genome-wide significant association ( $p < 5 \times 10^{-8}$ ).



**Fig. 2** The proportion of independent genetic associations shared between age-related diseases and traits. Proportions are relative to the total number of independent genetic associations identified for the trait on the horizontal axis; darker blue indicates a greater proportion of shared genetic associations. Cells marked with an asterisk are statistically significant after adjusting for multiple testing of all possible pairwise combinations of traits. Traits are ordered by the total number of implicated genetic association clusters. See SI Fig 2 for the number of overlaps between each combination.

**Clustering overlapping GWAS hits identifies independent genetic associations.** To estimate the number of independent genetic signals for the ageing-related traits, we clustered together GWAS hits that may share a causal variant. To define potentially shared signals, we used co-localisation evidence where summary statistics were available (requiring a posterior probability of co-localisation greater than 0.8), otherwise overlap of variants derived from Linkage Disequilibrium (LD) expansion or fine mapping (see Methods).

Around 12% of hits ( $n = 3,524$ ) were only identified in a single GWAS. The remaining associations were grouped into 3,158 clusters of which 2,152 contained associations to one trait and 1,006 clusters involved more than one trait.

This approach aims to cluster together GWAS hits that correspond to a single genetic signal. However, in some cases, the resulting clusters containing multiple associations may still include multiple distinct genetic signals. To investigate whether clusters could be subdivided further, we applied, to each cluster, a community detection algorithm that aims to identify subclusters with weak co-localisation or LD evidence across subclusters and strong evidence within subclusters (see Methods). Most clusters consisted of just one community (77% of all clusters, 49% of multi-trait clusters) or two communities (19% of all clusters, 39% of multi-trait clusters). The largest number of communities within a single cluster was 9 (cluster ID 21) consisting of 299 nodes from 15 traits and was most strongly associated with the gene encoding APOE.

**Shared clusters identify genetic correlations between related traits.** Using these clusters of genetic associations to age-related traits, we tested whether the overlap in genetic signals between pairs of traits was larger than expected by chance given the total number of genetic signals linked to each trait (one-sided Fisher's exact

test, see Methods) (Fig. 2). Overall, 96 of the 406 combinations have a  $p$ -value  $< 0.05$ , of which 55 combinations are significant after adjusting for multiple testing (see Methods). This suggests that the set of ageing-related diseases and traits are enriched with genetically correlated traits.

Most of the significant pairwise overlaps are consistent with previously identified relationships between ageing-related diseases and traits, indicating that the clustering method captures relationships between correlated traits. The most significant overlap is between lipid measurement and metabolic syndrome. Metabolic syndrome is a combination of metabolic abnormalities including altered blood lipid levels, hypertension, obesity, and insulin resistance, which are risk factors for type 2 diabetes and cardiovascular disease<sup>21</sup>. As expected, metabolic syndrome has significant overlaps with all these related traits, as well as age-related macular degeneration, gait measurement, and grip strength measurement. The next most significant overlap is between lean body mass and grip strength measurement, the diagnostic measurements used to define sarcopenia, a common ageing-related condition characterised by loss of skeletal muscle mass, strength and function<sup>22</sup>. This is followed by the highly correlated traits lean body mass and obesity, as well as type 2 diabetes and obesity.

Other overlapping traits reflect known clinical associations, including hypertension with chronic kidney disease and type 2 diabetes<sup>23,24</sup>, and age-related macular degeneration with metabolic syndrome and type 2 diabetes. There is also a significant overlap between loci associated with heel bone mineral density and osteoporosis. Gait measurement and grip strength measurement significantly overlap and are known to negatively correlate with multimorbidity<sup>25,26</sup>. Atherosclerosis is a known life-threatening comorbidity of COPD, and while these diseases are currently treated separately, recent work has proposed identifying and treating shared underlying mechanisms<sup>27</sup>; we find shared genetic associations that could shed light on such mechanisms.

We also find novel correlations between age-related diseases and traits. We find a significant overlap between adult-onset asthma and rheumatoid arthritis; some epidemiological studies have found a positive association between asthma and rheumatoid arthritis, but the relationship is poorly understood<sup>28,29</sup>. We also find a significant overlap between osteoarthritis and glaucoma. While inflammatory forms of arthritis are thought to increase the risk of glaucoma due to intraocular pressure, and a possible common autoimmune-mediated pathological pathway has been proposed for rheumatoid arthritis and glaucoma<sup>30,31</sup>, there is no established link between osteoarthritis and glaucoma.

The single lead variant associated with Lewy body dementia (19\_44908684\_T\_C, rs429358) is part of cluster 21 that is linked to *APOE* and involves 9 age-related traits including age-related macular degeneration, Alzheimer's disease, atherosclerosis, as well as longevity.

Longevity has significant overlaps with healthspan, atherosclerosis, stroke, obesity, osteoarthritis, type 2 diabetes, and lipid measurement. The longevity associations are derived from GWAS on parental lifespan (13 studies), survival to exceptional age (7 studies), and 1 multivariate analysis of ageing traits (healthspan, parental lifespan or longevity). While parental lifespan and longevity have been found to have distinct genetic associations<sup>32</sup>, the combined analysis provides more power to identify associations. Of the 21 clusters that include a longevity trait, most were associated with the combined study of combined ageing (13/21), whereas 4 contained only longevity and 4 contained only parental lifespan.

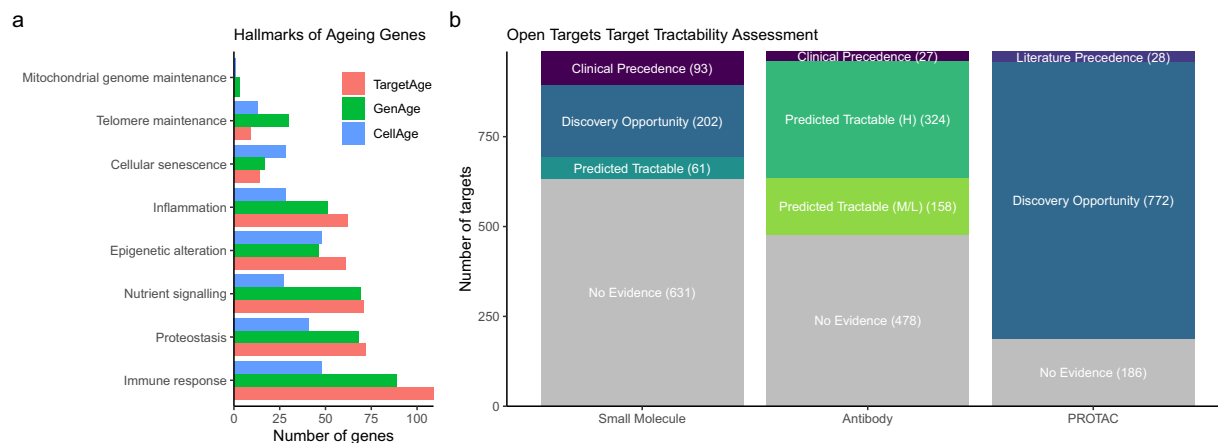
These results are consistent with previous findings on the shared underlying genetics of ageing-related diseases and traits. We hypothesise that genes implicated with multiple ageing-related traits could represent shared aetiology that may relate to underlying ageing processes.

**Shared GWAS associations between traits implicate 995 genes in multiple age-related morbidities.** We mapped each independent genetic association cluster to possible causal genes using the Open Targets Genetics 'locus-to-gene' (L2G) score for each locus in the cluster<sup>33</sup> (Methods). Of the 1,006 clusters involving more than one trait, 796 (79%) contain at least one predicted causal gene (L2G score  $> 0.5$ ) with 178 of these supported by co-localisation with protein quantitative trait loci (pQTL) or gene expression quantitative trait loci (eQTL). Overall, there are 1,021 predicted causal genes linked to at least one multimorbidity cluster, of which 995 are protein-coding, which we refer to as the "TargetAge" gene set. Overall, 290 (29%) of TargetAge genes are annotated to at least one Hallmark of Ageing ( $p = 1.40 \times 10^{-6}$ ).

The majority of TargetAge genes are linked to one ( $n = 905$ , 91%) or two clusters ( $n = 80$ , 8%). There are eight genes (*MPPED2*, *APOB*, *LDLR*, *TBX3*, *ZFH3*, *BCL2L11*, *SMAD3*, and *IRS1*) linked to three clusters, and two genes, *PPARG* and *LRMDA*, are linked to four clusters.

Of the genes linked to more than two clusters, *IRS1* (insulin receptor substrate 1) and *PPARG* (peroxisome proliferator activated receptor gamma) have previously been linked to ageing and are included in GenAge, a manually curated database of ageing-related genes due to longevity-altering effects in model organisms<sup>34</sup>. Knockout of the gene homologous to *IRS1* resulted in increased lifespan in both mice<sup>35</sup> and drosophila<sup>36</sup>. *PPARG* is an important regulator of several ageing-related pathways and has genetic links to type 2 diabetes, atherosclerosis, and longevity; mice with lowered expression of the homologous gene have reduced lifespan<sup>37</sup>. A further four genes have links to the Hallmarks of Ageing, including *TBX3* (T-box transcription factor 3) a regulator of cellular senescence<sup>38</sup>. *SMAD3* (SMAD family member 3) is a regulator of inflammation and an inhibitor of *PPARG* expression<sup>39</sup>; decreased expression of the orthologue *Smad3* has been linked to ageing, neuroinflammation, and neurodegeneration in mouse models<sup>40</sup>. *LDLR* (low density lipoprotein receptor) and *APOB* (apolipoprotein B) are involved in proteostasis and nutrient signalling. *APOB* has been linked to longevity in humans and mice, and may play a role in the pro-longevity effects of dietary restriction<sup>41</sup>.

The remaining four genes that were linked to more than two clusters, but do not have established links to GenAge or the Hallmarks of Ageing, are *MPPED2* (metallophosphoesterase domain containing 2), *ZFH3* (zinc finger homeobox 3), *BCL2L11* (BCL2 like 11), and *LRMDA* (leucine rich melanocyte differentiation associated). Although it does not have a known role in inflammation, a *MPPED2* polymorphism has been associated



**Fig. 3** (a) The proportion of genes annotated with at least one GO term corresponding to each Hallmarks of Ageing category, for the TargetAge gene set (red), GenAge (green) and CellAge (blue). (b) Tractability categorisations when considering Small Molecule, Antibody, or PROTAC therapeutic approaches. Categorisations are based on the existence of an approved or experimental drug (Clinical Precedence), literature reports for PROTAC (Literature Precedence), and different levels of experimental and computational evidence.

with altered systemic inflammation and adverse outcomes in trauma patients<sup>42</sup>. The transcription factor *ZFHX3* is associated with atrial fibrillation, an ageing-related condition, and has a potential role in inflammation<sup>43,44</sup>. *LRMDA* is involved in melanocyte differentiation, and *BCL2L11* is a mediator of apoptosis and is associated with tumorigenesis<sup>45</sup>.

**TargetAge genes overlap with known age-related genes and the hallmarks of ageing.** We compared the TargetAge gene set with three sets of genes with known links to ageing. Two are manually curated databases of human genes linked to ageing or longevity (GenAge Human, 307 genes<sup>34</sup>), and cellular senescence (CellAge, 279 genes<sup>46</sup>). The third is a set of 4,527 genes annotated with Gene Ontology (GO) biological process terms representing each of the Hallmarks of Ageing (see Methods).

The most highly represented hallmark is the immune inflammatory response (11% of genes, Fig. 3a). The TargetAge set of multimorbidity genes also overlaps with both the GenAge Human and CellAge gene sets (Fig. 1b). Of the 995 TargetAge genes, 34 (3.4%) are in GenAge Human (representing 11% of the GenAge Human database) ( $p = 1.4 \times 10^{-5}$ ), and 34 (3.4%) are in CellAge (representing 12% of the CellAge database) (Fig. 1b) ( $p = 2.0 \times 10^{-6}$ ). Whilst these significant overlaps indicate shared biology coverage between TargetAge, GenAge, and CellAge, the sets are mostly distinct, which suggests that there are elements of unique biology that are being captured by each set.

**TargetAge genes include drug targets and drug development opportunities.** TargetAge genes are implicated in multiple age-related diseases and are enriched in ageing processes. They may therefore represent attractive opportunities for therapeutic intervention, provided they are tractable for drug development. The Open Targets tractability assessment categorises genes into levels of potential tractability as drug targets by integrating information such as the clinical precedence, structural characterisation, protein family, and cellular location (Methods).

We used the Open Targets tractability assessment to explore the drug development opportunities within the TargetAge gene set (Fig. 3b). Of the 995 targets in the TargetAge set, 86 have associated drugs with approved indications, and a further 42 have associated drugs currently in clinical trials. There are approved small molecule drugs for 65 targets, therapeutic antibodies for 9 targets, and 22 are targeted by therapeutics of other modalities. One target, oestrogen receptor 1 (*ESR1*), is the target of an investigational proteolysis-targeting chimeras (PROTAC) treatment for Metastatic ER+/HER2- Breast Cancer. Targets with existing drugs may have efficacy and safety information available and may represent potential drug repurposing opportunities.

For the development of a small molecule drug, targets must have a suitable binding site. There is evidence of ligand binding for 202 TargetAge targets (categorised as “Discovery Precedence”), and a further 61 targets are predicted to have a binding pocket or are members of the druggable genome (categorised as “Predicted Tractable”) (see Methods). The druggable genome represents the subset of human genes thought to be amenable to drug development, including those closely related to drug targets, secreted proteins, and members of key drug target protein families<sup>47</sup>.

For a therapeutic antibody approach, a target is categorised as “Predicted Tractable” if its subcellular localisation is accessible to an antibody, such as location in the plasma membrane. There are 324 high confidence “Predicted Tractable” targets that are annotated to accessible locations based on high confidence experimental evidence, while 158 medium confidence targets are either annotated with medium confidence evidence or have predicted Signal Peptide or Trans-membrane regions (see Methods).

Proteolysis-targeting chimeras (PROTACs) are an emerging drug modality in which proteins are targeted for ubiquitin-directed degradation<sup>48</sup>. There are literature reports of successful PROTAC approaches for 28 targets, and a further 772 represent possible PROTAC opportunities based on the PROTACtable Genome assessment<sup>48</sup>, which considers the cellular location, evidence of ubiquitination and small-molecule ligand binding sites, and the availability of protein half-life information (see Methods).

Chemical probes are potent, selective, cell permeable compounds that modulate protein function and thereby facilitate target validation and early stage drug discovery<sup>49,50</sup>. High quality chemical probes are available for 38 TargetAge targets, characterised by the expert-curated Chemical Probes Portal<sup>49</sup>, the precompetitive pharmaceutical collaboration Open Science Probes<sup>51</sup>, or the public-private partnership Structural Genomics Consortium<sup>49</sup>. Of these, 11 targets do not have existing drugs in clinical development, representing potential novel drug discovery opportunities (SI Table 3). Furthermore, around one fifth of the TargetAge genes (n = 138) have at least one potential chemical probe identified by Probe Miner, a public resource that systematically assesses and ranks small molecules for potential use as chemical probes based on public bioactivity data<sup>52</sup>.

Overall, 128 (13%) TargetAge genes are the target of an approved or investigational drug. A further 526 (53%) have experimental evidence of binding pockets or are predicted to be tractable by small molecule or antibody modality approaches. The remaining 341 (34%) have no existing tractability evidence, representing underexplored genes which may reveal novel biological insights and therapeutic opportunities.

## Discussion

We systematically combined 1,394 GWAS of age-related traits to identify clusters where one or more of these traits shared a common genetic signal. To develop the genetic clusters, we made use of genetic co-localisation (when summary statistics were available and credible sets overlapped) and LD-expansion (in the absence of summary statistics) from Open Targets Genetics. We identified 1,006 clusters in which a shared causal variant is implicated in multiple age-related traits. From these clusters, we created the TargetAge gene set of 995 candidate causal genes, of which 29% have a known annotation to at least one Hallmark of Ageing. They are particularly enriched in immune response pathways; dysregulation of these pathways and chronic inflammation, a hallmark of ageing referred to as immunosenescence, is thought to contribute to age-related morbidity. This set has a statistically significant overlap with the manually curated ageing-related gene sets GenAge and CellAge, with evidence of shared biological processes, but 935 of these genes are novel.

The TargetAge genes and genetic analysis outputs are presented in an interactive web application (Fig. 1d). The multimorbidity genetic clusters and candidate genes are displayed alongside evidence of links to ageing and predicted tractability information. Individual comorbidity clusters reveal potentially shared genetic signals between multiple age-related traits and offer further avenues for investigation. Where summary statistics are available, co-localisation evidence indicates where associations have the same direction of effect.

The application facilitates the exploration and prioritisation of potential drug development opportunities for age-related multimorbidity. Some may represent drug repurposing opportunities: 13% of TargetAge genes are targeted by drugs with investigational or approved indications. A further 18% have experimental evidence of binding pockets, and a further 34% are predicted tractable either by small molecule or antibody therapeutics. Of these, 11 targets have high quality chemical probes already available and no existing drugs in clinical development. The remaining 34% have no existing tractability evidence, representing underexplored genes which may reveal novel biological insights and therapeutic opportunities. Together this analysis suggests there is substantial underexplored opportunity for development of drugs for age-related multi-morbidities.

Some clusters identified in this analysis reflect established relationships between commonly occurring comorbidities. For example, a locus associated with atrial fibrillation and heel bone mineral density is linked to thyroid hormone receptor beta (*THRB*) - one of several receptors for thyroid hormones - in cluster 1307. Overstimulation of this receptor by thyroid hormone can cause hyperthyroidism<sup>53</sup>. In adults, hyperthyroidism is associated with increased incidence of atrial fibrillation - with older patients at greater risk<sup>54</sup> - as well as osteoporosis and increased risk of bone fracture, although the underlying mechanism is not fully elucidated<sup>55</sup>. The combination of atrial fibrillation and heel bone mineral density is represented in a further 14 clusters.

We note several clusters of cardiovascular diseases linked to a common therapeutically actionable target. For instance, cluster 209 represents an important locus for cardiovascular diseases, including coronary artery disease, peripheral arterial disease, and large artery stroke, and coronary atherosclerosis. The locus is strongly linked to the gene encoding endothelin receptor type A (*EDNRA*), the target of several approved small molecule drugs for pulmonary artery hypertension and type 2 diabetes indications, suggesting a potential repositioning opportunity for other cardiovascular diseases. Another cluster involving vascular diseases and stroke is linked to phospholipid phosphatase 3 (*PLPP3*). This gene is thought to play a role in vascular homeostasis through negative regulation of pro-inflammatory cytokines, leucocyte adhesion, cell survival and migration in aortic endothelial cells<sup>56</sup>. This locus co-localises with eQTLs in monocytes, which is consistent with a functional role in inflammation and vascular disease, and *PLPP3* is predicted tractable by an antibody therapeutic approach.

In some cases, shared genetic associations may represent distal regulators of the causal gene. Cluster 644 involves a common non-coding variant rs9349379 in the third intron of phosphatase and actin regulatory protein 1 (*PHACTR1*). A study by Gupta *et al.* identified this as the causal variant for associations with five vascular diseases, with a putative mechanism acting through distal regulatory effects on endothelin-1 (*EDN1*), a gene located 600 kb upstream of *PHACTR1*<sup>57</sup>. TargetAge replicates this association: cluster 644 links this variant to coronary artery disease, myocardial infarction, intermediate coronary syndrome, and coronary atherosclerosis. Furthermore, *PHACTR1* is implicated in a second cluster, which is associated with ischemic stroke, large artery stroke, and hand grip strength in addition to coronary artery disease and myocardial infarction.

We also find several clusters with shared therapeutically tractable targets linking measures of adiposity (e.g., body mass index) with type 2 diabetes (T2DM) risk, metabolic syndrome, and osteoarthritis. For

example, *SEC16B* – a protein expressed mostly in the small intestine that is predicted tractable by antibody therapeutics – links obesity with T2DM and metabolic syndrome. *SEC16B* contributes to the intracellular trafficking of vesicles containing lipids and proteins from the endoplasmic reticulum to the Golgi apparatus for further processing and secretion<sup>58</sup>. A recent experimental study showed that intestinal *SEC16B* knockout mice had a significantly lower serum triglycerides after a high-fat meal, improved glucose clearance, and were protected from high fat diet-induced obesity<sup>59</sup>. Our study, along with this experimental evidence, therefore, supports prioritization of intestinal *SEC16B* as a therapeutic target for obesity and impaired glucose tolerance.

Another example of a shared target is Filamin A interacting protein (*FILIP1*), implicated in a locus associated with increased lean body mass and increased risk of osteoarthritis (cluster 220), which is highly expressed in heart and skeletal muscle tissue, and is predicted tractable by a therapeutic antibody. We also identified shared targets that were not predicted tractable via small molecule or antibody approaches, but potentially tractable by proteolysis-targeting chimeras (PROTACs) – an emerging drug modality<sup>48</sup>. An example of the latter is *CMIP* (c-Maf inducing protein) – a negative regulator of T-cell signalling<sup>60</sup>. *CMIP* is linked to cluster 57, associated with lower triglycerides, higher adiponectin (which was inversely associated with T2DM in an observational study<sup>61</sup>), lower risk of T2DM, and lower risk of metabolic syndrome.

Our approach is highly inclusive with the aim of harnessing as much information as possible to uncover potentially novel relationships between ageing-related traits. Some of the proposed multimorbidity clusters are likely to include genetic signals that overlap but have distinct causal mechanisms, particularly in regions of low recombination and high LD, or regions with many genetic associations. Studies with full summary statistics can help discriminate distinct genetic signals that exist within a single cluster by reducing the number of possible causal variants (fine mapping) and removing connections between loci that are unlikely to share a causal variant despite sharing some overlapping tag variants (co-localisation). However, studies with summary statistics represent only around 10% of the GWAS studies available for the investigated ageing-related traits. Increased availability of comprehensive summary statistics would therefore improve our ability to distinguish overlapping but distinct signals. Community detection and visual inspection can help evaluate whether an individual cluster represents a shared genetic signal; the presence of multiple communities within the cluster may suggest that it includes distinct genetic signals. Furthermore, our findings may be biased to European populations due to the bias of the underlying data, although we expect the underlying biological mechanisms implicated to be relevant across human groups. Details of individual GWAS studies contributing to each cluster can be explored in the web application, including the sample size and ancestry of the study and replication populations. Where clusters contain multiple associations for each trait, this indicates that the association has been replicated in multiple studies.

The methodology described here demonstrates one way in which large-scale tools such as Open Targets Platform and Open Targets Genetics can be used to systematically combine evidence from difference sources and use co-localisation data at scale for multiple diseases and traits. Rather than understanding and treating ageing-related diseases seen in multimorbid patients in isolation, there is growing evidence that better health outcomes may be achieved by recognising that diseases often co-occur non-randomly around common genetic, biological, or environmental pathways<sup>8,62</sup>. Such Geroscience, multimorbidity focussed approaches are relevant beyond age-related diseases, and the methodology described here could be applied to other clinically relevant combinations of diseases thought to share common genetic or biological pathways.

We have shown that our approach is able to highlight a range of opportunities both in terms of existing therapeutic agents against linked targets and also novel targets which might allow treatment for multimorbidity. In the present study, we have not explored the causal mechanisms linking the associated SNPs to the increased or decreased risk of each ageing-related outcome. This is an important next step for exploring any potential therapeutic or repurposing opportunities. For the subset of loci that colocalise with pQTLs and eQTLs, the relationship between the gene product and the outcome can be inferred from the direction of effect on the colocalised protein expression or abundance. This indicates whether a therapeutic approach would require activation or inhibition of the gene product, and whether any of the existing drugs would be suitable for repurposing. Individual clusters of interest can be interrogated in this way using the information presented in the application and in the Open Targets Platform and Genetics Portal. It will be important to develop tool molecules for those novel targets identified in order to better understand their biology and to validate their possible role in disease. Ultimately, clinical trials will be required to fully validate the associated proposed hypotheses in an ageing and multimorbidity context.

## Methods

**Genetic analysis.** *Data.* Target and GWAS data were obtained from Open Targets Platform version 22.02 (<https://platform.opentargets.org/downloads>)<sup>16</sup> and Open Targets Genetics version 210608 (<https://genetics-docs.opentargets.org/data-access/data-download>)<sup>17</sup>. From Open Targets Platform, targets were retrieved for which there is GWAS evidence linking the gene to any of the age-related trait EFO codes, or their descendant terms in the Open Targets EFO slim, excluding terms that contain the word “juvenile” or fall under the therapeutic areas “genetic, familial or congenital disease” (OTAR\_0000018), “pregnancy or perinatal disease” (OTAR\_0000014), or “cell proliferation disorder” (MONDO\_0045024). For a full list of EFO codes, and the corresponding number of GWAS and genome-wide significant associations, see Figshare (All Traits Details)<sup>20</sup>. For all the GWAS studies contributing to this evidence, study information and lead variants were retrieved from Open Targets Genetics (Figshare, GWAS IDs)<sup>20</sup>.

*Lead to tag variant expansion.* The lead variant represents just one possible causal variant for each genetic association. These are expanded to a set of possible causal variants, referred to as tag variants. Open Targets Genetics



uses two methods to perform this expansion: Linkage Disequilibrium (LD) expansion and fine-mapping expansion.

**Linkage disequilibrium (LD) expansion.** Where summary statistics are not available, the tag variants include all those in LD with the lead variant ( $r^2 > 0.7$ ). In Open Targets Genetics, LD is calculated for all studies using the 1000 Genomes Phase 3 (1KG) genotypes as a reference. Where ancestry information is known for the GWAS study population, the most closely matching 1KG super-population is used - or a weighted-average across relevant super-populations for mixed ancestry study populations - otherwise European ancestry is assumed.

**Fine mapping expansion.** Details of fine-mapping conducted for Open Targets Genetics are provided elsewhere<sup>33</sup> (<https://genetics-docs.opentargets.org/our-approach/assigning-traits-to-loci>). In brief, where summary statistics were available, Genome-wide Complex Trait Analysis Conditional and Joint Analysis (GCTA-COJO; v1.91.3<sup>63</sup>) was used to identify conditionally independent loci (where both the marginal and conditional p-values were less than  $5 \times 10^{-8}$ ) using genotypes from the UK Biobank population down-sampled to 10 K study participants as the linkage-disequilibrium (LD) reference for the conditional analysis. Approximate Bayes factors were computed using the conditional summary statistics and posterior probabilities (PP) were derived from the Bayes factors for all SNPs within a  $\pm 500$ Kb window assuming a single causal variant. Where summary statistics were not available, we used the Probabilistic Identification of Causal SNPs (PICS) method<sup>64</sup> with LD from 1000 Genomes phase 3 data to estimate the PP for each variant. In both cases, any variant with a PP  $> 0.1\%$  in the credible set is retained.

**GWAS association overlap.** We used two methods to determine whether associations from different GWAS studies may share a causal variant and therefore represent the same signal, depending on the availability of full summary statistics for those studies.

Where summary statistics are available, Open Targets Genetics co-localisation analysis is used<sup>17,65</sup>. For each locus, this method integrates over evidence from all variants within a  $\pm 500$ Kb window in each study to evaluate which of these four hypotheses is most likely: no association with either trait (H0), association only with trait 1 (H1), association only with trait 2 (H2), association with both traits via two independent SNPs (H3), or association with both traits through a shared causal SNP (H4). We use a cut-off of H4  $> 0.8$  to define the association for study 1 and study 2 as sharing the same causal variant.

For most cases, summary statistics are not available for one or both studies. In this case, for lead variants less than 500Kb apart, the associations are considered to represent the same signal if at least one of the tag variants is shared between the two lead variants. Using a more stringent overlap criterion increased the overall number of clusters and decreased the number of clusters involving more than one age-related trait, but does not substantially affect the structure of the graph (SI Fig. 1a). When considering the number of communities detected within each cluster, more stringent criteria reduced the outliers with a high number of communities but did not substantially affect the overall distribution, suggesting the clusters are relatively robust to this criterion (SI Fig. 1b). For inclusivity, we use the least stringent definition of an overlap.

**Clustering genetic associations.** We construct an undirected, unweighted graph in which each node is a trait-associated locus from a single GWAS study, represented by the GWAS study and the lead variant. Nodes are connected by an edge where these genetic associations may share a causal variant, either through co-localisation analysis (where summary statistics are available) or tag variant overlap (see *GWAS association overlap*, above).

Graph construction and analysis was carried out in R (version 4.0.2) using igraph (version 1.2.6) (Csardi and Nepusz 2006) and visnetwork (version 2.0.9)<sup>66</sup> for graph visualisation. Disconnected clusters within the graph represent independent genetic signals, and are identified by calculating the maximal connected components, using the “clusters” function (also known as “components”) with default parameters. To identify substructure within each cluster, which may indicate separate signals, we identify communities using “cluster\_louvain” implementation of the Louvain community detection method with default parameters<sup>67</sup>.

To test the statistical significance of the genetic overlap between each combination of traits, raw p-values were obtained using one-sided Fishers exact test for each pair, implemented using *fisher.exact* in R. P-values adjusted for multiple testing (406 possible combinations of 29 traits) were obtained using BH (also known as FDR) correction implemented in the *p.adjust* function<sup>68</sup>.

**Assigning causal genes to GWAS loci.** Loci were assigned to likely causal genes using the Open Targets Genetics locus-to-gene (L2G) score<sup>33</sup>. L2G is a machine learning method that combines predictive functional genomics features – including predicted pathogenicity, expression quantitative trait loci (eQTL), protein quantitative trait loci (pQTL), genomic distance, and regulatory regions – and outperforms models based on genomic distance alone<sup>33,69,70</sup>. This score ranges from 0 to 1 and is assigned using a supervised machine learning model, which is trained on 445 gold standard GWAS loci with high-confidence causal genes. We only consider genes with an L2G score of at least 0.5 and assign each locus to the gene with the highest L2G score.

**Ageing-related gene sets.** GenAge is a manually curated database of genes thought to be involved in ageing or longevity that includes 307 human genes (GenAge Human) and 2,205 genes from model organisms (GenAge Models)<sup>34</sup>. Human genes are included in the database if there is experimental evidence linking the gene to ageing through gene manipulation experiments in human, mammal, or animal models.

CellAge is a manually-curated database of 279 genes found to induce or inhibit cellular senescence through gene manipulation experiments in human cell models<sup>46</sup>. Cellular senescence is one of the pathways associated with ageing and age-related pathology, and a major target for therapeutic intervention<sup>71</sup>.

The Hallmarks of Ageing are processes that manifest during normal ageing, for which experimental manipulation can accelerate or decelerate the ageing process<sup>9</sup>. This includes causes of cellular damage (genomic instability, telomere attrition, epigenetic alterations, and loss of proteostasis), responses to damage (deregulated nutrient sensing, mitochondrial dysfunction, and cellular senescence), and processes that lead to the ageing trait (stem cell exhaustion, altered intracellular communication). We curated a subset of Gene Ontology (GO) biological process terms that fall under each of the hallmarks of ageing (SI Table 2).

**Tractability assessment.** Target tractability assessment information was retrieved from Open Targets, which uses a pipeline based on Mendez *et al.*<sup>72</sup> and PROTAC tractability workflow from Schneider *et al.*<sup>48</sup>.

Targets are classified as “Clinical Precedence” if there are drugs approved or in clinical development based on curated mechanisms of action from ChEMBL<sup>73</sup>.

For small molecule modality, targets are categorised as “Discovery Precedence” if there is evidence of ligand binding (either high quality compounds with bioactivity data in ChEMBL or co-crystallisation with a ligand) and “Predicted Tractable” if the target is predicted to have a binding pocket (drugEBIity score  $\geq 0.7$ ) and/or is a member of the druggable genome.

For antibody modality, targets categorised as “Predicted Tractable high confidence” are localised to the plasma membrane, extracellular matrix, or secreted, according to high confidence “Subcellular location” terms in UniProt or “Cellular component” terms in GO. “Predicted Tractable Medium to low confidence” targets are annotated to these locations with medium confidence or have predicted Signal Peptide or Trans-membrane regions.

For PROTAC modality, targets with literature reports of successful PROTAC degrader are classified as “Literature Precedence”. Targets are classified as “Discovery Opportunity” if they meet the following criteria: evidence of a ubiquitylation site (UniProt, PhosphoSitePlus, mUbiSiDa, or diglycine proteomics experiments); available experimental data on protein half-life; small-molecule ligand binders ( $\geq 10 \mu\text{M}$  activity reported in ChEMBL), are secreted, extracellular, or located in the cell cytoplasm, cytosol or nucleus (“Subcellular location” terms in UniProt or “Cellular component” terms in GO).

*Curated chemical probes.* Information on available high quality chemical probes is retrieved through Open Targets. Chemical probes released by the Structural Genomics Consortium (<https://www.thesgc.org/chemical-probes>) and through Open Science Probes (<https://www.sgc-ffm.uni-frankfurt.de/>) are open access with no restrictions on use and are required to meet the following criteria: *in vitro* potency of  $< 100 \text{ nm}$ ,  $> 30$ -fold selectivity compared to related proteins within the same target family, demonstrated on-target effects in cells at  $< 1 \mu\text{M}$ , no PAINS elements, and (more recently) an inactive close analogue suitable for use as a control<sup>49</sup>. The Chemical Probes Portal (<https://www.chemicalprobes.org/>) provides expert guidance and ratings for chemical probes; probes are included in Open Targets if they receive a rating of 4 or 3 stars, indicating a recommended or best available probe for a target.

*Predicted chemical probes.* Probe Miner ranks all bioactive compounds for a particular protein target based on six scores: potency, selectivity, cellular activity (as a proxy for cell permeability), structure-activity relationships (SAR), the availability of inactive analogues, and pan-assay interference compound (PAINS) prediction<sup>52</sup>.

## Data availability

Full details of the resulting targets and genetic clusters are provided as four CSV tables and one R data file via Figshare, along with a data dictionary describing the content of each file<sup>20</sup>. Targets are annotated with the strength of the genetic link to each age-related trait (the Open Targets genetic association score), known links to ageing (GenAge, CellAge, and Hallmarks of Ageing), and the tractability assessment. Information on each GWAS study and lead variants associated with each cluster is included, such as the trait studied, sample size, effect size, colocalisation evidence, most likely causal gene and L2G score. All other results of the analysis are available within the paper and Supplementary Materials.

In addition, we developed a web application (<https://targetage.shinyapps.io/TargetAge/>) to facilitate visualisation and exploration of the TargetAge genes identified as genetically implicated in multiple age-related traits, and the underlying multimorbidity clusters and genetic associations.

All data used for this analysis is previously published and publicly available, as described in the Methods. GWAS, L2G and colocalisation data was obtained from Open Targets Genetics version 210608 (<https://genetics-docs.opentargets.org/data-access/data-download>)<sup>17</sup>. The ageing-related diseases and traits, and the GWAS study IDs used in the analysis are provided in Figshare (GWAS IDs)<sup>20</sup>. Target annotation data – including Gene Ontology (GO) annotations, tractability assessments, clinical precedence, and chemical probe and TEP information – were retrieved from Open Targets Platform version 22.02 (<https://platform.opentargets.org/downloads>)<sup>16</sup>.

The ageing-related gene sets CellAge (Build 2)<sup>46</sup> and GenAge (Build 20)<sup>34</sup> were downloaded from the “Human Ageing Genomics Resources” website (<https://genomics.senescence.info/download.html>).

## Code availability

All software used in the analysis have been described in the paper and are freely available. Custom code for the analysis and the web application is available on GitHub at <https://github.com/CMD-Oxford/targetage-pipeline> and <https://github.com/CMD-Oxford/TargetAgeApp>.

Received: 19 May 2023; Accepted: 22 August 2023;

Published online: 25 September 2023

## References

- Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45–56 (2018).
- Office for National Statistics. National population projections: 2020-based interim. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2020basedinterim#toc> (2022).
- Great Britain. Parliament. House of Lords. Science and Technology Committee. Ageing: Science, Technology and Healthy Living : 1st Report of Session 2019–21. (Dandy Booksellers Limited, 2021).
- NIHR Dissemination Centre. Multi-morbidity predicted to increase in the UK over the next 20 years. <https://evidence.nihr.ac.uk/alert/multi-morbidity-predicted-to-increase-in-the-uk-over-the-next-20-years>, <https://doi.org/10.3310/signal-00572> (2018).
- Masnoon, N., Shakib, S., Kalisch-Ellett, L. & Caughey, G. E. What is polypharmacy? A systematic review of definitions. *BMC Geriatr.* **17**, 230 (2017).
- Chaib, S., Tchkonja, T. & Kirkland, J. L. Cellular senescence and senolytics: the path to the clinic. *Nat. Med.* **28**, 1556–1568 (2022).
- Ermogenous, C., Green, C., Jackson, T., Ferguson, M. & Lord, J. M. Treating age-related multimorbidity: the drug discovery challenge. *Drug Discov. Today* **25**, 1403–1415 (2020).
- Whitty, C. J. M. *et al.* Rising to the challenge of multimorbidity. *BMJ* **368**, l6964 (2020).
- López-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Belloy, M. E., Napolioni, V. & Greicius, M. D. A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron* **101**, 820–838 (2019).
- Dönertaş, H. M., Fabian, D. K., Fuentealba, M., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nature Aging* 1–13, <https://doi.org/10.1038/s43587-021-00051-5> (2021).
- Pun, F. W. *et al.* Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. *Aging* **14**, 2475–2506 (2022).
- Barzilai, N., Crandall, J. P., Kritchevsky, S. B. & Espeland, M. A. Metformin as a Tool to Target Aging. *Cell Metab.* **23**, 1060–1065 (2016).
- Dönertaş, H. M., Fuentealba Valenzuela, M., Partridge, L. & Thornton, J. M. Gene expression-based drug repurposing to target aging. *Aging Cell* **17**, e12819 (2018).
- King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).
- Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2020).
- Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
- Melzer, D., Pilling, L. C. & Ferrucci, L. The genetics of human ageing. *Nat. Rev. Genet.* **21**, 88–101 (2020).
- Malone, J. *et al.* Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**, 1112–1118 (2010).
- West, C. E. *et al.* Integrative GWAS and co-localisation analysis suggests novel genes associated with age-related multimorbidity. *figshare*, <https://doi.org/10.6084/m9.figshare.c.6601510> (2023).
- Rochlani, Y., Pothineni, N. V., Kovelamudi, S. & Mehta, J. L. Metabolic syndrome: pathophysiology, management, and modulation by natural compounds. *Ther. Adv. Cardiovasc. Dis.* **11**, 215–225 (2017).
- Cruz-Jentoft, A. J. *et al.* Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* **48**, 16–31 (2019).
- Sun, D. *et al.* Type 2 Diabetes and Hypertension. *Circ. Res.* **124**, 930–937 (2019).
- Tatsumi, Y. & Ohkubo, T. Hypertension with diabetes mellitus: significance from an epidemiological perspective for Japanese. *Hypertens. Res.* **40**, 795–806 (2017).
- Calderón-Larrañaga, A. *et al.* Multimorbidity and functional impairment-bidirectional interplay, synergistic effects and common pathways. *J. Intern. Med.* **285**, 255–271 (2019).
- Bohannon, R. W. Grip Strength: An Indispensable Biomarker For Older Adults. *Clin. Interv. Aging* **14**, 1681–1691 (2019).
- Brassington, K., Selemidis, S., Bozinovski, S. & Vlahos, R. Chronic obstructive pulmonary disease and atherosclerosis: common mechanisms and novel therapeutics. *Clin. Sci. (Lond.)* **136**, 405–423 (2022).
- Rolfes, M. C., Juhn, Y. J., Wi, C.-I. & Sheen, Y. H. Asthma and the risk of rheumatoid arthritis: An insight into the heterogeneity and phenotypes of asthma. *Tuberc. Respir. Dis. (Seoul)* **80**, 113–135 (2017).
- Kim, S. Y., Min, C., Oh, D. J. & Choi, H. G. Increased risk of asthma in patients with rheumatoid arthritis: A longitudinal follow-up study using a national sample cohort. *Sci. Rep.* **9**, 6957 (2019).
- Kim, S. H., Jeong, S. H., Kim, H., Park, E.-C. & Jang, S.-Y. Development of Open-Angle Glaucoma in Adults With Seropositive Rheumatoid Arthritis in Korea. *JAMA Netw Open* **5**, e223345 (2022).
- Geyer, O. & Levo, Y. Glaucoma is an autoimmune disease. *Autoimmun. Rev.* **19**, 102535 (2020).
- Timmers, P. R. H. J., Wilson, J. F., Joshi, P. K. & Deelen, J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat. Commun.* **11**, 3570 (2020).
- Mounjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
- Tacutu, R. *et al.* Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).
- Selman, C. *et al.* Evidence for lifespan extension and delayed age-related biomarkers in insulin receptor substrate 1 null mice. *FASEB J.* **22**, 807–818 (2008).
- Clancy, D. J. *et al.* Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science* **292**, 104–106 (2001).
- Argmann, C. *et al.* Ppargamma2 is a key driver of longevity in the mouse. *PLoS Genet.* **5**, e1000752 (2009).
- Kumar, P. P. *et al.* Coordinated control of senescence by lncRNA and a novel T-box 3 co-repressor complex. *Elife* **3**, (2014).
- Tan, N. S. *et al.* Essential role of Smad3 in the inhibition of inflammation-induced PPARbeta/delta expression. *EMBO J.* **23**, 4211–4221 (2004).
- Liu, Y., Yu, L., Xu, Y., Tang, X. & Wang, X. Substantia nigra Smad3 signaling deficiency: relevance to aging and Parkinson's disease and roles of microglia, proinflammatory factors, and MAPK. *J. Neuroinflammation* **17**, 342 (2020).
- Mutlu, A. S., Duffy, J. & Wang, M. C. Lipid metabolism and lipid signals in aging and longevity. *Dev. Cell* **56**, 1394–1407 (2021).
- Schimunek, L. *et al.* MPPED2 Polymorphism Is Associated With Altered Systemic Inflammation and Adverse Trauma Outcomes. *Front. Genet.* **10**, 1115 (2019).
- Zaw, K. T. T. *et al.* Association of ZFH3 gene variation with atrial fibrillation, cerebral infarction, and lung thromboembolism: An autopsy study. *J. Cardiol.* **70**, 180–184 (2017).
- Tomomori, S. *et al.* Maintenance of low inflammation level by the ZFH3 SNP rs2106261 minor allele contributes to reduced atrial fibrillation recurrence after pulmonary vein isolation. *PLoS One* **13**, e0203281 (2018).

45. Dai, H., Meng, X. W., Ye, K., Jia, J. & Kaufmann, S. H. Chapter 7 - Therapeutics targeting BCL2 family proteins. in *Mechanisms of Cell Death and Opportunities for Therapeutic Development* (ed. Liao, D.) 197–260, <https://doi.org/10.1016/B978-0-12-814208-0.00007-5> (Academic Press, 2022).
46. Avelar, R. A. *et al.* A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* **21**, 91 (2020).
47. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, (2017).
48. Schneider, M. *et al.* The PROTACtable genome. *Nat. Rev. Drug Discov.* **20**, 789–797 (2021).
49. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541 (2015).
50. Gerry, C. J. & Schreiber, S. L. Chemical probes and drug leads from advances in synthetic planning and methodology. *Nat. Rev. Drug Discov.* **17**, 333–352 (2018).
51. Müller, S. *et al.* Donated chemical probes for open science. *Elife* **7**, (2018).
52. Antolin, A. A. *et al.* Objective, Quantitative, Data-Driven Assessment of Chemical Probes. *Cell Chem Biol* **25**, 194–205.e5 (2018).
53. Brent, G. A. Mechanisms of thyroid hormone action. *J. Clin. Invest.* **122**, 3035–3043 (2012).
54. Agner, T., Almdal, T., Thorsteinsson, B. & Agner, E. A reevaluation of atrial fibrillation in thyrotoxicosis. *Dan. Med. Bull.* **31**, 157–159 (1984).
55. Delitala, A. P., Scuteri, A. & Doria, C. Thyroid Hormone Diseases and Osteoporosis. *J. Clin. Med. Res.* **9**, (2020).
56. Touat-Hamici, Z. *et al.* Role of lipid phosphate phosphatase 3 in human aortic endothelial cell function. *Cardiovasc. Res.* **112**, 702–713 (2016).
57. Gupta, R. M. *et al.* A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* **170**, 522–533.e15 (2017).
58. Budnik, A., Heesom, K. J. & Stephens, D. J. Characterization of human SEC16B: indications of specialized, non-redundant functions. *Sci. Rep.* **1**, 77 (2011).
59. Shi, R., Lu, W., Tian, Y. & Wang, B. Intestinal SEC16B modulates obesity by regulating chylomicron metabolism. *Mol. Metab.* **70**, 101693 (2023).
60. Oniszczuk, J. *et al.* CMIP is a negative regulator of T cell signaling. *Cell. Mol. Immunol.* **17**, 1026–1041 (2020).
61. Li, S., Shin, H. J., Ding, E. L. & van Dam, R. M. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA* **302**, 179–188 (2009).
62. MacMahon, S. *et al.* Multimorbidity: a priority for global health research. *The Academy of Medical Sciences: London, UK* **127** (2018).
63. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1–3 (2012).
64. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
65. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
66. Almende B. V., Thieurmel, B. & Robert, T. visNetwork: Network Visualization using 'vis.js' Library. <https://CRAN.R-project.org/package=visNetwork> (2019).
67. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
68. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
69. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
70. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* <https://doi.org/10.1038/s41588-023-01443-6> (2023).
71. Amaya-Montoya, M., Pérez-Londoño, A., Guatibonza-García, V., Vargas-Villanueva, A. & Mendivil, C. O. Cellular Senescence as a Therapeutic Target for Age-Related Diseases: A Review. *Adv. Ther.* **37**, 1407–1424 (2020).
72. Brown, K. K. *et al.* Approaches to target tractability assessment - a practical perspective. *Medchemcomm* **9**, 606–613 (2018).
73. Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).

## Acknowledgements

C.E.W., B.V., J.M.L., G.W., C.J.G. and L.L. are funded by the UK SPINE knowledge exchange network. B.M. is funded by the Kennedy Trust for Rheumatology Research. L.S. acknowledges the support provided by the Sir Henry Wellcome fellowship [220457/Z/20/z]. M.K., M.J.F., J.S., D.O. and I.D. are funded by Open Targets. J.L. is funded by the MRC-Versus Arthritis Centre for Musculoskeletal Ageing Research [MR/K00414X/1]. This research was funded in part by a Wellcome Trust [Grant number 206194]. For the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any author-accepted manuscript version arising from this submission, and associated datasets are published under CC0. We thank the Member States of the European Molecular Biology Laboratory for funding.

## Author contributions

C.E.W. wrote the manuscript with contributions from B.D.M., I.D., A.R.L., M.K., L.S., J.M.L. and J.S., C.E.W. designed and performed the analysis with the help of M.K., L.S., J.S., D.O. and M.J.F. and D.O., and interpreted the results with contributions from all authors. B.D.M., I.D., A.R.L., B.V. and C.B. conceived and supervised the study.

## Competing interests

J.S. is now an employee of Illumina. M.K. is now an employee of Variant Bio. The authors declare no other competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02513-4>.

**Correspondence** and requests for materials should be addressed to C.E.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023