Supporting Information

Methods

Imaging Data

In 2006 through 2010, 503,000 adults (aged 40-70 years) were recruited from the general population in the United Kingdom into a prospective cohort study (Sudlow et al. 2015). The UK Biobank aims to scan 100,000 participants by the end of 2023 with various imaging modalities (Petersen et al. 2013); the current analysis includes 31,494 participants for whom lumbar spine DEXA imaging scans (Lunar iDXA densitometer; GE Healthcare, Chicago, Illinois) were collected for body composition and bone mineral density assessments (April 2014-September 2019). At the imaging assessment visit, information was collected on a range of demographic and lifestyle factors, including ethnicity, education, occupation, alcohol consumption, smoking status, socioeconomic status, and physical activity. Various measurements were also taken, including height, weight, waist and hip circumferences, and blood pressure. Systolic blood pressure and diastolic blood pressure were measured twice after the participant had been at rest for at least 5 minutes in the seated position by using a digital sphygmomanometer (Omron 705 IT; OMRON Healthcare Europe B.V., Hoofddorp, Netherlands) with a suitably sized cuff; the average of the two systolic blood pressure measures was used in all analyses. Further details about the procedural characteristics for the imaging data have been published online[1].

Annotation of Abdominal Aortic Calcification:

The baseline abdominal aortic calcification for 1,000 randomly chosen participants was manually quantified by four annotators using the 24-point scheme described by Kauppila and coworkers (Kauppila et al. 1997). Calcification scores of the anterior and posterior wall of the abdominal aorta over 1st lumbar spine to 4th lumbar spine (L1 to L4) were recorded. Calcification score of each section ranges from 0 to 3. Score 0 represents without any calcification; score 1 represents calcification length less than 1/3 of vertebra; score 2 represents the calcification length spanned from 1/3 to 2/3 of vertebra; score 3 represents calcification length greater than 2/3 of vertebra. AAC score is the sum of calcification score from L1 to L4 with a maximum of 24. AAC score 0 represents no abdominal aortic calcification in lateral lumbar x-ray and AAC score 24 represents a most

[1] http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/DXA_explan_doc.pdf

severe degree of abdominal aortic calcification in lateral lumbar X-ray scan. The median annotation values for 1000 images of these images were used to train and test the machine learning models (read Machine Learning section in Methods). The median annotation scores were stable even on repeat assessments by all annotators (**Figure S2**). The median annotation scores were also assessed for accuracy by comparing associations with biomarkers from the same participants in the MrOS cohort (Cawthon et al. 2016). We also created a test dataset of manually annotated 300 randomly chosen images using the procedure above to assess the accuracy of the machine learning pipelines.

Machine Learning:

We developed an ensemble machine learning approach that combined two pipelines to score aortic calcification levels based only on lumbar spine DEXA scans. Both models developed as part of this work estimate AAC scores via the following 3 steps (details in the methods sections):

1. Segmentation of the lower spine region.
2. Localization of the Aortic region using a spine-curve fitting method.
3. Regression on the localized region to predict the calcification levels from the aortic region.

The DEXA images analyzed in this manuscript were downloaded in three batches. The DEXA images were not all identical in size, but were similar in scale: 444.7(mean) +/- 27.4(SD) pixels wide and 940.0(mean) +/- 8.3(SD) pixels high. The second batch of images that we analyzed had a different distribution of dimensions: 755.1(mean) +/- 65.3(SD) pixels wide and 1665.9(mean) +/- 114.7(SD) pixels high. The third batch of images was similar to the second: 781.5(mean) +/- 43.9(SD) pixels wide and 1654.8(mean) +/- 104.5(SD) pixels high. All images in the second and third batches were rescaled to 55% of their original size along each axis.

Pipeline 1

Detection of the Vertebrae Using Segmentation

Previous studies (Han et al. 2018; Fan et al. 2019; Lessmann et al. 2019) have successfully used a fully convolutional neural network to segment out the vertebrae and determine the spinal curvature. The segmentation model in Pipeline 1 employs a similar approach using a U-Net to achieve semantic segmentation of the Pelvis and the lower spine.

In order to localize the region of the DEXA scan corresponding to the Aorta, correct anatomic locations of the Pelvis (**P**) and 4 vertebrae (**L3-L5**, **S1**) are needed. The network to achieve the necessary segmentation is based on a U-Net architecture. The network architecture consists of a contracting path (left side) and an expansive path (right side) as shown in **Figure S4**. The

contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a batch normalization layer, a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. Each downsampling step doubles the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path (residual connection), and two 3x3 convolutions, each followed by a ReLU. At the final layer a 1x1 convolution is used to map each 16- component feature vector to one of 3 classes - spine, pelvis or background.

The UKBB dataset has a total of 31,494 lumbar spine DEXA scans of patients available to score. Of those, 200 images were randomly chosen for manual annotations of **P** and **L3-L5, S1**. The png images of the scans were loaded on an open sourced annotation tool, QuPath (Bankhead et al. 2017), and the relevant anatomical locations were marked using the polygon tool. The user annotations were converted to binary masks for the 3 classes - Pelvis, vertebrae and background . Of the 200 images, 175 were used for training and cross-validation while the rest were reserved as an unseen test set to quantify the segmentation performance.

Simple data augmentations in the form of crop and zoom, left-right flips and slight rotations (upto +/- 10 degrees) were used to augment the images in the training dataset. The network was trained with the multi-class cross-entropy loss using an Adam optimizer. The evaluation metric is the mean IoU for each of the 3 classes which is defined as (**Figure S5**):

$$IoU = \frac{true-positives}{true-positives+false-negatives+false-positives}$$

Aortic Region Extraction

Once the segmentations of **P** and **L3-L5, S1** are completed, the aortic region is localized and extracted via the following steps:

1. Run the segmentations through post-processing steps to:
    a. Eliminate false positives for **P** (determined through co-localization with **L3-L5, S1**).
    b. Fix broken or missing vertebrae using median height and width estimates for the different vertebrae
2. Determine the centroids of each of **L3, L4, L5, S1** and **P**. Fit a spline curve to pass through each of these points to determine the spinal curvature
3. Along the spinal curvature, move to the right by a fixed offset $\mathbf{A}_{\mathbf{off}}$ and extract a rectangular region whose width is $\mathbf{A}_{\mathbf{width}}$ and whose height is determined by the vertical distance between the centroids of **L3** and **P**. Crop and/or pad the images to be of size 196x196 pixels.

As illustrated in **Figure S6**, at the output of this step, the Aortic regions will be localized and extracted from the original CT images. These regions are then passed on to the Regression module to be converted into a calcification score.

Regression

The final step in translating the DEXA images to a calcification score is to run a regression model that maps the extracted aortic regions to a score. The regression model consists of a backbone feature extractor followed by a fully connected layer and an output layer with a single node that outputs the calcification score. The backbone network is typically borrowed from the classification task on the ImageNet database. This paper explores 3 different backbone feature extractors:

1. InceptionV3 (Szegedy et al. 2015)
2. ResNet50 (He et al. 2015)
3. Custom convolutional neural network consisting of 4 layers of convolution, batch normalization and a ReLu activation layer each.

Of the 1300 user annotated calcification levels, 1000 images were used for training and cross-validation while the other 300 were held out as an unseen test set. As illustrated in **Figure S7**, the ground truth data is highly skewed towards the lower calcification scores creating a high degree of data imbalance. In order to introduce some degree of balance, the dataset was augmented in a stratified manner. Augmentation routines include horizontal and vertical flips, random rotations, zoom and crop, random brightness changes, random contrast changes as well as random hue changes. During agile augmentation, images with higher scores were augmented with different combinations of the routines to produce as many as 32 variations of every image while images with lower calcification scores were augmented only once or not at all.

Since the training set was quite small, both InceptionV3 and ResNet backbones overfit very readily to the dataset and performed quite poorly on the validation sets. The best results on the validation set were obtained when using the custom convolutional network as the backbone feature extractor. All results presented in this paper were, therefore, generated using this custom backbone network shown in the figure.

During training, the regression model used a weighted mean squared error metric wherein errors in the higher calcification scores were weighted higher than those in the lower scores, once again with the intention of offsetting the high degree of skew in the ground truth score distributions. The network is trained with an Adam optimizer and the evaluation metric is the correlation between the predicted scores and the median user scores. The final correlation score between the manually annotated AAC scores and the predicted AAC scores on an unseen validation set is shown in **Figure S7**.

Pipeline 2

The second pipeline we developed in this manuscript seeks to score the degree of abdominal aortic calcification through analysis of a DEXA image. The analysis occurs in three steps:

1. identification of the spine, and use of the labelled vertebrae to identify & extract the small region within the image that will contain the aorta;
2. classification of the aortic sub-image using three models (one classifying calcified-vs-not using a low threshold for calcification, another classifying the same using a high threshold for calcification, and a third classifying the extent of background noise in the image into low, medium, or high categories);
3. use of the scores assigned to each category by the above classification models to assign a calcification score using an ML regression model.

Detection of the Vertebrae Using Segmentation

The clinical scoring system for AAC involves the assignment of points based on the intensity with which the walls of the abdominal aorta can be visualized in an x-ray image adjacent to the L1-L4 lumbar vertebrae. Here, we attempted to isolate the relevant portion of the DEXA image for such evaluation by first mapping the spine and then capturing a smaller image adjacent to the L3 & L4 vertebrae. For our study, DEXA images were available in bulk. But they are not ideally suited to the detection of AAC: high levels of background intensity would often appear adjacent to the L1 & L2 vertebrae, often due to visualization of the rib cage. We therefore sought to isolate just the part of the image adjacent to vertebrae L3 & L4. That strategy is illustrated by the examples in Figure S8, with the lumbar vertebrae labelled in green and the desired aortic region circled in orange. This goal was achieved through two sequential processes as described below:

The foundation of this AAC scoring system is the annotation of the spine and its constituent vertebrae in the DEXA image. That annotation defines the landmarks that are used to isolate the portion of the image that contains the aorta, a task that cannot be performed independently of these landmarks because the aorta is only clearly visible in these images when it is extensively calcified.

This section describes the steps taken to annotate as many individual vertebrae as possible while minimizing the false-positive annotation of other parts of the image as vertebrae. This task was initially approached with a simple object-detection model. In order to correct mistakes made during the initial segmentation step, we chose to implement a series of analyses & models to detect & correct errors. The end product is an analysis pipeline that is summarized in **Figure S8**.

Initially, vertebrae are boxed using an object-detection model that is applied with a very low score threshold ("draw vertebra boxes"; few false negatives, many false positives). Incorrect annotations that are far from the spine are easily detected and removed using the unusual

vertebra-to-vertebra angles that are produced ("trim off-axis vertebrae"). Each remaining vertebra is then scrutinized by a classification model that is trained using high-quality versus low-quality annotations that were produced by the initial object-detection model, i.e. a model to specifically address the first model's errors ("remove low-quality vertebrae"). Then, missing vertebrae (either missed initially or removed because they were poorly defined) are detected and filled in, using the context of the nearby annotations to increase sensitivity and specificity versus the initial object-detection model. In the first case, skipped vertebrae are detected ("classify vertebra pairs") and filled in by a model appropriate to the number of consecutive vertebrae that were missed ("fill-in one" or "fill-in multi"). In the second case, missing vertebrae at the bottom of the spine are detected ("have we hit bottom?") and filled in ("extend down by one"). The performance of each substep is summarized in **Figure S9**.

Aortic Region Extraction

After identifying all the vertebrae in the spine, the region of the scan adjacent to the L3 and L4 lumbar vertebrae was extracted: this was the region of the scan with the abdominal aorta. Given an outline of the spine, it is straightforward to determine which vertebra is which given a single reference point. In these DEXA images, the border between the L4 and L5 vertebrae was often coincident with the top edge of the spine. An object-detection model was trained by transfer learning from the ssd_mobilnet_v1 model**.** The best box returned by the model was used, irrespective of its score. Performance was measured using IoU (0.893 and 0.818 for training and test sets respectively).

The aortic images were created by stacking rectangular images gathered from adjacent to the L3 and L4 vertebrae, using the following four-step process:

a. definition of a vector pointing towards the aorta for each vertebra;
b. definition offset distances and image widths using the dimensions of the lumbar bounding boxes and aortic vectors;
c. definition of sub-image heights & extraction of the aortic sub-images adjacent to L3 and L4; and
d. rightening & stacking the L3- and L4-adjacent sub-images into a single output image.

For each input image (full DEXA scans), the end result of the analyses described in the links above was a smaller image depicting the regions adjacent to the L3 vertebra (above the white line) and the L4 vertebra (below the white line). The inability to identify both of those vertebrae resulted in no output aortic image.

Regression

The overall distribution of AAC values is highly skewed towards little or no aortic calcification. This property was observable in the training set, with most rater-generated scores at zero (Figure 1). We developed two models, each described below, in order to address the problem of sparsity of training data for high-calcification scores. The first model focuses on the lower end of the distribution, and was trained to distinguish between images with zero-value versus non-zero

AAC scores. The second focuses on the higher end of the score distribution, where I used a more-efficient but less-precise-than-scoring method to enrich a larger test data set for high-calcification images.

a) **Model 1:** For the training data, four raters scored calcification as an integer, and the median value was taken of those four scores. For this low-threshold AAC model, any image with a rater-median score of 0.5 or greater was considered "calcified". That approximately split the training set (264 "calcified" images, 332 "non-calcified" images). The ROC curves to measure the accuracy of this model is shown in **Figure S10** while Cohen's kappa for these models were 0.58 and 0.33 for the training and test sets respectively.
b) **Model 2:** For the purpose of helping to develop higher-threshold calcification models, we designated a set of 5000 "sandbox" images that were non-overlapping with the validation and training sets, and were therefore of potential use to the model developers as training data, but for which they would not provide manual ratings. We used those data by iteratively applying a low-threshold model to those images, sorting out the "calcified" images, then manually enriching those images for yet-higher calcification values by selecting the apparently-more-calcified images from image pairs until we had sufficient data to train another model. We repeated this process until we had arrived at a training set with 170 "calcified" and 4654 "non-calcified" images, at a threshold that we estimated to be at approximately score=5.
   For evaluation, the original training and test sets suffered from sparsity of high-scoring data, making our evaluation of their performance sensitive to statistical noise. The ROC curves and Cohen's kappa values for a) the actual, "sandbox"-enriched training set (yellow - 0.71); b) the original training set, with a score threshold of 5 (green - 0.73), and c) the original test set, again with a score threshold of 5 (blue - 0.53) (**Figure S10**).

For the final output value of the model (AAC score), we built and trained a small regression model to input the probability scores from the classification models above and output an AAC estimate. For the binary classification tasks, we used just one of the two outputs. The structure of the model is shown in **Figure S11**. This model has 52 trainable parameters. We experimented with many model structures and multiple attempts at training the model described above, evaluating performance using the "test" set. The statistics for the final model are shown for an unseen validation set in **Figure S12**.

Ensemble Model

Due to the paucity of labeled data and to avoid overfitting by using the same labeled data for training both pipelines and the ensemble model, we used an unsupervised ensemble method to combine the scores from both pipelines. In particular, we took the mean calcification levels predicted by both pipelines as the ensemble prediction. We tested the accuracy of the ensemble method on 300 test images that were not used for training either model and the accuracy of the ensemble model was higher than the accuracy of either pipeline (Table 2). Hence, we used the

ensemble method to quantify calcification for all participants within the cohort and these scores were used for all downstream analysis in this article.

Associations with Biomarkers/Physiological Markers

We used linear regression to examine the association of aortic calcification with various biomarkers and physiological markers. For these models, univariate associations of aortic calcification with each marker were evaluated with two different models. In model 1, all univariate associations were performed after adjusting for age and sex, while in model 2, we also adjusted for BMI, Townsend deprivation index, and race in addition to age and sex. We chose these factors based on their associations with traditional cardiovascular outcomes and/or their association with aortic calcification. All p-values were calculated using the two-tailed t-statistic of the estimated association. The associations were considered to be significant after multiple hypothesis testing (i.e., Bonferroni Correction with p-value < 1.2e-4). These estimations were calculated using the statsmodels package v.0.9.0 in python (Seabold and Perktold 2010).

Common variant genome wide association study

We used the UKBB imputed genotypes (Bycroft et al. 2018), excluding SNPs with a minor allele frequency <1% and poor imputation quality (info value <0.9). We removed particpants who were not Caucasian, exhibited sex chromosome aneuploidy, heterozygosity outliers, or genotype call rate outliers. In total, we considered 9,572,557 SNPs and 19,051 individuals (Supplementary Table 1) for genetic analysis.

To conduct the genetic association study, we used BOLT-LMM (Loh et al. 2015) and standardized machine-learned AAC, including genotype SNP chip (Illumina vs Affimetrix), sex, age, age^2, and recruitment center as fixed effect covariates and genetic relatedness derived from genotyped SNPs as a random effect to control for population structure and relatedness. We verified that the test statistics showed no inflation compared to the expectation using the genomic control lambda coefficient (1.048) and the intercept (1.0061, s.d. 0.0087) of linkage disequilibrium (LD) score regression (LDSC) (B. K. Bulik-Sullivan et al. 2015).

At the locus on chromosome 2, in order to ensure this association was not spurious due to poor genotype calls, we manually verified the raw genotype clusters for the tag variant, rs7600800. We did not find evidence of poor quality genotype calls at this locus.

Genetic architecture of AAC

Identification of distinct association signals

We performed approximate conditional analysis using GCTA (Yang et al. 2012), considering all variants that passed quality control measures and were within 500kb of the locus index variant. As a reference panel for LD calculations, we used genotypes from 5,000 UKBB participants (Bycroft et al. 2018) that were randomly selected after filtering for unrelated, Caucasian participants. We excluded the major histocompatibility complex (MHC) region due to the complexity of LD structure at this locus (GRCh37::6:28,477,797-33,448,354; see https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC). For each locus, we considered variants with locus-wide evidence of association ($p\text{-value}_{joint}<10^{-6}$) to be conditionally independent.

Construction of genetic credible sets

For each distinct signal, we calculated credible sets (Wellcome Trust Case Control Consortium et al. 2012) with 95% probability of containing at least one variant with a true effect size not equal to zero. We first computed the natural log approximate Bayes factor (Wakefield 2007), Λ_j , for the j th variant within the fine-mapping region:

$$\Lambda_j = ln\left(\sqrt{\frac{V_j}{V_{j+\omega}}}\right)\frac{\omega\beta^2}{2V_j(V_j+\omega)}$$

where β_j and V_j denote the estimated allelic effect (log odds ratio for case control studies) and corresponding variance. The parameter ω denotes the prior variance in allelic effects and is set to $(0.2)^2$ for case control studies (Wakefield 2007) and $(0.15\sigma)^2$ for quantitative traits (Giambartolomei et al. 2014), where σ is the standard deviation of the phenotype estimated using the variance of coefficients ($Var(\beta_j)$), minor allele frequency (f_j), and sample size (n_j; see the sdY.est function from the coloc R package (Giambartolomei et al. 2014)):

$$2n_j f_j(1-f_j) \sim \sigma^2 \frac{1}{Var(\beta_j)} - 1$$

Here, σ^2 is the coefficient of the regression, estimating σ such that $\sigma = \sqrt{\sigma^2}$.

We calculated the posterior probability, π_j, that the j th variant is driving the association, given l variants in the region, by:

$$\pi_j = \frac{(1-\gamma)\Lambda_j}{\sum_{k=0}^{l}\Lambda_k}$$

where γ denotes the prior probability for no association at this locus and k indexes the variants in the region (with k=0 allowing for the possibility of no association in the region). We set γ=0.05

to control for the expected false discovery rate of 5%, since we used a threshold of $p\text{-value}_{marginal}<5x10^{-8}$ to identify loci for fine-mapping. We note that setting γ=0 generates credible sets as proposed by The Wellcome Trust Case Control Consortium et al. (Wellcome Trust Case Control Consortium et al. 2012) and is suitable when one is very confident of the identified loci (e.g., replicated across many studies).

To construct the credible set, we (i) sorted variants by increasing Bayes factors (natural log scale), (ii) included variants until the cumulative sum of the posterior probabilities was >=1−c, where c corresponds to the credible set cutoff of 0.95.

AAC heritability estimates

We estimated the heritability of each trait using the restricted maximum likelihood method (Yang et al. 2010), as implemented in BOLT-LMM with the --reml option.

Genetic correlation of AAC with other phenotypes

We estimated the genetic correlation of AAC with phenotypes using an LDSC-based method (B. Bulik-Sullivan et al. 2015), as implemented in the LD Hub web resource (Zheng et al. 2017). For this analysis and all other analyses using LDSC, we followed the recommendation of the developers and (i) removed variants with imputation quality (info) <0.9 because the info value is correlated with the LD score and could introduce bias, (ii) excluded the major histocompatibility complex (MHC) region due to the complexity of LD structure at this locus (GRCh37::6:28,477,797-33,448,354; see https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC), and (ii) restricted to HapMap3 SNPs (International HapMap 3 Consortium et al. 2010). We restricted our analysis to the 836 traits analysed in European populations. Of these, 754 successfully gave an estimate of genetic correlation.

Partitioning of ACC heritability

We used LDSC to partition the heritability of AAC according to functional categories (Finucane et al. 2015) as well as tissue/cell type specific annotations (Finucane et al. 2018).

For functional categories, we used the baseline v2.2 annotations provided by the developers (https://data.broadinstitute.org/alkesgroup/LDSCORE). Following Finucane et al. (Finucane et al. 2018), we calculated tissue specific enrichments using a model that includes the full baseline annotations as well as annotations derived from (i) chromatin information from the NIH Roadmap Epigenomics (Roadmap Epigenomics Consortium et al. 2015) and ENCODE (ENCODE Project Consortium 2012) projects (including the EN-TEx data subset of ENCODE which matches many of the GTEx tissues, but from different donors), (ii) tissue/cell type specific expression markers from GTEx v6p (GTEx Consortium et al. 2017) and other datasets (Fehrmann et al. 2015; Pers et al. 2015), and (iii) immune cell type expression markers from the ImmGen Consortium (Heng, Painter, and Immunological Genome Project Consortium 2008).

For each annotation set, we controlled for the number of tests using the Storey and Tibshirani procedure (Storey and Tibshirani 2003). As noted by (Finucane et al. 2015), although heritability is non-negative, the unbiased LDSC heritability estimate is unbounded; thus, it is possible for the estimated heritability, and therefore enrichment, to be negative (e.g., if the true heritability is near zero and/or the sampling error is large due to small sample sizes).

In order to enable visualization, we grouped tissue/cell types into systems (e.g., "blood or immune", "central nervous system"). These groupings and labels were the same as those used in Finucane et al. (Finucane et al. 2018), except for (i) the immune expression labels which we extended to include "stem cells" and "stromal cells" according to the ImmGen cell type classifications (http://www.immgen.org) and (ii) the "pancreas" label which we replaced with an "endocrine" label composed of the following tissue/cell types: adrenal gland, ovary, pancreas, pituitary, prostate, testis, thyroid, adrenal cortex, adrenal gland, endocrine gland, gonads, granulosa cells, islets of langerhans, and glucagon sensing cells.

Genetic colocalization of AAC with other phenotypes

We performed colocalization analysis using the coloc R package (Giambartolomei et al. 2014) using default priors and all variants within 500kb of the index variant. As performed by Guo et al. (Guo et al. 2015), we considered two genetic signals to have strong evidence of colocalization if PP3+PP4≥0.99 and PP4/PP3≥5 and suggestive evidence of colocalization if PP3+PP4≥0.9 and PP4/PP3≥3. For gene expression colocalizations, we used summary statistics from GTEx v7 (GTEx Consortium et al. 2017). For disease and quantitative trait colocalizations, we used UKBB summary statistics of PheCodes (Zhou et al. 2018), normalized quantitative traits (http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank). For analysis we selected UKBB phenotypes where the minimum p-value within the +-500kb region around the locus tag SNP was $<5x10^{-8}$.

Follow up analysis at the rs2107595 locus

To assess the correlation between *TWIST1* and *HDAC9* expression and calcification visible on histological imaging, we downloaded the sample annotations from https://storage.googleapis.com/gtex_analysis_v7/annotations/GTEx_v7_Annotations_SampleAttributesDS.txt. We defined a participant's vascular tissue as calcified if the annotation 'calcification' appeared in the 'Pathology categories' field. We downloaded expression data from GTEx v7 (dbGaP Accession phs000424.v7.p2). For each arterial tissue (coronary, aorta, or tibial) and gene (*TWIST1* or *HDAC9*), we used a logistic regression model, adjusted for age and sex, to assess the association between expression levels and calcificaiton. We used the Bonferroni procedure to correct for multiple models by multiplying each p-value by 6.
We explored *Twist1* expression patterns in single cell expression data from mouse aorta (Kalluri et al. 2019) (Figure S17).

In addition to the colocalizations described in the main text, we tested for colocalization with lipid-related quantitative traits directly measured in UKBB, even though no genetic association (p-value<$5x10^{-8}$) was found at this locus, due to the importance of lipids for coronary artery disease (CAD)-related phenotypes. We found no evidence of colocalization (Supplementary Tables 4, 5, 6, 7), suggesting that the genetic effect at the rs2107595 locus on AAC and CAD-related phenotypes is not directly related to lipid biology. We also repeated the colocalization analysis for CAD and blood pressure traits using genetic studies that did not include UKBB participants (Nikpay et al. 2015), (Hoffmann et al. 2017). We found similarly strong evidence of colocalization for both traits (Supplementary Table 7).

Finally, given the strong colocalization of AAC with CAD and SBP signals as well as the substantially larger sample sizes of these traits ($n_{effective}$ for CAD=76,054 and SBP=340,159), we performed fine-mapping at this locus using associations with CAD and SBP in UKBB. Two SNPs, rs2107595 and rs57301765, that are in strong LD with each other (1000GENOMES:phase_3:GBR r^2>0.99) constituted the 95% credible sets for both traits.

Prognostic Analyses of Aortic Calcification:

Association of different diseases with AAC

We used Cox proportional hazards models to examine the association of aortic calcification with various diseases. The 10th edition of the international classification of diseases (ICD10) diagnosis codes and date of initial diagnoses were extracted from electronic health records of all participants. The ICD10 codes are hierarchically organized and were created for insurance billing purposes. ThePheWAS codes attempt to group the different ICD10 codes into medically meaningful groups (Denny et al. 2010). We converted the ICD10 codes at level 2 hierarchy into PheWAS codes before associating with the predicted aortic calcification. Any participant diagnosed with the PheWAS code prior to baseline was removed from analysis before calculating associations with aortic calcification. For these models, univariate associations of aortic calcification with each marker were evaluated with two different models. In model 1, all univariate associations were performed after adjusting for age and sex, while in model 2, we also adjusted for BMI, Townsend deprivation index, and race in addition to age and sex. We chose these factors based on their associations with traditional cardiovascular outcomes and/or their association with aortic calcification. All p-values were calculated using the two-tailed t-statistic of the estimated association. These estimations were calculated using the statsmodels package v.0.9.0 in python (Seabold and Perktold 2010).

Comparison of risk from Aortic Calcification and LDL:

We used Cox proportional hazards models (Cox 2018) to compare the risk for acute myocardial infarction from aortic calcification and LDL with the following model:

$$\lambda(t) = \lambda_0(t)\, e^{\alpha}$$

where:

$$\alpha = \alpha_0 + \alpha_{LDL} \times LDL + \alpha_{triglyceride} \times triglyceride + \alpha_{AAC} \times AAC + \alpha_{LDL.AAC} \times LDL \times AAC$$

where α_0 represents the baseline risk, α_{LDL} represents the risk due to increase of LDL, α_{AAC} represents the risk due to increase in aortic calcification, $\alpha_{triglyceride}$ represents the risk due to increase in triglycerides, and $\alpha_{LDL.AAC}$ was used to measure the risk due to the interaction between LDL and triglycerides. The LDL and triglyceride levels were logged and standardized while aortic calcification levels were standardized before measuring the risk for acute MI events.

Acute MI events were defined as the first occurence of I21, I22, I23, I24.1 or I25.2 from the ICD10 codes in the electronic health records similar to (Millett, Peters, and Woodward 2018). Any participant diagnosed with acute MI events prior to baseline was removed from analysis before comparing risks from aortic calcification and LDL. We created four different models to correct the risk for statin usage that reduces LDL without reducing systolic blood pressure completely (Figure S22 and S23) in addition to a naive model in which no statin correction was performed. The first model was a naive model built for statin nonusers and the estimates for LDL risk are confounded by survivorship bias due to the nonrandomness of statin usage. In model 2, we adjusted the LDL and triglyceride levels for statin users by adding 1.25 mmol/L to the LDL levels for statin users while no adjustment was performed for non-statin users (Nissen et al. 2005) (Figure S22). In model 3, we adjusted the LDL and triglyceride levels for statin users by dividing the measured LDL for statin users by 0.65 while no adjustment was performed for non-statin users (Nissen et al. 2005) (Figure S22). Finally, in model 4, we imputed the LDL levels for statin users based on systolic blood pressure, diastolic blood pressure, pulse, age, sex, and measured blood biomarker (albumin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, bilirubin, urea, calcium, creatinine, cystatin C, gamma glutamyltransferase, glucose, glycated haemoglobin (HbA1c), insulin growth factor, phosphate, rheumatoid factor. Testosterone, sex hormone binding globin, total protein, urate, and vitamin D) levels in serum. Imputation was performed using multiple imputation by chained equations (MICE) (Azur et al. 2011). In the MICE procedure a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data. To evaluate the sensitivity of the results to the imputed values, we performed the risk of LDL and aortic calcification with ten different sets of LDL and triglyceride imputations and each MICE imputation was performed with 50 iterations to relax from initial estimates. The estimated risk scores were not sensitive to the imputation set as shown in Figure S24.

Supporting Information Text

Manual Annotation of Abdominal Aortic Calcification

Four annotators quantified the abdominal aortic calcification by assessing the digitized baseline lateral DEXA scans of lumbar spine using a visual semiquantitative method (Kauppila et al. 1997) for 1300 randomly chosen participants. In this method, the severity of calcific deposits in the anterior and posterior walls of the abdominal aorta adjacent to each of the first four lumbar vertebrae (L1-L4) were assessed individually, using the midpoint of the intervertebral space above and below the vertebrae as boundaries. Severity scores for each of these eight segments (0–3) were added to yield an AAC score (0-24).

We utilized 1000 images for training and testing both machine learning pipelines while 300 images were kept aside for estimating accuracy on a validation dataset. The inter-annotator variability was measured over the training data initially (**Table S1**). In general, the inter-annotator variability was lowest for highly calcified and low/noncalcified individuals while annotations were most variable for participants with intermediate levels of calcification (**Figure S1**). To assess the effect of intra-annotator variability of the median annotation scores, we chose 136 images with the highest inter-annotator variability within the training dataset (i.e., the images on which there was the largest disagreement within different annotators) and all four annotators re-annotated these images. While the inter-annotator variability remained within patients with high calcification scores, the reannotated median calcification score showed a Pearson correlation of 0.93 with the original calcification scores indicating that the median calcification score was a pretty reliable indicator of calcification within the cohort (**Figure S2**). The calcification scores for the 300 images in the validation dataset were also highly correlated across annotators as well as with the median calcification score (**Tables S2**). The manually annotated median calcification scores also have similar trends of correlations with biomarkers in the MrOS cohort with expert annotated calcification scores (Szulc et al. 2014) (**Table S3**).

Automation of Estimating Abdominal Aortic Calcification

The UKBB has collected whole body DEXA scans for 31,494 of the participants to estimate their bone mineral density. As the previous sections explain, the calcification of the Aorta is visible in these scans and can be quantified based on the scoring methodology illustrated in (de Bie et al. 2017) (**Figure S8**). Coronary arteries are visible in DEXA scans only when calcified.

The first approach this paper took towards scoring these images was to train a deep regression model that mapped the complete lumbar spine CT images on to user annotated AAC scores. The model used a ResNet50 architecture (He et al. 2015) as backbone and had a fully connected layer followed by a linear output layer to score the calcification levels in the images. However, due to the highly variable background noise in the images coupled with the high

degree of noise in the user annotated scores, the model performed quite poorly as can be seen in **Figure S3** below even on a limited range of scores.

The automatic scoring methodology adopted in both models of this paper, therefore, relies on a segmentation methodology of large structures that would always be visible in these scans such as the Pelvis and the Spine in order to first localize the region where the Aorta can be seen. Once localized, any calcification that is seen in the Aortic region is quantified and mapped to a score via a regression model. Both pipelines developed as part of this work estimates AAC via 3 steps (details in methods).

The main goal of the computer vision and machine learning exercise was to find a way to automate and scale the scoring of aortic calcifications to the entire UKBB dataset in the absence of a large volume of ground truth user annotations. The segmentation, region extraction and regression architectures were all chosen with this constraint in mind. Since the annotated dataset was quite small, we explored 2 separate machine learning models to automate the scoring of these images. Eventually, the scores from these 2 models were averaged to create an ensemble score. As expected, the ensemble score had a higher correlation score to the median annotated scores than either of the 2 models separately.

With this ensemble method, we have managed to score a total of 29,957 DEXA images with only 200 segmentation annotations and 1300 calcification score annotations. Given that the UKBB imaging modality is highly consistent, the models given out in the github page can very well be used to infer the scores on future DEXA scans as they become available.

SI Figures

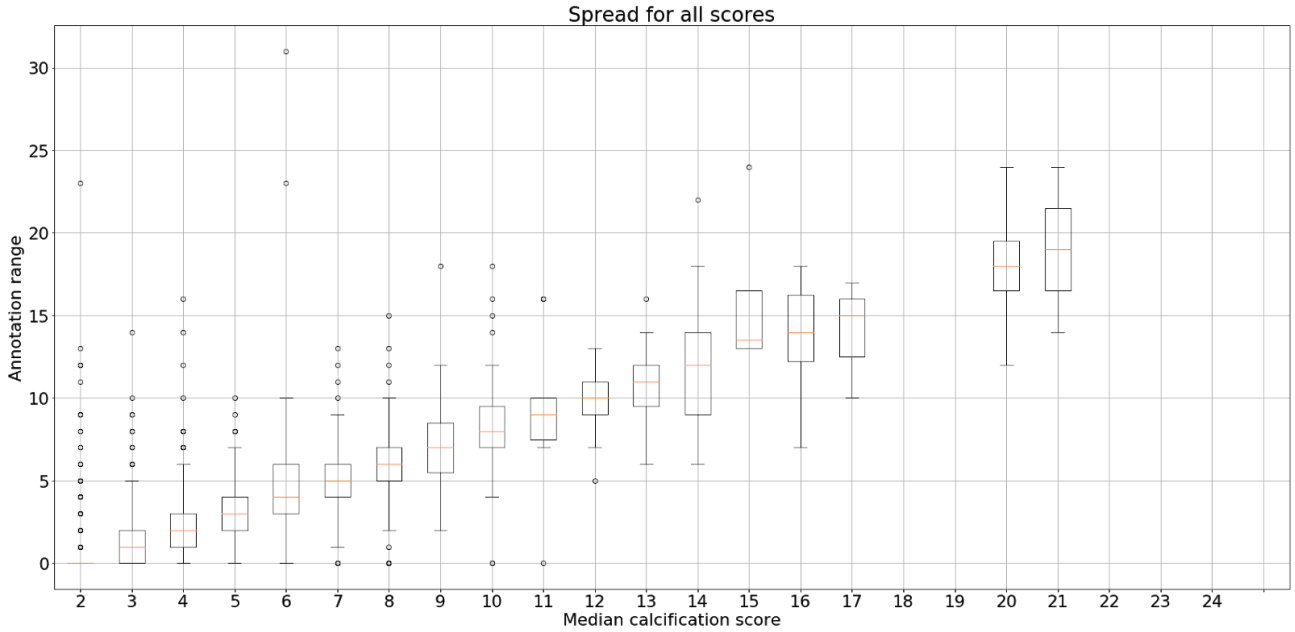


Figure S1: The variability in calcification across annotators increases for medium level calcified participants.

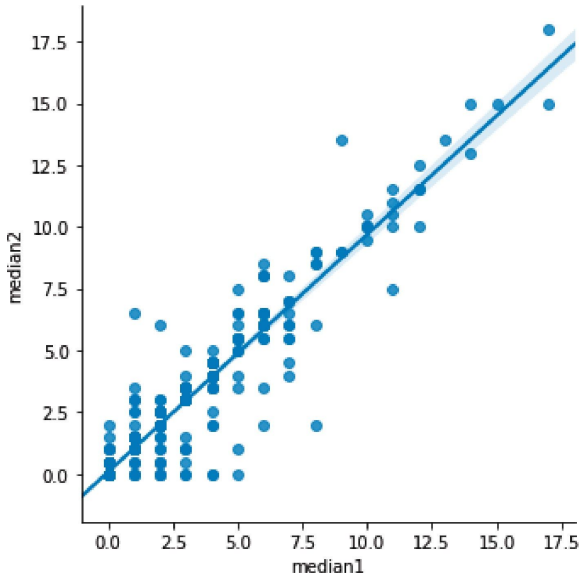


Figure S2: The median scores for 136 manually annotated scans from two rounds of annotation are highly correlated.

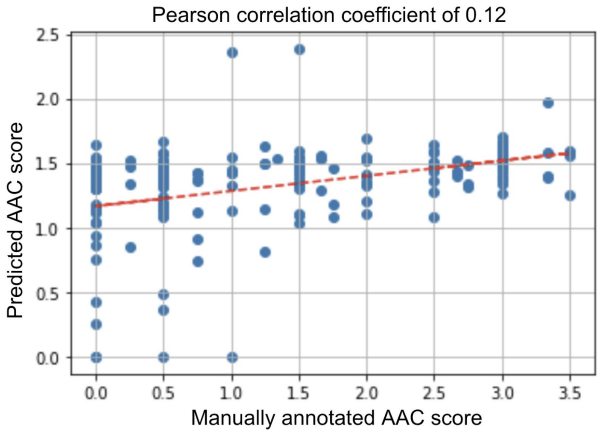


Figure S3: Conventional deep learning regression approaches with ResNet50 architecture performed poorly on a test set while quantifying aortic calcification from the complete lumbar spine DEXA scans.

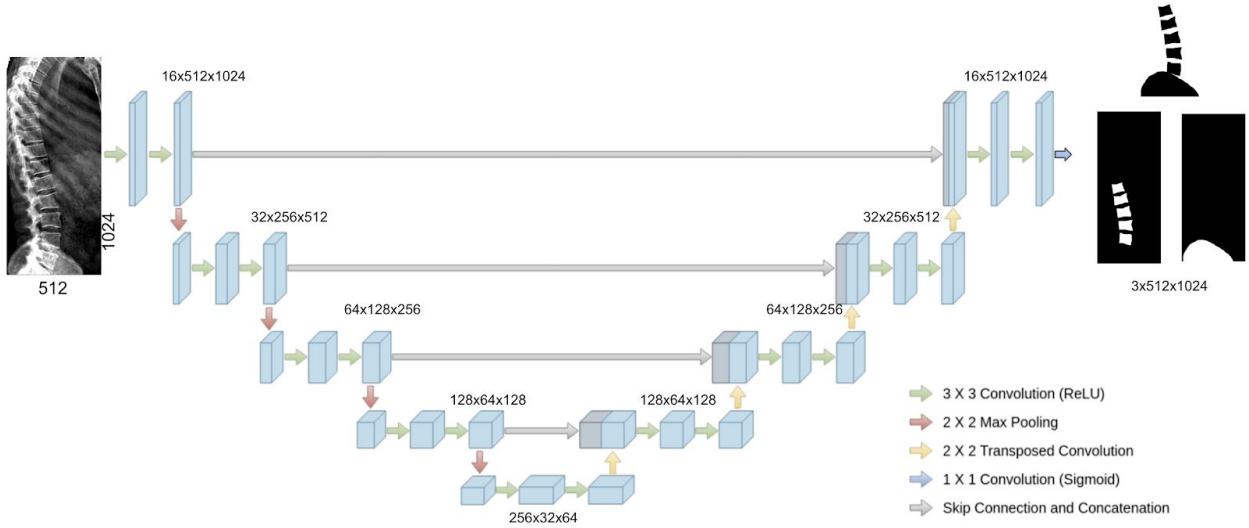


Figure S4: U-net architecture used for segmenting relevant spine and pelvis regions in pipeline 1. The input DEXA image is shown on the left and the binary masks for the 3 output classes - background, lower spine and pelvis - are shown on the right.

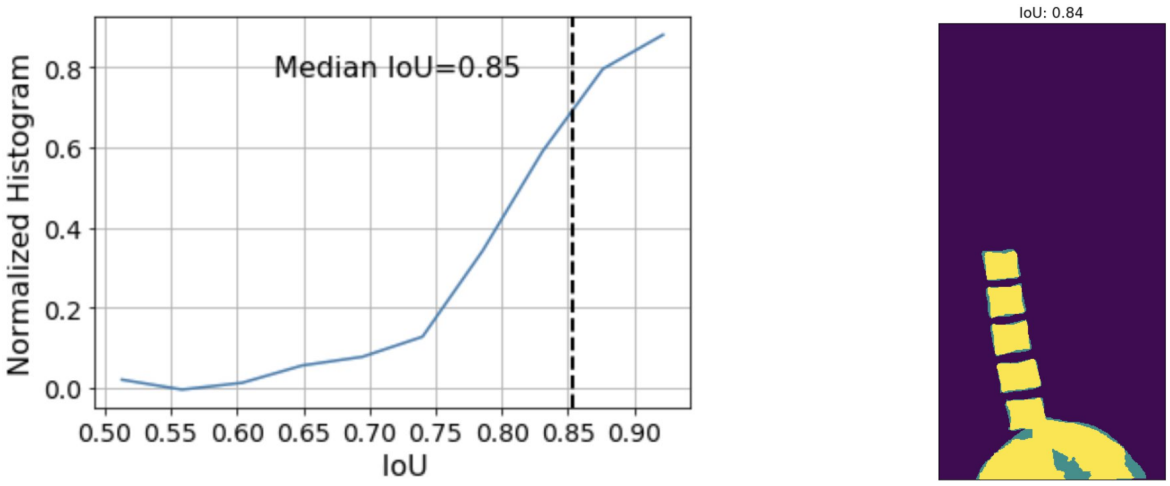


Figure S5: (a) shows the histogram of IoU values obtained across the test set along with the median value. **(b)** shows an overlay of the ground truth segmentation mask and the predicted segmentation mask for a typical example image whose IoU is close to the median value. The yellow regions show the regions of overlap while the blue regions show gaps in the prediction mask.

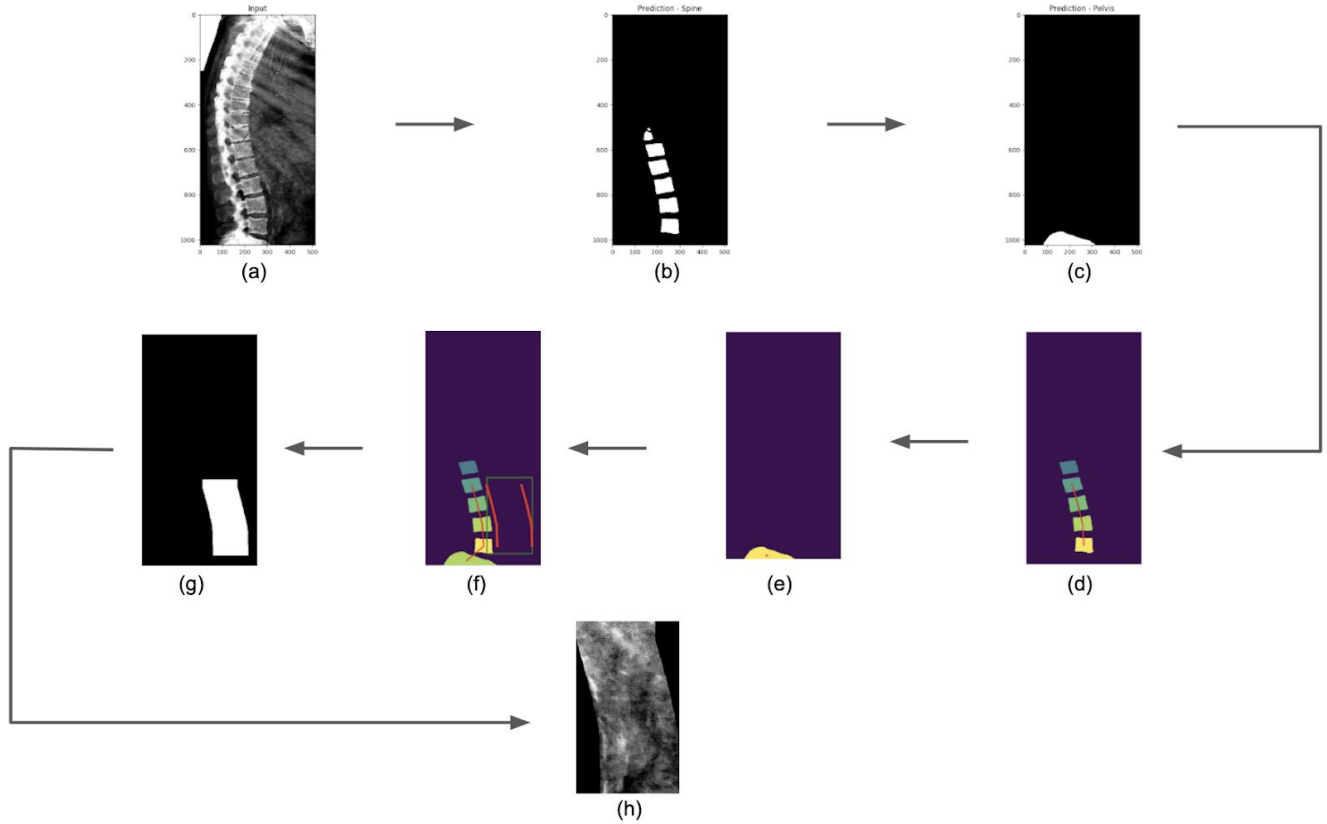


Figure S6: Schematic representation of the steps to extract abdominal aortic region from lumbar spine DEXA scans. (a) shows the input image. (b)-(c) show the predicted segmentation masks of the lower spine and the pelvis. (d)-(f) show the predicted

centroids of the different vertebrae (connected), predicted centroid of the pelvis and the overall spinal curvature together with the estimated aortic region to the right of the spine. (g) shows the binary mask of the predicted aortic region. (h) shows the aortic region extracted from the original input image. This is then fed into a regression model to score the calcification level.

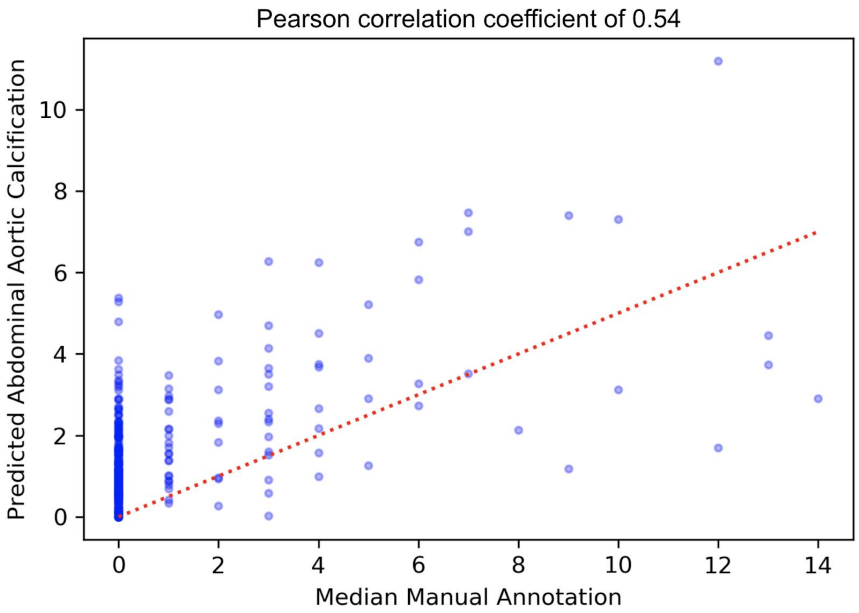


Figure S7: The predicted AAC from pipeline 1 as compared to manually annotated calcification scores for the validation dataset.

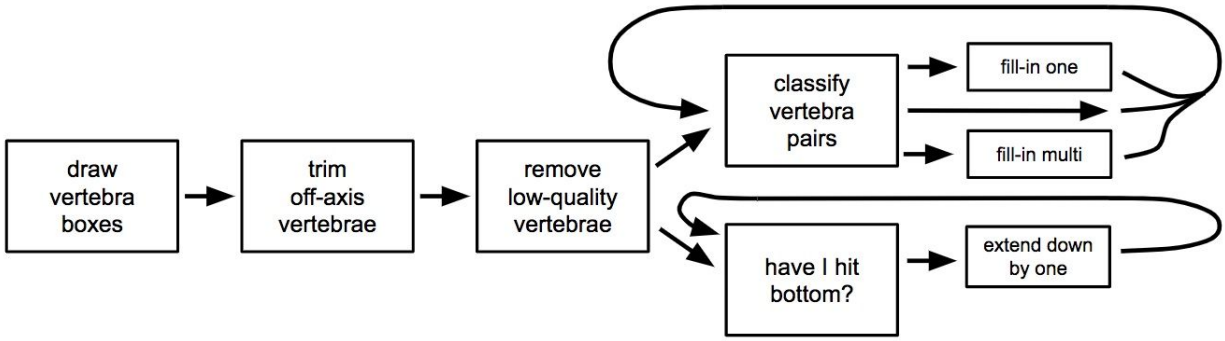
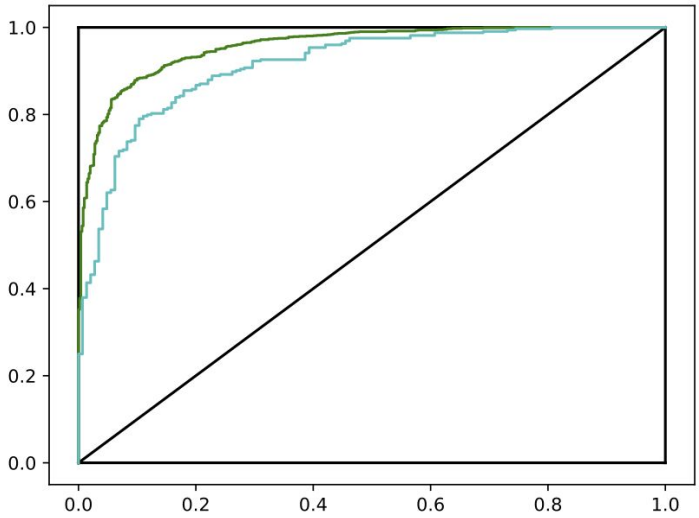


Figure S8: Segmentation steps within machine learning pipeline 2 for predicting AAC from DEXA scans.

A.



B.

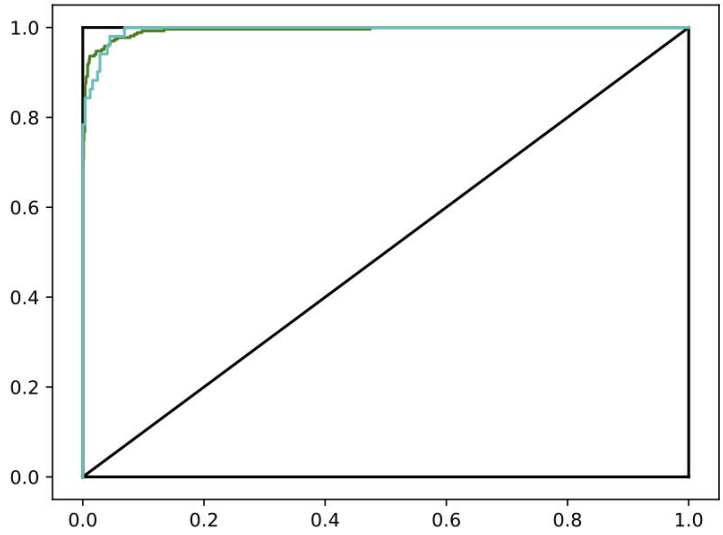
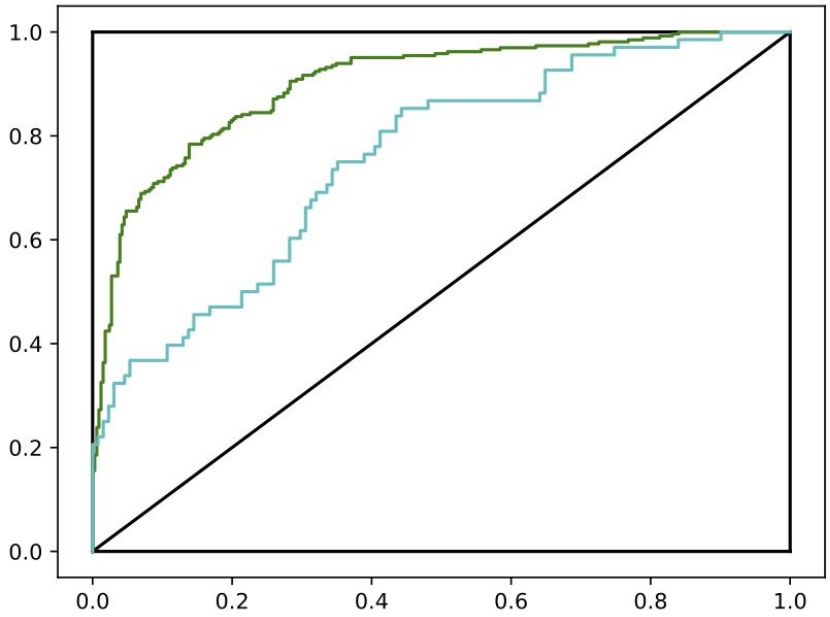


Figure S9: A. ROC curve measuring the accuracy of the models for eliminating bad vertebrae in step b of pipeline 2 for training dataset (green) and test dataset (green). The Cohen's kappa values for its performance on the training set (green) and test set (blue) are 0.743 and 0.638 respectively. B. ROC curve measuring the accuracy of the models for extending vertebra to the bottom of the spine in step e of the segmentation step of pipeline 2 for training dataset (green) and test dataset (green). The Cohen's kappa values for its performance on the training set (green) and test set (blue) are 0.917 and 0.874 respectively.

A)



B)

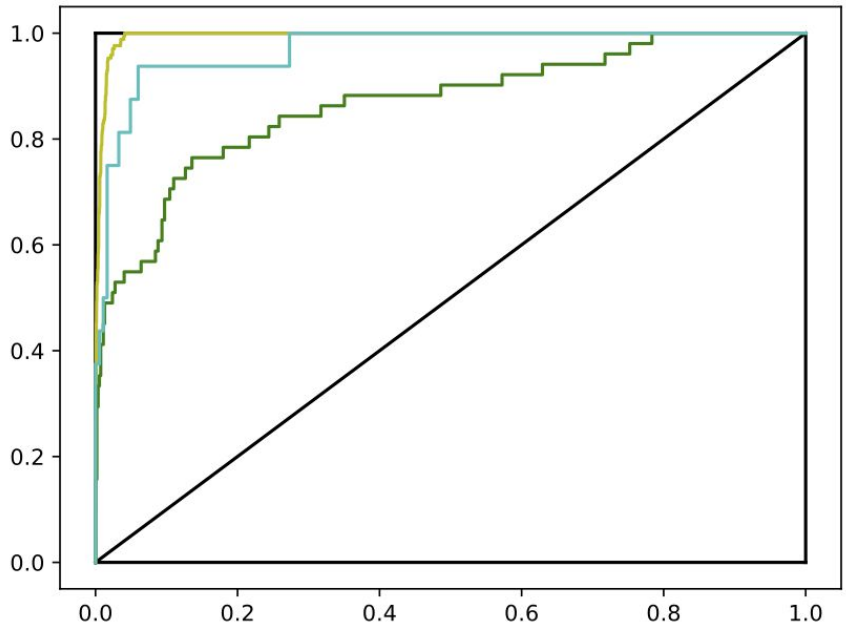


Figure **S10**: A. ROC curve measuring the accuracy of the model 1 that classifies calcified aorta from non-calcified aorta. The Cohen's kappa values for training (green) and test sets (blue) are 0.58 and 0.33 respectively. B. ROC curve measuring the accuracy of the model 2 that classifies high-threshold aortic calcification. The original training and test sets are shown in green and blue respectively while the sandbox enriched training set is shown in yellow in B. The Cohen's kappa values for original training, sandbox enriched training and test sets are 0.53, 0.71, and 0.73 respectively.

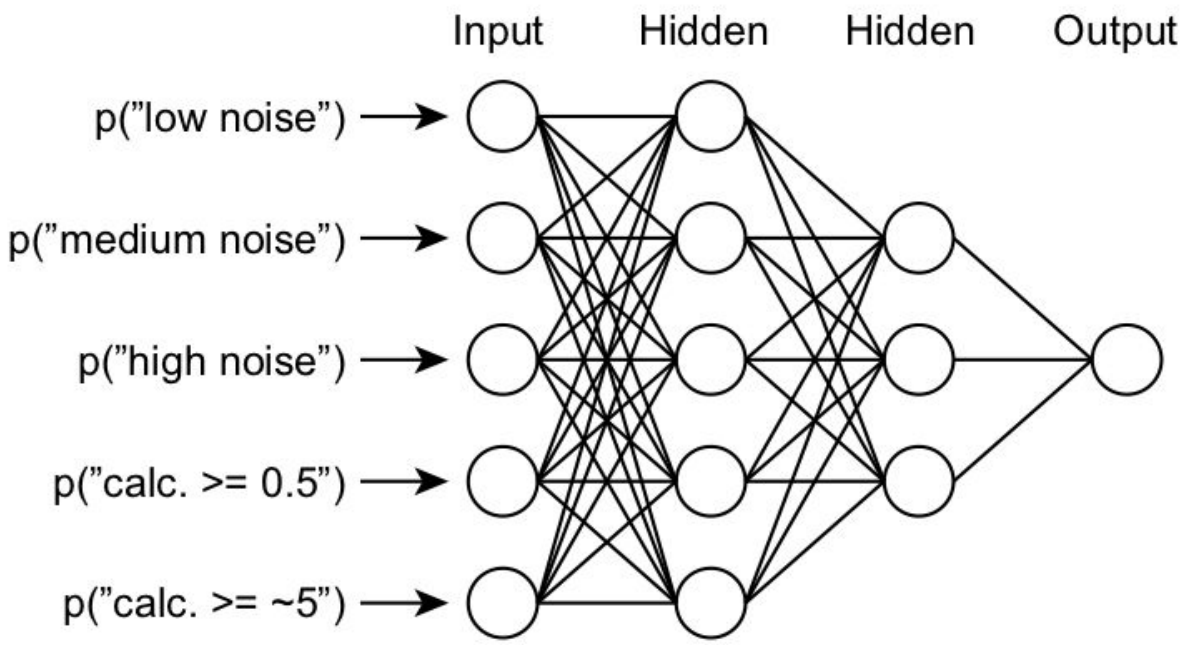


Figure S11: Architecture of neural network used to regress calcification values from output of models 1 and 2 for classifying calcified images from noncalcified images.

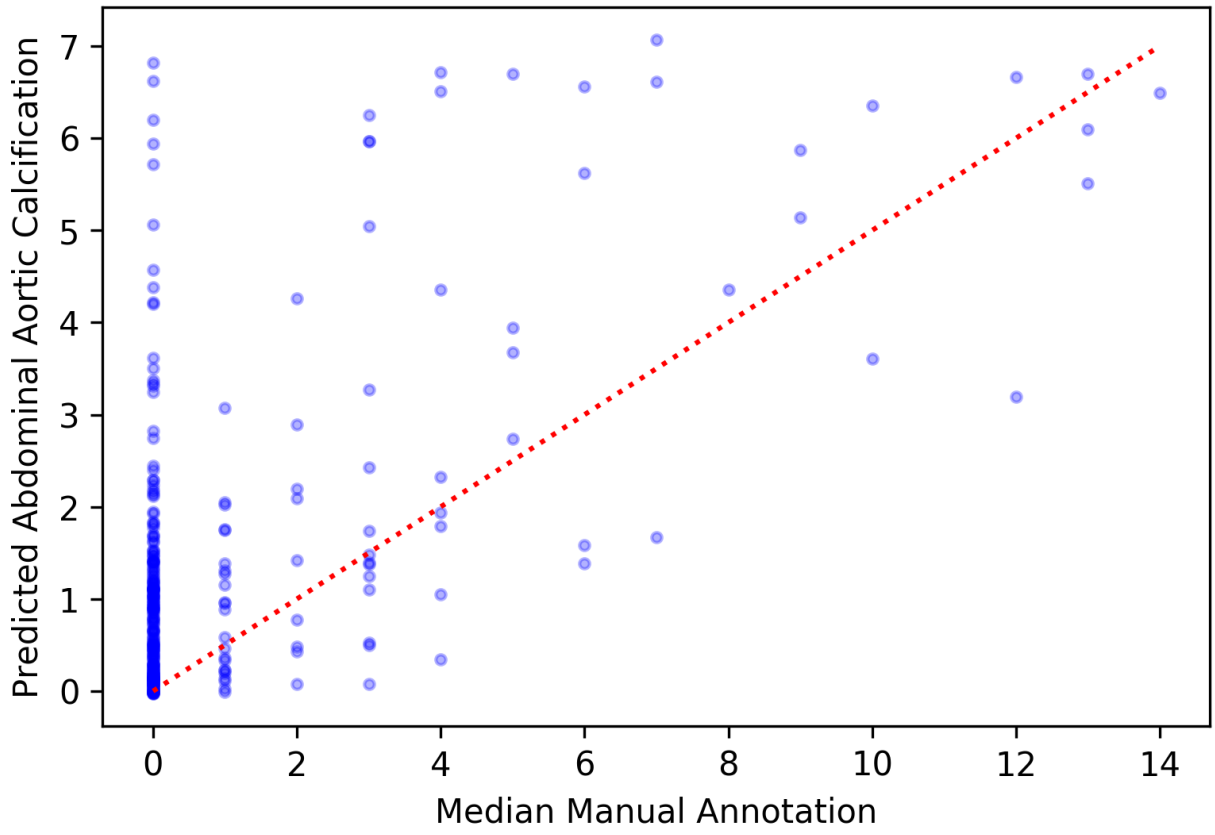


Figure S12: The comparison of predicted AAC to median annotation scores for the validation dataset from pipeline 2.

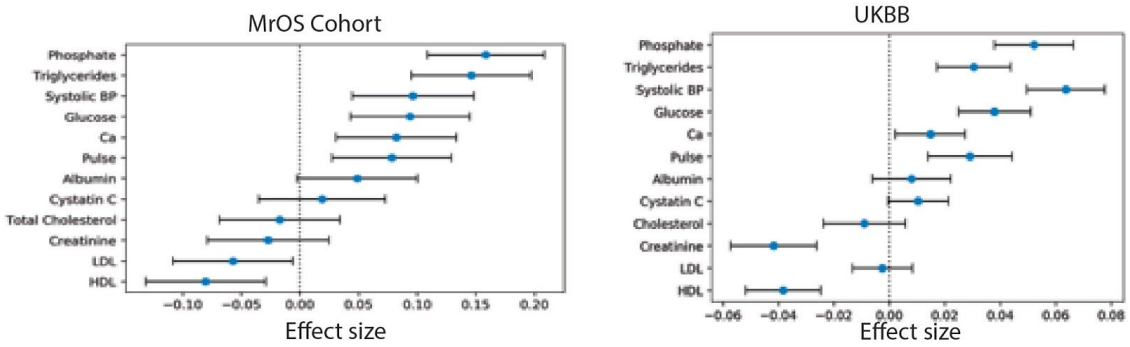


Figure S13: Association of biomarkers with AAC in the MrOS cohort and in this study. The biomarkers that are significantly associated with AAC in MrOS cohort are also significantly associated with predicted AAC in this study even though the effect sizes are slightly lower in the UK biobank cohort. All associations are calculated after adjusting for age and sex.

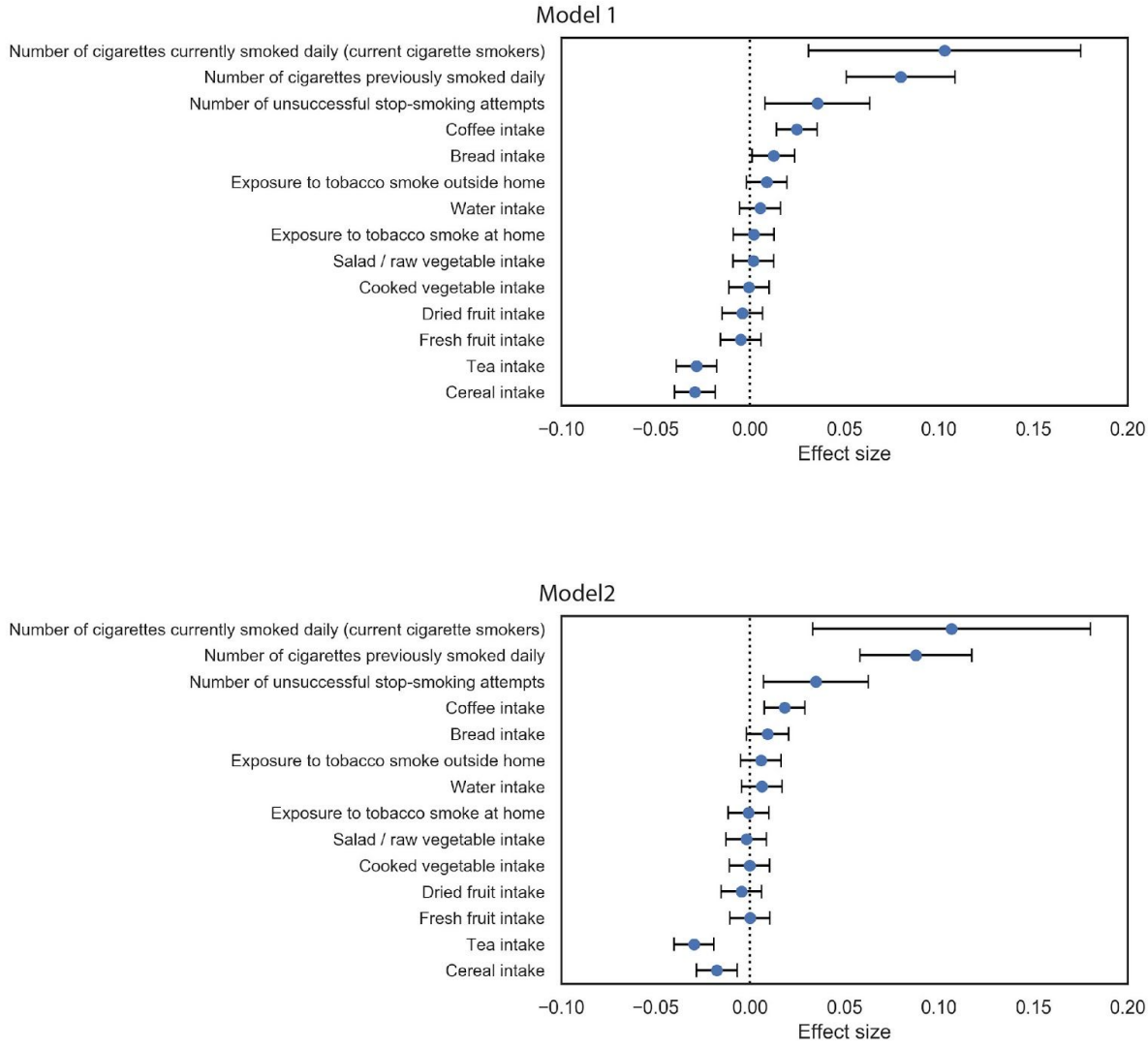


Figure S14: Association of predicted AAC with diet. Univariate regression analysis of risk factors at baseline for predicted AAC after adjusting for age and sex in model 1 and after adjusting for socioeconomic factors, BMI, and smoking status in addition to adjusting for age and sex in model 2. The blue dots represent mean effect size while the intervals represent standard errors for the effect size.

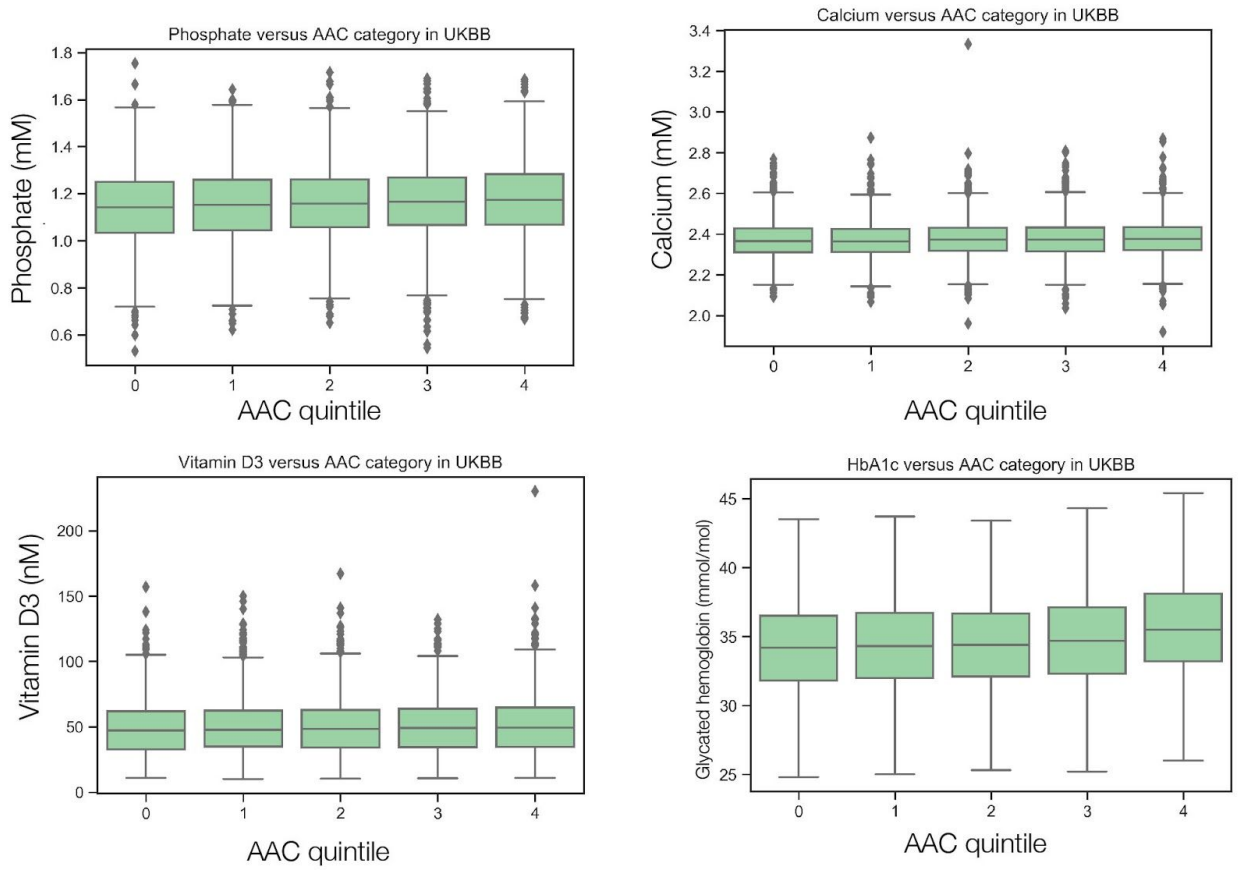


Figure S15: Box plot showing the level of different risk factors (serum Phosphate, Calcium, Vitamin D3, and glycated Hemoglobin HbA1c) of participants stratified according to level of calcification (higher AAC score represented by larger group number). There is very little change in the distribution of different biomarkers as a function of AAC group (except for HbA1c levels in the highest quintile of AAC) and most participants would be within clinically healthy levels for these biomarkers. The center line at each age represents the median biomarker level for all participants in that calcification category while the lower and upper boundaries of the box indicate the 25th and 75th percentile of the serum biomarker level for that calcification category. The lower and upper ends of the lines denote the 95% confidence interval of the serum biomarker level for that calcification category.

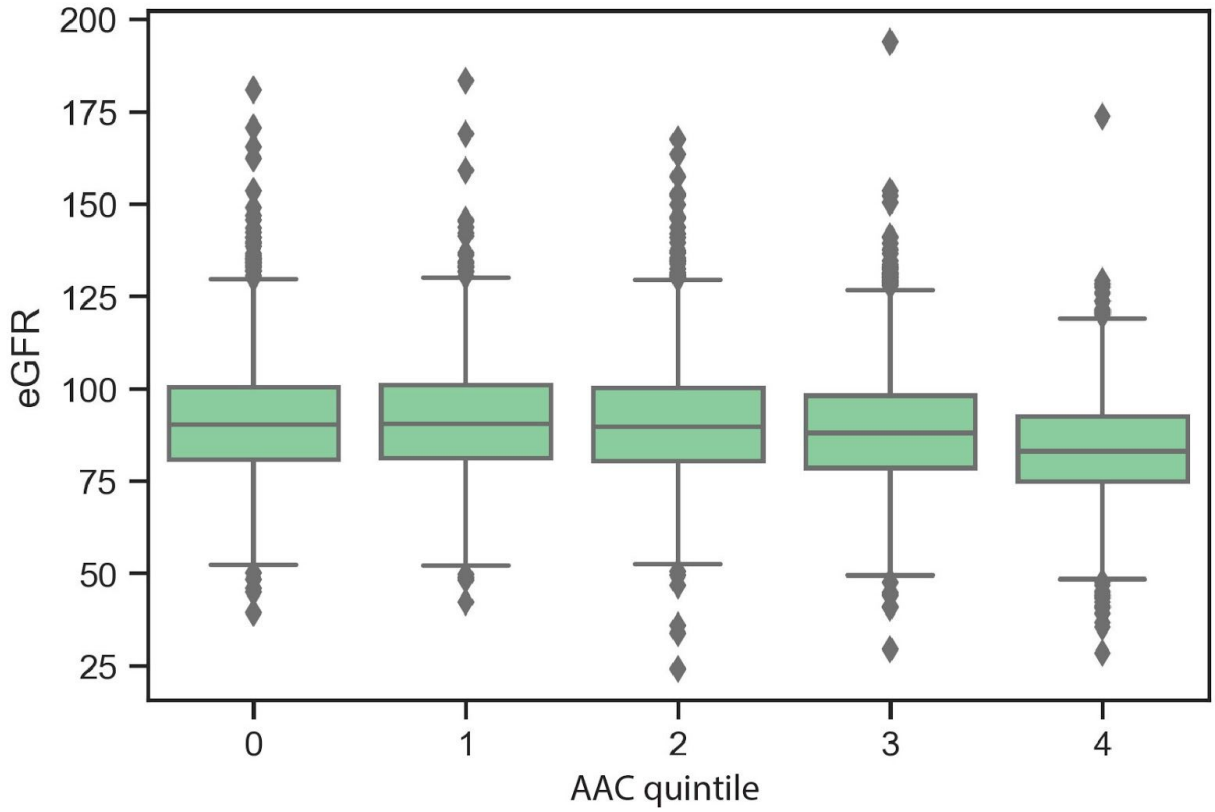


Figure S16: Box plot showing the variation of estimated glomerular filtration rate (eGFR) of participants grouped according to level of (higher AAC score represented by larger group number). The eGFR was calculated from Cystatin-C levels using the formula (Grubb et al. 2014). The eGFR for the participants with the highest levels of AAC (categories 3 and 4) tend to be lower, indicating that their kidneys are not functioning as efficiently as the other participants on average. The center line at each age represents the median biomarker level for all participants in that calcification category while the lower and upper boundaries of the box indicate the 25th and 75th percentile of the eGFR for that calcification category. The lower and upper ends of the lines denote the 95% confidence interval of the eGFR for participants within a particular calcification category.

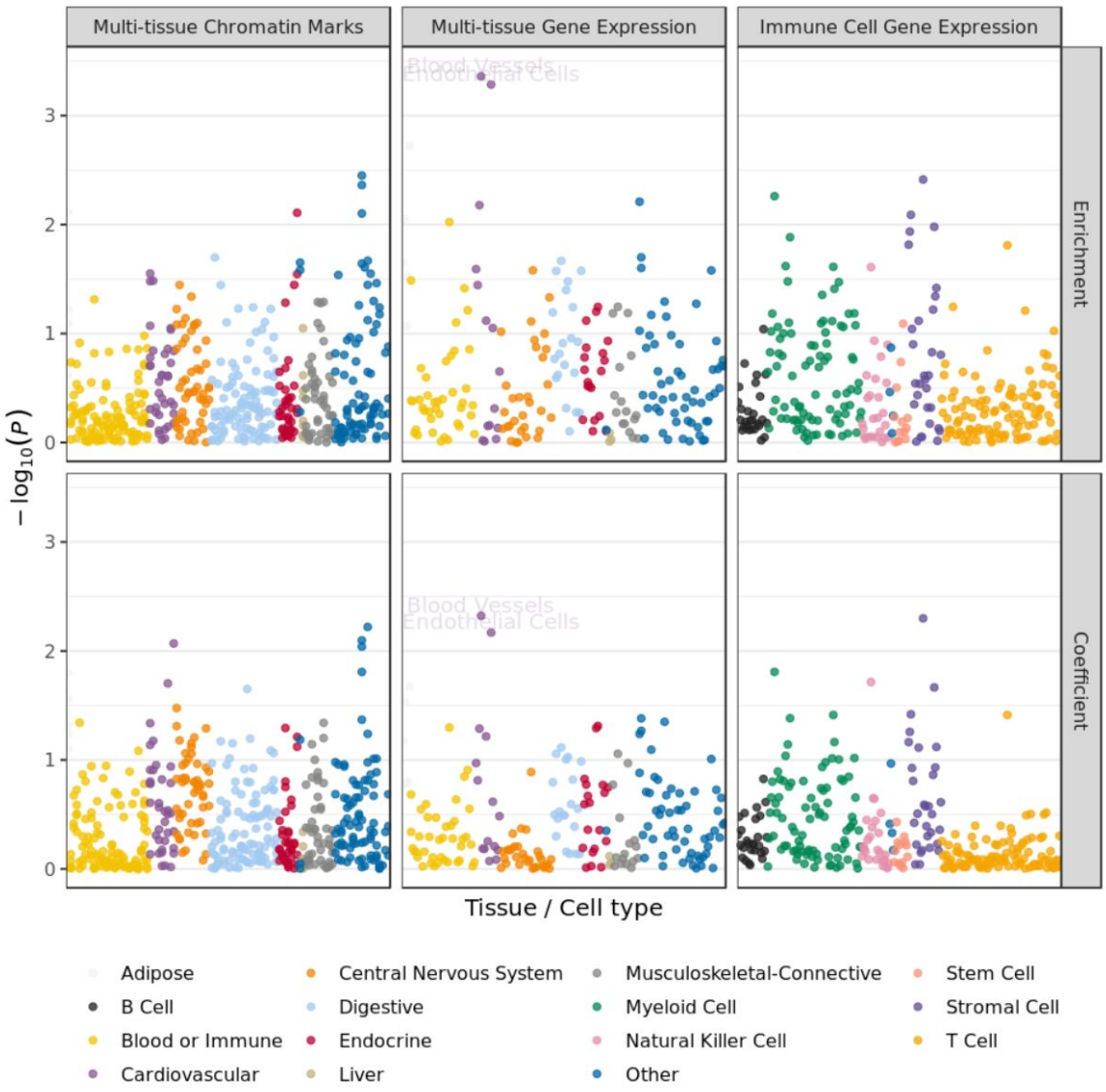


Figure S17 Heritability enrichment of AAC across different genomic annotations. P-values (y-axis) of either total enrichment (top row facet) or positive regression coefficients (tau; bottom row facet) across annotations (x-axis) and datasets (column facets). Enrichment is calculated by dividing the proportion of heritability explained by SNPs in an annotation by proportion of SNPs in an annotation ($\text{proportion}_{h2}/\text{proportion}_{SNPs}$). Regression coefficients represent the average contribution of an annotation to per-SNP heritability, correcting for annotations in the baseline model (Methods).

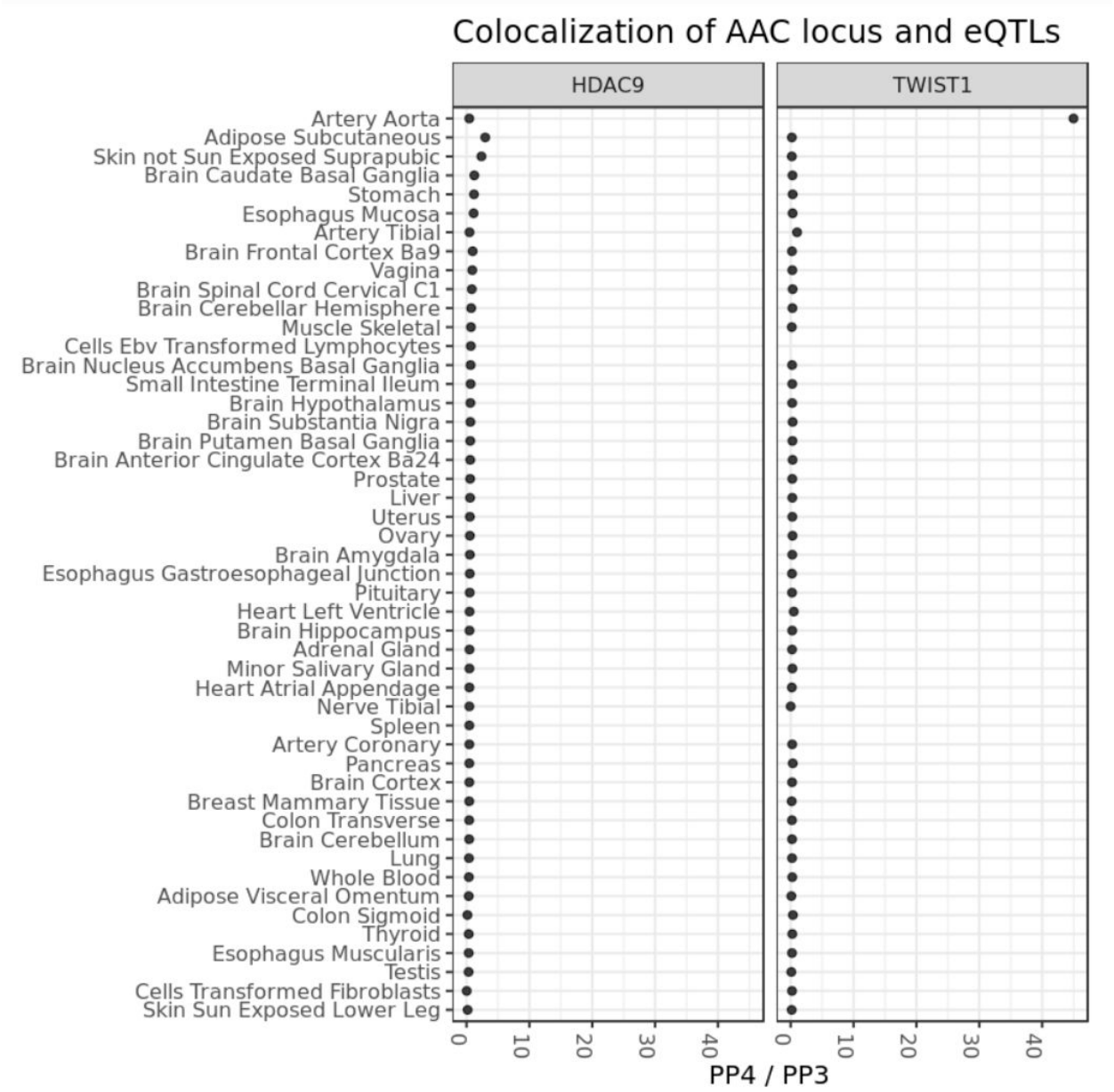


Figure S18. Colocalization of AAC and eQTLs for proximal genes (facets) across various tissues (y-axis). The x-axis shows the posterior probability of configuration 4 (PP4; one shared SNP associated with trait 1 and trait 2) over the posterior probability of configuration 3 (PP3; two independent SNPs associated with trait 1 and trait 2). Missing points for *TWIST1* due to *TWIST1* not having sufficient expression in “spleen” or “Epstein-Barr virus transformed lymphocytes” to be tested in the GTEx v7 eQTL analysis.

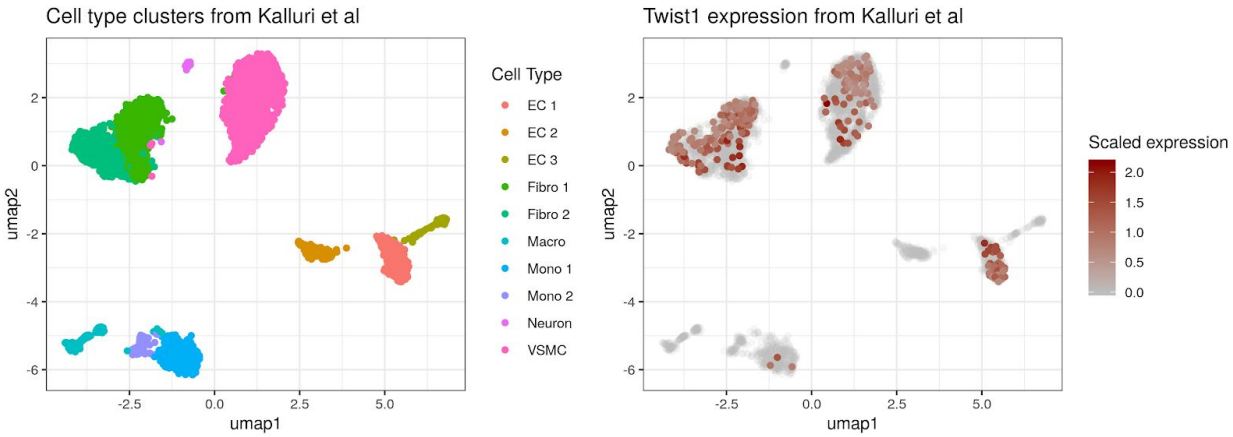
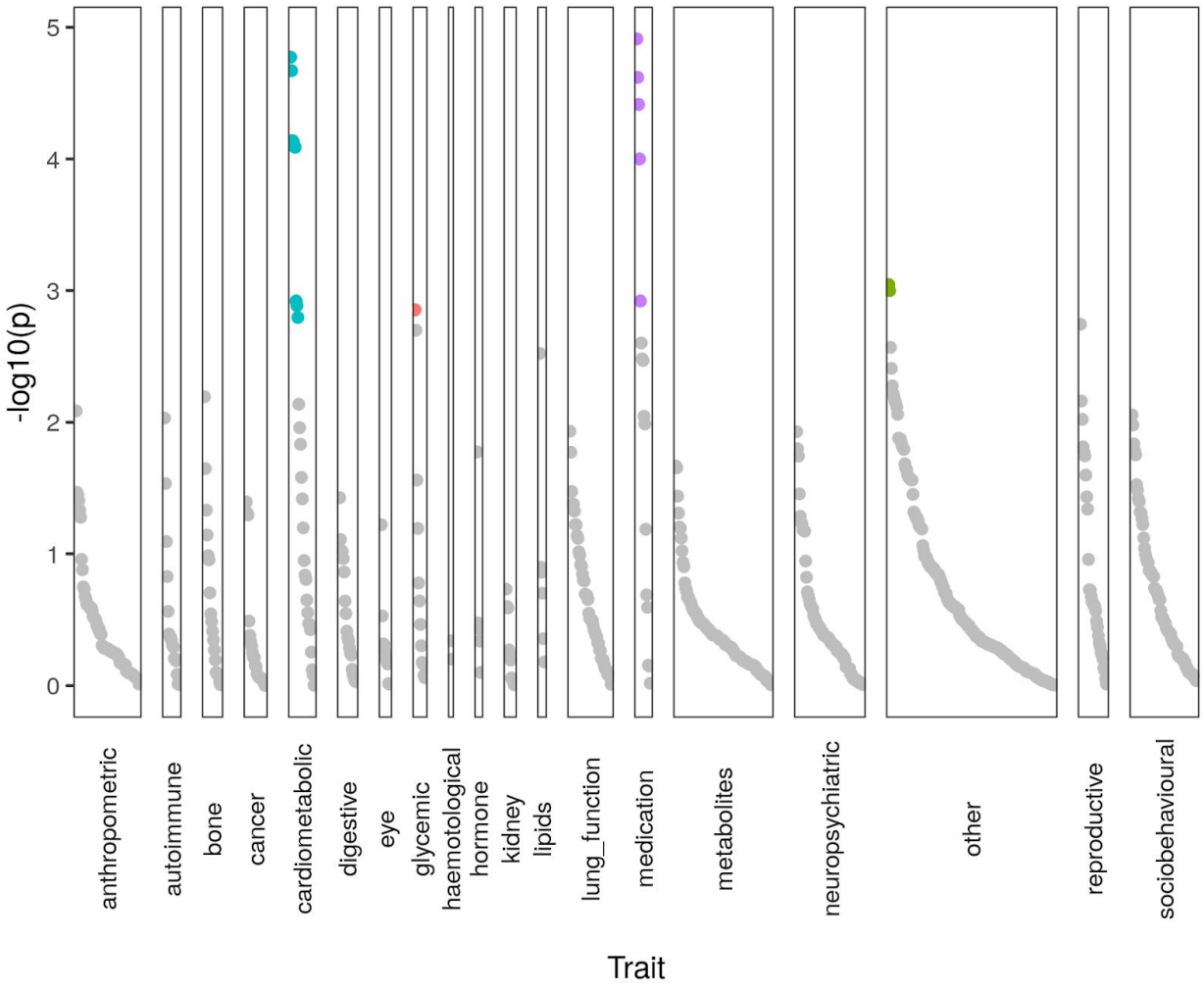


Figure S19: Expression of *Twist1* in mouse aorta cell subtypes, with clusters as identified in (Kalluri et al. 2019). EC1 corresponds to a subset of endothelial cells characterized by expression of genes involved in extracellular matrix organization. This corresponds to a cluster identified in (Lukowski et al. 2019) as having a mesenchymal phenotype, quiescence, and high mitochondrial content, preceding the others two more differentiated states in pseudotime.



Supplementary Figure S20: P-value for genetic correlation between AAC and 754 traits, computed using LD Score Regression. Points are colored if they are significant after adjustment for multiple testing (Benjamini-Hochberg q-value < 0.05).

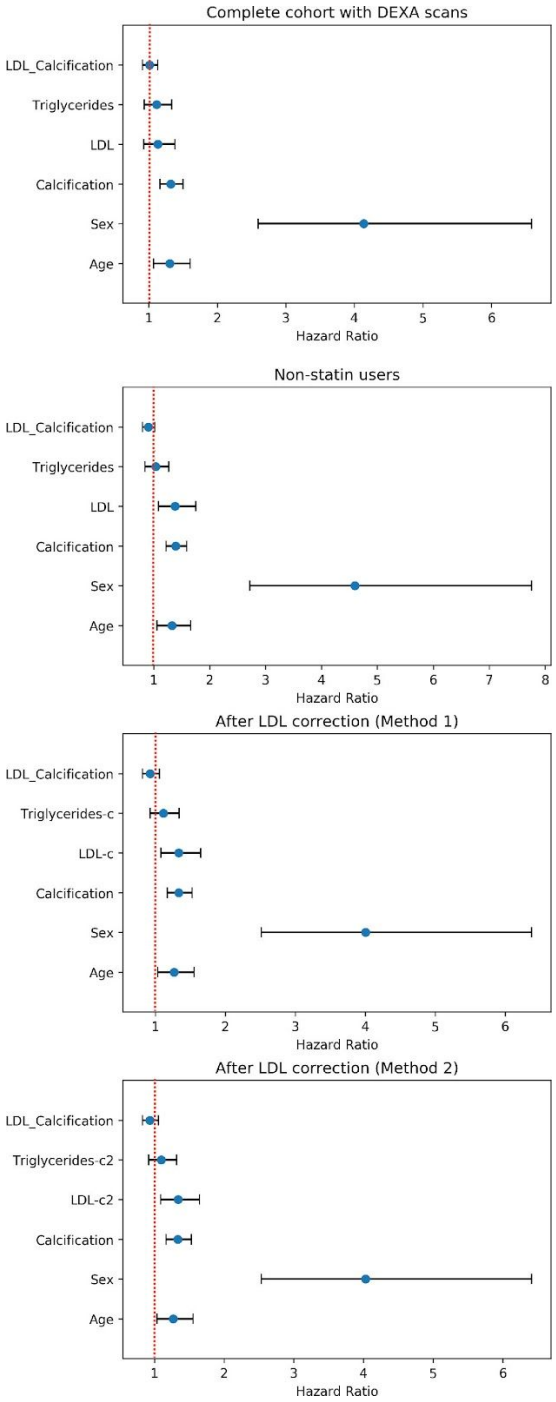


Figure S21: The hazard ratios for different covariates in different Cox proportional hazard models for predicting acute MI events in the future for (A) complete cohort with experimentally measured LDL levels, (B) nonstatin users in cohort with experimentally measure LDL levels, (C) complete cohort with statin corrected LDL levels (method 1), and (D) complete cohort with statin corrected LDL levels (method 2).

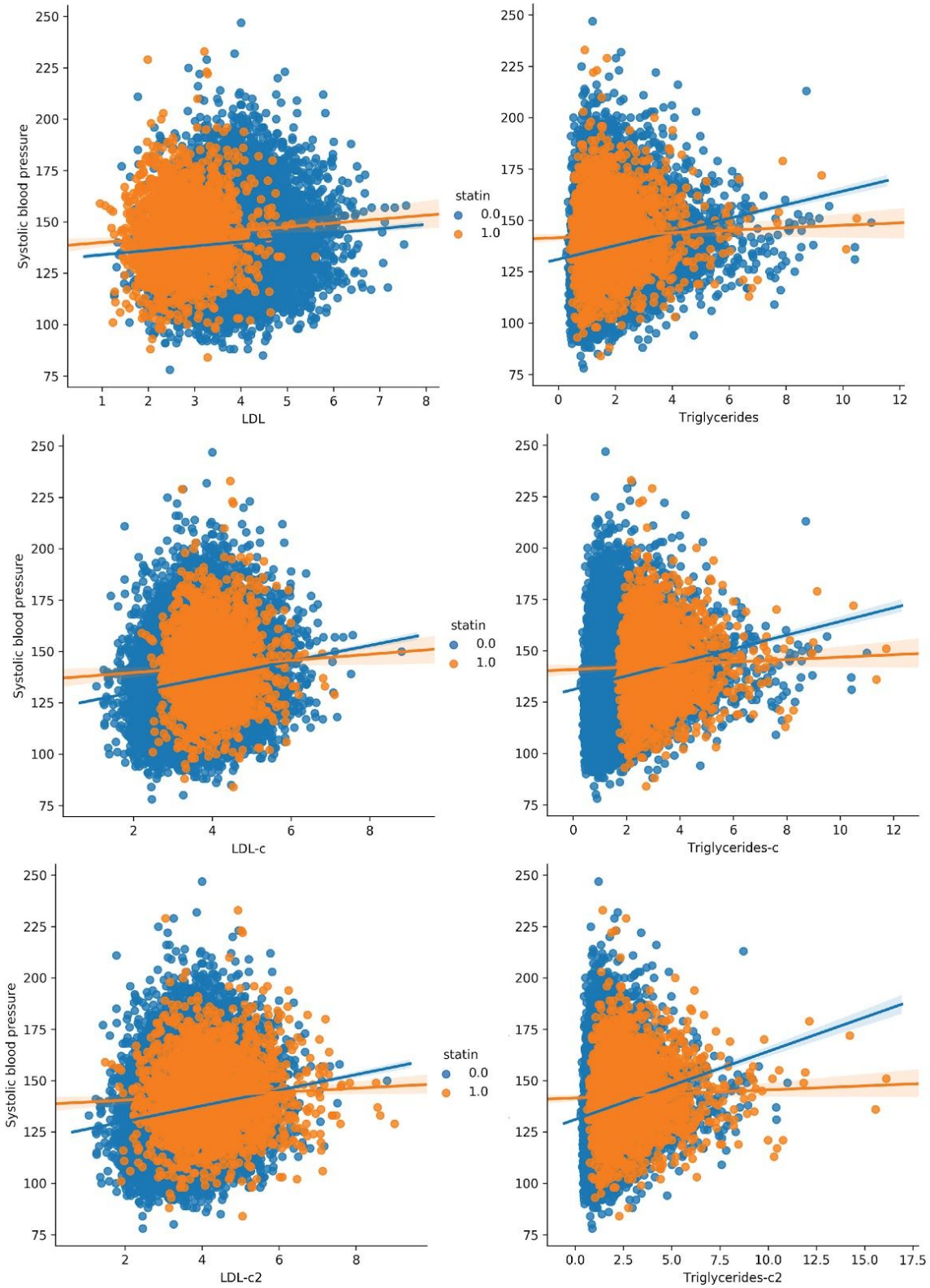


Figure S22: Effect of statins on LDL-Blood Pressure relationship. The relationship between baseline LDL and Triglyceride levels to blood pressure is plotted for all UK biobank participants before and after statin correction (c1 and c2 represent the two statin

correction methods). Blue represents people who don’t take statin medication are shown in blue while statin takers are shown in orange. The uncorrected LDL levels are lower than that predicted by blood pressure for statin takers and the correction methods reduce this discrepancy.

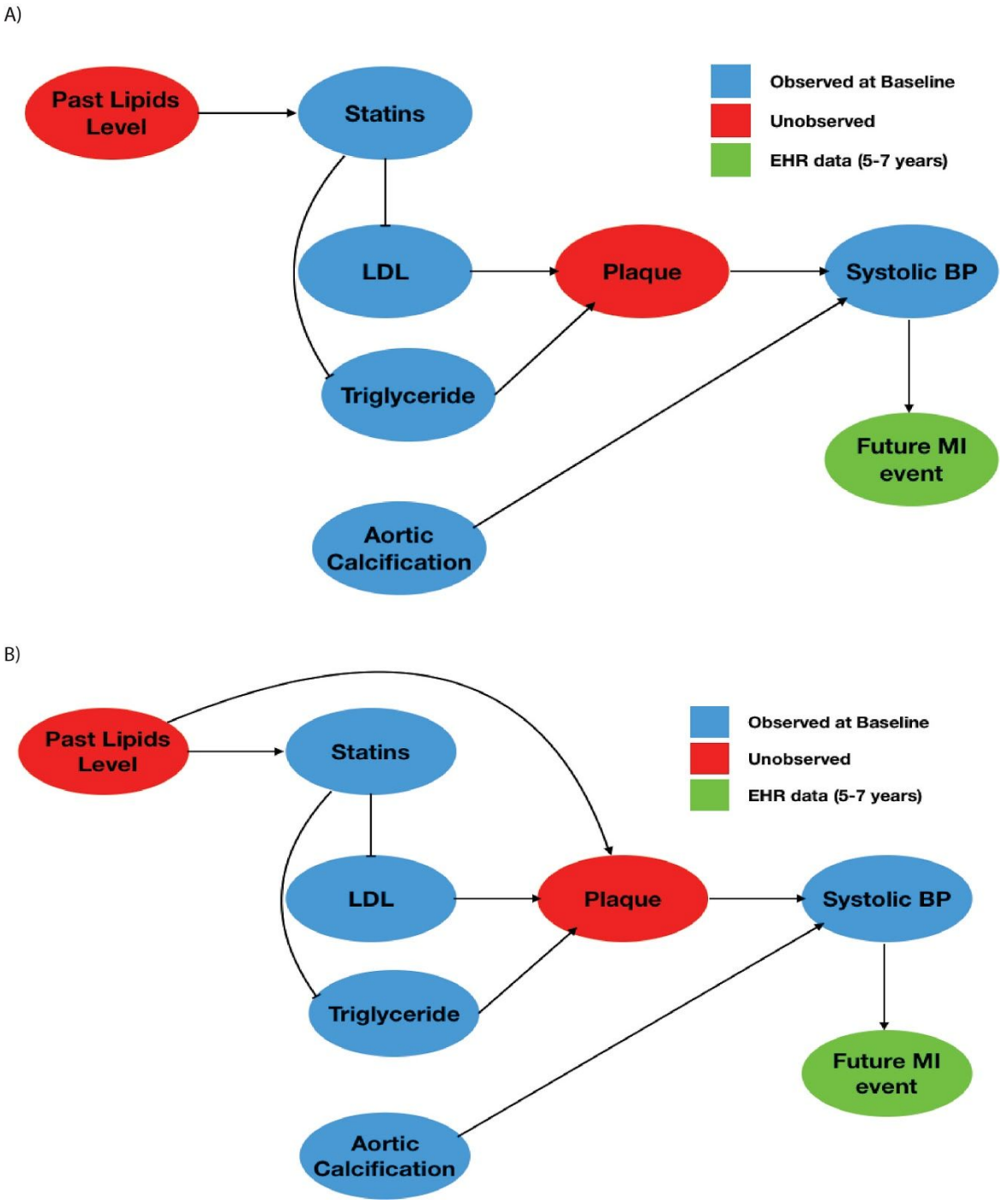


Figure S23: Graphical Models for MI event based on LDL and AAC. The relationship between different lipid levels and their predictions of future MI events. LDL and triglyceride levels affect plaque formation increasing systolic blood pressure and the probability of future myocardial infarction events. Aortic calcification forms a parallel independent pathway for systolic blood pressure increase. Statins reduce LDL and triglyceride levels and in (A) do not affect acute MI events as they only affect blood

pressure through LDL levels. However, in **Figure S22,** we show that statin takers do have higher blood pressure than that predicted by LDL levels alone and this is because statins do not remove old plaque that is determined by past lipid levels (B). Hence, we apply statin correction to LDL levels to get an unbiased estimate of LDL risk and compare it to Aortic Calcification risk in this article.

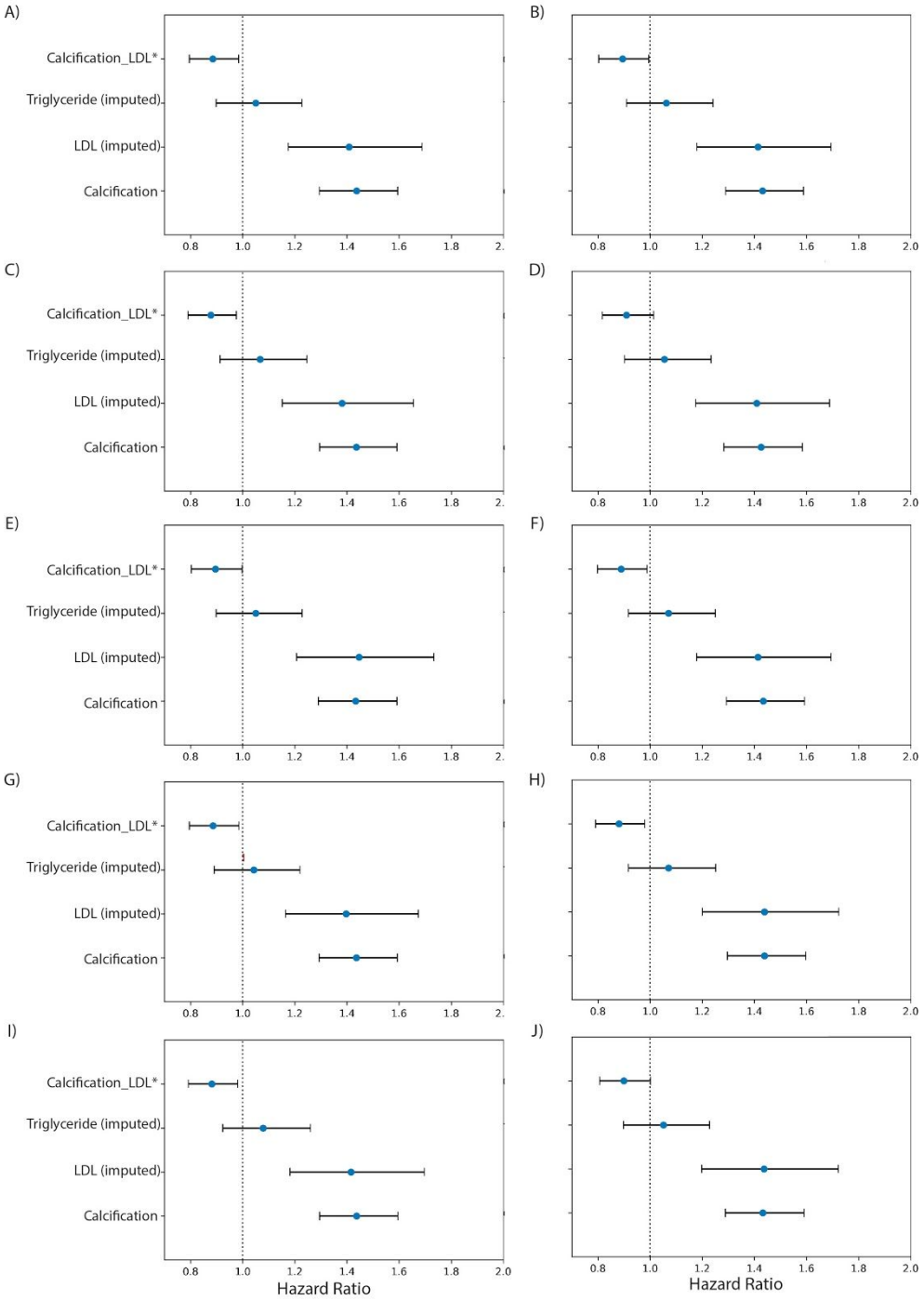


Figure S24: Comparison of LDL risk to Calcification risk with LDL imputation for statin users. Ten random initializations with MICE were used to impute the LDL level for statin users while LDL levels for non-statin users were used unadjusted. Cox proportional hazards were used to compare the hazard from LDL and aortic calcification over the UK

biobank population. In spite of the variability from imputation, the hazard ratios for LDL and calcification risk levels were similar across all 10 models.

Supplementary Tables

A)

	Annotator 1	Annotator 2	Annotator 3	Annotator 4	Median
Annotator 1	1	0.84	0.72	0.74	0.92
Annotator 2	0.84	1	0.75	0.78	0.93
Annotator 3	0.72	0.75	1	0.69	0.85
Annotator 4	0.74	0.78	0.69	1	0.86
Median	0.92	0.93	0.85	0.86	1

B)

	Annotator 1	Annotator 2	Annotator 3	Annotator 4	Median
Annotator 1	1	0.61	0.56	0.53	0.74
Annotator 2	0.61	1	0.57	0.53	0.8
Annotator 3	0.56	0.57	1	0.55	0.79
Annotator 4	0.53	0.53	0.55	1	0.75
Median	0.74	0.8	0.79	0.75	1

Table S1: Variability of manual annotation over the training set for the 4 different annotators measured using (A) Pearson Correlation and (B) Spearman correlation.

A.

	Annotator 1	Annotator 2	Annotator 3	Annotator 4	Median
Annotator 1	1	0.86	0.78	0.75	0.94
Annotator 2	0.86	1	0.73	0.67	0.93
Annotator 3	0.78	0.73	1	0.7	0.85
Annotator 4	0.75	0.67	0.7	1	0.73
Median	0.94	0.93	0.85	0.73	1

B.

	Annotator 1	Annotator 2	Annotator 3	Annotator 4	Median
Annotator 1	1	0.68	0.67	0.56	0.83
Annotator 2	0.68	1	0.71	0.51	0.88
Annotator 3	0.67	0.71	1	0.55	0.84
Annotator 4	0.56	0.51	0.55	1	0.57
Median	0.83	0.88	0.84	0.57	1

Table S2: Variability of manual annotation over the validation set for the 4 different annotators measured using (A) Pearson Correlation and (B) Spearman correlation.

	MrOS Cohort	UKBB Cohort
Phenotype	Correlation	Correlation
Age	0.3	0.28
Systolic Blood Pressure	0.15	0.18
Pulse	0.09	0.02
Glucose	0.08	0.06
Cystatin C	0.12	0.09
HDL	-0.06	-0.01
Phosphate	0.17	0.05
Triglycerides	0.1	0.04

Table S3: Correlation of annotated abdominal aortic calcification with different phenotypes across MrOS and UKBB cohorts. Only phenotypes significantly correlated with aortic calcification in MrOS cohort are displayed in the above table.

	Model 1		Model 2	
Biomarker	Effect size	P-value	Effect size	P-value
Albumin	0.006 (0.006)	0.3	0.003 (0.006)	0.6
Alkaline phosphatase	-0.001 (0.006)	0.9	0.004 (0.006)	0.5
Alanine aminotransferase	0.005 (0.006)	0.4	0.02 (0.006)	5.50E-03
Apolipoprotein A	-0.03 (0.006)	1.30E-06	-0.05 (0.006)	9.70E-13
Apolipoprotein B	0.02 (0.006)	1.50E-04	0.03 (0.006)	4.90E-06
Aspartate aminotransferase	0.005 (0.006)	0.4	0.01 (0.006)	0.06
Direct bilirubin	-0.008 (0.006)	0.2	-0.005 (0.006)	0.4
Urea	-0.008 (0.006)	0.2	0.0003 (0.006)	0.95
Calcium	0.02 (0.006)	1.00E-03	0.02 (0.006)	1.00E-03
Cholesterol	-0.008 (0.006)	0.2	-0.008 (0.006)	0.2
Creatinine	-0.04 (0.006)	1.40E-07	-0.02 (0.006)	1.00E-03
C-reactive protein	0.01 (0.006)	0.09	0.02 (0.006)	0.001
Cystatin C	0.01 (0.006)	0.04	0.02 (0.006)	0.003
Gamma glutamyltransferase	0.02 (0.006)	9.20E-05	0.03 (0.006)	1.70E-06
Glucose	0.03 (0.006)	1.40E-07	0.04 (0.006)	7.10E-10
Glycated haemoglobin (HbA1c)	0.07 (0.006)	1.10E-36	0.08 (0.006)	1.20E-40
HDL	-0.04 (0.006)	3.20E-10	-0.06 (0.007)	1.40E-19
IGF-1	-0.006 (0.006)	0.3	-0.004 (0.006)	0.5
LDL	-0.002 (0.006)	0.7	0.002 (0.006)	0.7
Lipoprotein A	0.001 (0.006)	0.8	0.003 (0.006)	0.4
Oestradiol	0.004 (0.011)	0.7	-0.004 (0.03)	0.9
Phosphate	0.05 (0.006)	4.60E-16	0.04 (0.006)	4.00E-13
Rheumatoid factor	-0.002 (0.006)	0.9	-0.003 (0.006)	0.9
Sex hormone binding globulin	0.02 (0.007)	0.01	0.0003 (0.006)	0.97
Total bilirubin	0.02 (0.006)	2.00E-03	-0.01 (0.006)	0.03
Testosterone	-0.02 (0.02)	0.3	-0.04 (0.02)	0.06
Total protein	0.01 (0.006)	0.07	0.02 (0.006)	1.00E-03
Triglycerides	0.02 (0.006)	9.20E-05	0.03 (0.006)	1.20E-07
Urate	0.004 (0.007)	0.6	0.02 (0.007)	4.00E-03

Vitamin D	-0.0005 (0.006)	0.9	-0.004 (0.006)	0.5

Table S4: Association of predicted AAC with biomarkers Univariate regression analysis of risk factors at baseline for predicted AAC after adjusting for age and sex in model 1 and after adjusting for socioeconomic factors, BMI, and smoking status in addition to adjusting for age and sex in model 2. The estimate of the effect size and the standard error of the estimated effect size are given along with p-values for each coefficient in a univariate fit.

Physiological Measure	Model 1		Model 2	
	Effect size	P-value	Effect size	P-value
Hand grip strength (left)	-0.03 (0.008)	1.20E-04	-0.03 (0.008)	8.35E-05
Hand grip strength (right)	-0.02 (0.008)	4.00E-03	-0.02 (0.008)	3.48E-05
Waist circumference	-0.03 (0.008)	9.00E-06	0.007 (0.012)	0.59
Hip circumference	-0.04 (0.008)	4.30E-12	-0.03 (0.01)	0.011
Standing height	-0.05 (0.008)	2.50E-09	-0.05 (0.008)	5.70E-11
Fasting time	0.02 (0.005)	3.00E-04	0.02 (0.005)	3.06E-03
Heel bone ultrasound T-score	-0.07 (0.016)	5.90E-04	-0.06 (0.016)	1.88E-04
Heel bone mineral density (BMD) T-score	-0.04 (0.008)	8.40E-09	-0.03 (0.008)	1.50E-05
Pulse rate	0.02 (0.006)	1.00E-03	0.02 (0.006)	5.75E-05
Number of correct matches in round	0.005 (0.005)	0.4	0.004 (0.005)	0.46
Number of incorrect matches in round	-0.001 (0.006)	0.8	-0.002 (0.005)	0.68
Time to complete round	-0.0002 (0.006)	0.97	0.0002 (0.006)	0.97
Number of times snap-button pressed	-0.01 (0.005)	0.08	-.0.009 (0.005)	0.1
Duration to first press of snap-button in each round	-0.007 (0.008)	0.4	-0.006 (0.008)	0.45
Number of days/week walked 10+ minutes	0.004 (0.005)	0.5	-0.002 (0.005)	0.76
Duration of walks	0.01 (0.04)	0.04	0.006 (0.006)	0.24
Number of days/week of moderate physical activity 10+ minutes	0.0006 (0.005)	0.9	-0.007 (0.005)	0.21
Duration of moderate activity	0.005 (0.006)	0.4	-0.001 (0.006)	0.85
Number of days/week of vigorous physical activity 10+ minutes	-0.004 (0.005)	0.4	-0.008 (0.005)	0.15
Duration of vigorous activity	-0.005 (0.007)	0.5	-0.006 (0.007)	0.34
Time spent watching television (TV)	0.02 (0.005)	1.40E-05	0.03 (0.006)	3.46E-06
Time spent using computer	0.02 (0.006)	4.00E-03	0.01 (0.006)	0.011
Time spent driving	0.01 (0.006)	8.00E-03	0.015 (0.006)	9.00E-03
Sleep duration	0.007 (0.005)	0.2	0.007 (0.005)	0.19
Forced vital capacity (FVC)	-0.04 (0.008)	1.10E-08	-0.06 (0.008)	1.88E-12

Forced expiratory volume in 1-second (FEV1)	-0.06 (0.008)	2.30E-16	-0.06 (0.008)	1.32E-16
Peak expiratory flow (PEF)	-0.03 (0.007)	8.30E-07	-0.03 (0.007)	2.25E-05
Heel quantitative ultrasound index (QUI)	-0.04 (0.008)	8.40E-08	-0.03 (0.008)	1.37E-05
Heel bone mineral density (BMD)	-0.04 (0.008)	8.20E-09	-0.03 (0.008)	1.39E-05
Heel Broadband ultrasound attenuation (BUA)	-0.04 (0.008)	1.60E-08	-0.03 (0.008)	2.94E-05
Ankle spacing width	-0.05 (0.009)	8.50E-08	-0.04 (0.01)	2.48E-04
Diastolic blood pressure	-0.003 (0.006)	0.6	0.01 (0.006)	0.08
Systolic blood pressure	0.06 (0.006)	6.10E-21	0.07 (0.006)	1.09E-29
Ankle spacing width (left)	-0.03 (0.011)	8.00E-03	-0.03 (0.013)	0.051
Pulse wave reflection index	-0.02 (0.01)	2.00E-02	-0.02 (0.009)	1.21E-02
Pulse wave peak to peak time	-0.003 (0.01)	0.7	-0.002 (0.01)	0.81
Signal-to-noise-ratio (SNR) of triplet (left)	-0.002 (0.009)	0.9	-0.0009 (0.009)	0.92
Signal-to-noise-ratio (SNR) of triplet (right)	-0.002 (0.009)	0.9	-0.001 (0.009)	0.89
Number of triplets attempted (left)	0.01 (0.009)	0.3	0.01 (0.009)	0.19
Number of triplets attempted (right)	0.01 (0.009)	0.3	0.01 (0.009)	0.21
Maximum digits remembered correctly	0.007 (0.017)	0.7	0.008 (0.016)	0.64
Time to complete test	0.01 (0.016)	0.5	0.007 (0.016)	0.67
Time to answer	0.004 (0.009)	0.7	0.003 (0.009)	0.76
Duration screen displayed	0.002 (0.009)	0.8	0.0003 (0.010)	0.97
Number of attempts	0.0003 (0.009)	0.97	-0.0003 (0.009)	0.98
ECG, heart rate	0.04 (0.014)	2.00E-03	0.05 (0.014)	1.00E-03
ECG, phase time	0.012 (0.013)	0.4	0.01 (0.013)	0.93
ECG, phase duration	0.03 (0.014)	0.04	0.03 (0.014)	0.03
ECG, number of stages in a phase	-0.007 (0.013)	0.6	-0.007 (0.013)	0.61
Maximum workload during fitness test	-0.07 (0.017)	7.90E-05	-0.07 (0.017)	3.32E-05
Maximum heart rate during fitness test	-0.007 (0.014)	0.6	-0.005 (0.014)	0.74
Number of trend entries	-0.04 (0.015)	6.00E-03	-0.04 (0.015)	2.00E-03
Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones	0.05 (0.007)	4.20E-13	0.05 (0.007)	5.46E-14

Weight (pre-imaging)	-0.07 (0.008)	8.20E-17	-0.10 (0.013)	3.72E-13
Height	-0.05 (0.01)	4.00E-07	-0.05 (0.01)	7.22E-08
Ventricular rate	0.03 (0.008)	3.40E-05	0.03 (0.008)	1.09E-05
P duration	-0.009 (0.008)	0.26	-0.007 (0.008)	0.37
QRS duration	-0.020 (0.008)	0.015	-0.020 (0.008)	0.03
Fluid intelligence score	-0.020 (0.009)	0.05	-0.01 (0.009)	0.14
Speech-reception-threshold (SRT) estimate (left)	0.02 (0.01)	0.09	0.01 (0.01)	0.14
Speech-reception-threshold (SRT) estimate (right)	0.02 (0.01)	0.07	0.01 (0.01)	0.16
Birth weight	-0.05 (0.007)	0.44	-0.008 (0.007)	0.21
Mean time to correctly identify matches	0.006 (0.006)	0.29	0.005 (0.006)	0.36
Smoking status	0.13 (0.005)	7.84E-119	-	-
Cascot confidence score	-0.01 (0.02)	0.55	-0.01 (0.02)	0.6
Number of fluid intelligence questions attempted within time limit	-0.003 (0.009)	0.8	0.001 (0.009)	0.9
Number of correct matches in round 1	-0.01 (0.008)	0.1	-0.01 (0.008)	0.17
Number of incorrect matches in round 1	0.006 (0.008)	0.74	0.003 (0.008)	0.65
Time to complete round 1	0.006 (0.008)	0.75	0.004 (0.008)	0.63
Errors before selecting correct item in numeric path (trail 1)	0.05 (0.05)	0.35	0.06 (0.05)	0.43
Errors before selecting correct item in alphanumeric path (trail #2)	-0.04 (0.03)	0.28	-0.04 (0.03)	0.24
Interval between previous point and current one in numeric path (trail #1)	0.008 (0.008)	0.28	0.009 (0.008)	0.26
Forced expiratory volume in 1-second (FEV1)	-0.08 (0.009)	1.23E-20	-0.08 (0.009)	9.94E-19
Forced vital capacity (FVC)	-0.06 (0.009)	8.98E-12	-0.07 (0.009)	3.14E-14
Duration to complete numeric path (trail #1)	0.01 (0.008)	0.087	0.01 (0.008)	0.16
Duration to complete alphanumeric path (trail #2)	0.04 (0.008)	1.19E-05	0.03 (0.008)	8.27E-05
Number of symbol digit matches made correctly	-0.04 (0.008)	4.23E-06	-0.03 (0.008)	1.00E-04
Fluid intelligence score 1	-0.02 (0.008)	0.01	-0.02 (0.008)	0.04
Number of symbol digit matches attempted	-0.04 (0.008)	4.13E-06	-0.03 (0.008)	8.60E-05

Value entered	-0.006 (0.008)	0.41	-0.004 (0.008)	0.61
Duration to entering value	0.007 (0.008)	0.38	0.004 (0.008)	0.57
Maximum digits remembered correctly 1	-0.02 (0.008)	0.035	-0.02 (0.008)	0.021
Total errors traversing numeric path (trail #1)	-0.002 (0.02)	0.92	0.0007 (0.02)	0.97
Total errors traversing numeric path (trail #2)	0.005 (0.01)	0.73	0.004 (0.01)	0.75
Body mass index (BMI)	-0.03 (0.006)	3.64E-10	-	-
Weight	-0.06 (0.006)	6.04E-18	-0.09 (0.014)	2.42E-11
Pulse wave Arterial Stiffness index	0.01 (0.01)	0.27	0.008 (0.01)	0.43
Heterozygosity	0.005 (0.006)	0.34	-0.008 (0.01)	0.4
Heterozygosity, PCA corrected	0.002 (0.006)	0.75	-0.0009 (0.006)	0.86
Genetic principal components	-0.01 (0.006)	0.03	0.018 (0.02)	0.35
Genetic relatedness pairing	-0.02 (0.03)	0.45	-0.02 (0.03)	0.5
Genetic relatedness factor	0.0004 (0.03)	0.99	0.002 (0.03)	0.94
Average X chromosome intensities for determining sex	-0.002 (0.01)	0.89	-0.003 (0.01)	0.76
Average Y chromosome intensities for determining sex	-0.007 (0.02)	0.63	-0.009 (0.02)	0.54
Minimum carotid IMT (intima-medial thickness) at 120 degrees	0.03 (0.02)	0.12	0.03 (0.02)	0.22
Mean carotid IMT (intima-medial thickness) at 120 degrees	0.05 (0.02)	0.03	0.04 (0.02)	0.07
Maximum carotid IMT (intima-medial thickness) at 120 degrees	0.04 (0.02)	0.08	0.03 (0.02)	0.12
Minimum carotid IMT (intima-medial thickness) at 150 degrees	0.02 (0.02)	0.25	0.02 (0.02)	0.34
Mean carotid IMT (intima-medial thickness) at 150 degrees	0.04 (0.02)	0.09	0.03 (0.02)	0.11
Maximum carotid IMT (intima-medial thickness) at 150 degrees	0.01 (0.02)	0.53	0.01 (0.02)	0.52
Minimum carotid IMT (intima-medial thickness) at 210 degrees	0.10 (0.02)	7.18E-05	0.09 (0.02)	1.03E-05
Mean carotid IMT (intima-medial thickness) at 210 degrees	0.10 (0.02)	1.73E-06	0.09 (0.02)	2.45E-05
Maximum carotid IMT (intima-medial thickness) at 210 degrees	0.09 (0.02)	3.35E-05	0.08 (0.02)	3.62E-04

Minimum carotid IMT (intima-medial thickness) at 240 degrees	0.05 (0.02)	0.02	0.04 (0.02)	0.08
Mean carotid IMT (intima-medial thickness) at 240 degrees	0.06 (0.02)	3.00E-03	0.03 (0.02)	0.02
Maximum carotid IMT (intima-medial thickness) at 240 degrees	0.07 (0.02)	1.00E-03	0.06 (0.02)	7.00E-03
Body fat percentage	-0.02 (0.007)	1.00E-03	0.05 (0.01)	4.93E-04
Whole body fat mass	-0.03 (0.006)	1.07E-09	-0.01 (0.02)	0.39
Whole body fat-free mass	-0.10 (0.01)	5.32E-23	-0.10 (0.01)	8.86E-15
Whole body water mass	-0.10 (0.01)	5.58E-23	-0.10 (0.01)	9.84E-15
Basal metabolic rate	-0.09 (0.009)	5.20E-23	-0.10 (0.01)	1.11E-13
Impedance of whole body	0.05 (0.008)	2.72E-12	0.04 (0.009)	9.99E-06
Impedance of leg (right)	0.06 (0.006)	8.07E-20	0.05 (0.008)	5.24E-10
Impedance of leg (left)	0.05 (0.006)	7.86E-18	0.04 (0.007)	1.23E-08
Impedance of arm (right)	0.05 (0.008)	9.81E-09	0.03 (0.01)	4.14E-04
Impedance of arm (left)	0.04 (0.008)	2.00E-07	0.03 (0.01)	3.00E-03
Leg fat percentage (right)	-0.02 (0.01)	0.08	0.10 (0.02)	1.08E-07
Leg fat mass (right)	-0.04 (0.007)	2.15E-08	0.006 (0.02)	0.75
Leg fat-free mass (right)	-0.10 (0.01)	9.36E-28	-0.10 (0.01)	3.60E-15
Leg predicted mass (right)	-0.10 (0.01)	6.93E-25	-0.11 (0.01)	2.85E-15
Leg fat percentage (left)	-0.03 (0.01)	0.019	0.10 (0.02)	1.52E-06
Leg fat mass (left)	-0.04 (0.007)	3.76E-09	-0.06 (0.02)	0.77
Leg fat-free mass (left)	-0.09 (0.01)	1.71E-22	-0.10 (0.01)	1.98E-13
Leg predicted mass (left)	-0.10 (0.01)	1.13E-22	-0.10 (0.01)	1.19E-13
Arm fat percentage (right)	-0.03 (0.008)	2.32E-05	0.05 (0.02)	2.32E-03
Arm fat mass (right)	-0.04 (0.006)	1.92E-11	-0.03 (0.01)	0.07
Arm fat-free mass (right)	-0.10 (0.01)	7.95E-20	-0.10 (0.01)	3.70E-12
Arm predicted mass (right)	-0.10 (0.01)	1.24E-20	-0.10 (0.01)	5.13E-12
Arm fat percentage (left)	-0.04 (0.008)	3.36E-06	0.04 (0.02)	0.014
Arm fat mass (left)	-0.04 (0.006)	1.74E-11	-0.02 (0.01)	0.07
Arm fat-free mass (left)	-0.09 (0.01)	5.45E-18	-0.09 (0.01)	1.60E-10
Arm predicted mass (left)	-0.08 (0.01)	1.70E-16	-0.07 (0.01)	3.54E-09
Trunk fat percentage	-0.02 (0.006)	1.00E-03	0.02 (0.009)	0.012
Trunk fat mass	-0.03 (0.006)	2.50E-09	-0.01 (0.01)	0.36
Trunk fat-free mass	-0.10 (0.01)	3.38E-20	-0.09 (0.01)	3.84E-13

Trunk predicted mass	-0.10 (0.01)	2.32E-20	-0.09 (0.01)	4.17E-13

Table S5: Association of predicted AAC with physiological markers. Univariate regression analysis of risk factors at baseline for predicted AAC after adjusting for age and sex in model 1 and after adjusting for socioeconomic factors, BMI, and smoking status in addition to adjusting for age and sex in model 2. The estimate of the effect size and the standard error of the estimated effect size are given along with p-values for each coefficient in a univariate fit.

CBC Measure	Model 1		Model 2	
	Effect size	P-value	Effect size	P-value
White blood cell (leukocyte) count	0.05 (0.006)	4.82E-20	0.05 (0.006)	1.23E-15
Red blood cell (erythrocyte) count	-0.03 (0.007)	1.90E-05	-0.01 (0.007)	0.07
Haemoglobin concentration	-0.01 (0.007)	0.07	-0.01 (0.007)	0.19
Haematocrit percentage	-0.01 (0.007)	0.15	-0.006 (0.007)	0.4
Mean corpuscular volume	0.03 (0.006)	5.00E-07	0.01 (0.006)	0.05
Mean corpuscular haemoglobin	0.02 (0.006)	8.70E-05	0.007 (0.006)	0.25
Mean corpuscular haemoglobin concentration	-0.004 (0.006)	0.49	-0.006 (0.006)	0.33
Red blood cell (erythrocyte) distribution width	-0.0004 (0.006)	0.95	0.002 (0.006)	0.71
Platelet count	0.02 (0.006)	5.30E-05	0.02 (0.006)	8.61E-05
Platelet crit	0.03 (0.006)	5.00E-06	0.03 (0.006)	1.87E-05
Mean platelet (thrombocyte) volume	-0.001 (0.006)	0.86	-0.003 (0.006)	0.65
Platelet distribution width	0.0003 (0.006)	0.96	0.002 (0.006)	0.67
Lymphocyte count	0.02 (0.006)	1.00E-03	0.02 (0.006)	4.00E-03
Monocyte count	0.04 (0.006)	1.41E-11	0.04 (0.006)	8.97E-11
Neutrophill count	0.05 (0.006)	4.57E-19	0.04 (0.006)	1.03E-14
Eosinophill count	0.02 (0.006)	1.74E-04	0.02 (0.006)	1.00E-03
Basophill count	0.01 (0.006)	0.03	0.0085 (0.006)	0.13
Nucleated red blood cell count	0.002 (0.006)	0.72	0.001 (0.006)	0.84
Lymphocyte percentage	-0.02 (0.006)	4.00E-03	-0.015 (0.006)	0.01
Monocyte percentage	0.002 (0.006)	0.75	0.005 (0.006)	0.4
Neutrophill percentage	0.01 (0.006)	0.035	0.01 (0.006)	0.08

Eosinophill percentage	0.004 (0.006)	0.43	0.004 (0.006)	0.47
Basophill percentage	0.008 (0.006)	0.16	0.005 (0.006)	0.36
Nucleated red blood cell percentage	0.0004 (0.006)	0.95	-0.0003 (0.006)	0.95
Reticulocyte percentage	0.02 (0.006)	6.00E-03	0.02 (0.006)	7.54E-05
Reticulocyte count	0.01 (0.006)	0.03	0.02 (0.006)	1.76E-04
Mean reticulocyte volume	0.05 (0.006)	0.42	-0.003 (0.006)	0.61
Mean sphered cell volume	0.02 (0.006)	4.00E-03	-0.0006 (0.006)	0.91
Immature reticulocyte fraction	0.02 (0.006)	5.90E-05	0.03 (0.006)	8.62E-08
High light scatter reticulocyte percentage	0.02 (0.006)	5.50E-05	0.04 (0.006)	2.87E-09
High light scatter reticulocyte count	0.02 (0.006)	4.53E-04	0.04 (0.006)	5.74E-09

Table S6: Association of predicted AAC with complete blood count markers. Univariate regression analysis of risk factors at baseline for predicted AAC after adjusting for age and sex in model 1 and after adjusting for socioeconomic factors, BMI, and smoking status in addition to adjusting for age and sex in model 2. The estimate of the effect size and the standard error of the estimated effect size are given along with p-values for each coefficient in a univariate fit.

Genetics supplementary tables (1-9)

References

Azur, Melissa J., Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. 2011. “Multiple Imputation by Chained Equations: What Is It and How Does It Work?” *International Journal of Methods in Psychiatric Research* 20 (1): 40–49.

Bankhead, Peter, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D. Dunne, Stephen McQuaid, et al. 2017. “QuPath: Open Source Software for Digital Pathology Image Analysis.” *Scientific Reports* 7 (1): 16878.

Bie, M. K. de, M. S. Buiten, J. I. Rotmans, M. Hogenbirk, M. J. Schalij, T. J. Rabelink, and J. W. Jukema. 2017. “Abdominal Aortic Calcification on a Plain X-Ray and the Relation with Significant Coronary Artery Disease in Asymptomatic Chronic Dialysis Patients.” *BMC Nephrology* 18 (1): 82.

Bulik-Sullivan, Brendan, Hilary K. Finucane, Verneri Anttila, Alexander Gusev, Felix R. Day, Po-Ru Loh, ReproGen Consortium, et al. 2015. “An Atlas of Genetic Correlations across Human Diseases and Traits.” *Nature Genetics* 47 (11): 1236–41.

Bulik-Sullivan, Brendan K., Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. 2015. “LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies.” *Nature Genetics* 47 (3): 291–95.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, et al. 2018. “The UK Biobank Resource with Deep Phenotyping and Genomic Data.” *Nature* 562 (7726): 203–9.

Cawthon, Peggy M., Mohammad Shahnazari, Eric S. Orwoll, and Nancy E. Lane. 2016. “Osteoporosis in Men: Findings from the Osteoporotic Fractures in Men Study (MrOS).” *Therapeutic Advances in Musculoskeletal Disease* 8 (1): 15–27.

Cox, David Roxbee. 2018. *Analysis of Survival Data*. Chapman and Hall/CRC.

Denny, Joshua C., Marylyn D. Ritchie, Melissa A. Basford, Jill M. Pulley, Lisa Bastarache, Kristin Brown-Gentry, Deede Wang, Dan R. Masys, Dan M. Roden, and Dana C. Crawford. 2010. “PheWAS: Demonstrating the Feasibility of a Phenome-Wide Scan to Discover Gene-Disease Associations.” *Bioinformatics* 26 (9): 1205–10.

ENCODE Project Consortium. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489 (7414): 57–74.

Fan, G., H. Liu, Z. Wu, Y. Li, C. Feng, D. Wang, J. Luo, W. M. Wells 3rd, and S. He. 2019. “Deep Learning-Based Automatic Segmentation of Lumbosacral Nerves on CT for Spinal Intervention: A Translational Study.” *AJNR. American Journal of Neuroradiology* 40 (6): 1074–81.

Fehrmann, Rudolf S. N., Juha M. Karjalainen, Małgorzata Krajewska, Harm-Jan Westra, David Maloney, Anton Simeonov, Tune H. Pers, et al. 2015. “Gene Expression Analysis Identifies Global Gene Dosage Sensitivity in Cancer.” *Nature Genetics* 47 (2): 115–25.

Finucane, Hilary K., Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, et al. 2015. “Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics.” *Nature Genetics* 47 (11): 1228–35.

Finucane, Hilary K., Yakir A. Reshef, Verneri Anttila, Kamil Slowikowski, Alexander Gusev, Andrea Byrnes, Steven Gazal, et al. 2018. “Heritability Enrichment of Specifically

Expressed Genes Identifies Disease-Relevant Tissues and Cell Types." *Nature Genetics* 50 (4): 621–29.
Giambartolomei, Claudia, Damjan Vukcevic, Eric E. Schadt, Lude Franke, Aroon D. Hingorani, Chris Wallace, and Vincent Plagnol. 2014. "Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics." *PLoS Genetics* 10 (5): e1004383.
Grubb, Anders, Masaru Horio, Lars-Olof Hansson, Jonas Björk, Ulf Nyman, Mats Flodin, Anders Larsson, et al. 2014. "Generation of a New Cystatin C-Based Estimating Equation for Glomerular Filtration Rate by Use of 7 Assays Standardized to the International Calibrator." *Clinical Chemistry* 60 (7): 974–86.
GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, et al. 2017. "Genetic Effects on Gene Expression across Human Tissues." *Nature* 550 (7675): 204–13.
Guo, Hui, Mary D. Fortune, Oliver S. Burren, Ellen Schofield, John A. Todd, and Chris Wallace. 2015. "Integration of Disease Association and eQTL Data Using a Bayesian Colocalisation Approach Highlights Six Candidate Causal Genes in Immune-Mediated Diseases." *Human Molecular Genetics* 24 (12): 3305–13.
Han, Zhongyi, Benzheng Wei, Ashley Mercado, Stephanie Leung, and Shuo Li. 2018. "Spine-GAN: Semantic Segmentation of Multiple Spinal Structures." *Medical Image Analysis* 50 (December): 23–35.
He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Deep Residual Learning for Image Recognition." *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1512.03385.
Heng, Tracy S. P., Michio W. Painter, and Immunological Genome Project Consortium. 2008. "The Immunological Genome Project: Networks of Gene Expression in Immune Cells." *Nature Immunology* 9 (10): 1091–94.
Hoffmann, Thomas J., Georg B. Ehret, Priyanka Nandakumar, Dilrini Ranatunga, Catherine Schaefer, Pui-Yan Kwok, Carlos Iribarren, Aravinda Chakravarti, and Neil Risch. 2017. "Genome-Wide Association Analyses Using Electronic Health Records Identify New Loci Influencing Blood Pressure Variation." *Nature Genetics* 49 (1): 54–64.
International HapMap 3 Consortium, David M. Altshuler, Richard A. Gibbs, Leena Peltonen, David M. Altshuler, Richard A. Gibbs, Leena Peltonen, et al. 2010. "Integrating Common and Rare Genetic Variation in Diverse Human Populations." *Nature* 467 (7311): 52–58.
Kalluri, Aditya S., Shamsudheen K. Vellarikkal, Elazer R. Edelman, Lan Nguyen, Ayshwarya Subramanian, Patrick T. Ellinor, Aviv Regev, Sekar Kathiresan, and Rajat M. Gupta. 2019. "Single-Cell Analysis of the Normal Mouse Aorta Reveals Functionally Distinct Endothelial Cell Populations." *Circulation* 140 (2): 147–63.
Kauppila, L. I., J. F. Polak, L. A. Cupples, M. T. Hannan, D. P. Kiel, and P. W. Wilson. 1997. "New Indices to Classify Location, Severity and Progression of Calcific Lesions in the Abdominal Aorta: A 25-Year Follow-up Study." *Atherosclerosis* 132 (2): 245–50.
Lessmann, Nikolas, Bram van Ginneken, Pim A. de Jong, and Ivana Išgum. 2019. "Iterative Fully Convolutional Neural Networks for Automatic Vertebra Segmentation and Identification." *Medical Image Analysis* 53 (April): 142–55.
Loh, Po-Ru, George Tucker, Brendan K. Bulik-Sullivan, Bjarni J. Vilhjálmsson, Hilary K. Finucane, Rany M. Salem, Daniel I. Chasman, et al. 2015. "Efficient Bayesian Mixed-Model Analysis Increases Association Power in Large Cohorts." *Nature Genetics* 47 (3): 284–90.
Lukowski, Samuel W., Jatin Patel, Stacey B. Andersen, Seen-Ling Sim, Ho Yi Wong, Joshua Tay, Ingrid Winkler, Joseph E. Powell, and Kiarash Khosrotehrani. 2019. "Single-Cell

Transcriptional Profiling of Aortic Endothelium Identifies a Hierarchy from Endovascular Progenitors to Differentiated Cells." *Cell Reports* 27 (9): 2748–58.e3.
Millett, Elizabeth R. C., Sanne A. E. Peters, and Mark Woodward. 2018. "Sex Differences in Risk Factors for Myocardial Infarction: Cohort Study of UK Biobank Participants." *BMJ* 363 (November): k4247.
Nikpay, Majid, Anuj Goel, Hong-Hee Won, Leanne M. Hall, Christina Willenborg, Stavroula Kanoni, Danish Saleheen, et al. 2015. "A Comprehensive 1,000 Genomes-Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease." *Nature Genetics* 47 (10): 1121–30.
Nissen, Steven E., E. Murat Tuzcu, Paul Schoenhagen, Tim Crowe, William J. Sasiela, John Tsai, John Orazem, et al. 2005. "Statin Therapy, LDL Cholesterol, C-Reactive Protein, and Coronary Artery Disease." *The New England Journal of Medicine* 352 (1): 29–38.
Pers, Tune H., Juha M. Karjalainen, Yingleong Chan, Harm-Jan Westra, Andrew R. Wood, Jian Yang, Julian C. Lui, et al. 2015. "Biological Interpretation of Genome-Wide Association Studies Using Predicted Gene Functions." *Nature Communications* 6 (January): 5890.
Petersen, Steffen E., Paul M. Matthews, Fabian Bamberg, David A. Bluemke, Jane M. Francis, Matthias G. Friedrich, Paul Leeson, et al. 2013. "Imaging in Population Science: Cardiovascular Magnetic Resonance in 100,000 Participants of UK Biobank - Rationale, Challenges and Approaches." *Journal of Cardiovascular Magnetic Resonance: Official Journal of the Society for Cardiovascular Magnetic Resonance* 15 (May): 46.
Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30.
Seabold, S., and J. Perktold. 2010. "Statsmodels: Econometric and Statistical Modeling with Python." *Of the 9th Python in Science Conference*. https://www.researchgate.net/profile/Josef_Perktold/publication/264891066_Statsmodels_Econometric_and_Statistical_Modeling_with_Python/links/5667ca9308ae34c89a0261a8/Statsmodels-Econometric-and-Statistical-Modeling-with-Python.pdf.
Storey, John D., and Robert Tibshirani. 2003. "Statistical Significance for Genomewide Studies." *Proceedings of the National Academy of Sciences of the United States of America* 100 (16): 9440–45.
Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Medicine* 12 (3): e1001779.
Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. "Rethinking the Inception Architecture for Computer Vision." *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1512.00567.
Szulc, Pawel, Terri Blackwell, John T. Schousboe, Douglas C. Bauer, Peggy Cawthon, Nancy E. Lane, Steven R. Cummings, Eric S. Orwoll, Dennis M. Black, and Kristine E. Ensrud. 2014. "High Hip Fracture Risk in Men with Severe Aortic Calcification: MrOS Study." *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research* 29 (4): 968–75.
Wakefield, Jon. 2007. "A Bayesian Measure of the Probability of False Discovery in Genetic Epidemiology Studies." *American Journal of Human Genetics* 81 (2): 208–27.
Wellcome Trust Case Control Consortium, Julian B. Maller, Gilean McVean, Jake Byrnes, Damjan Vukcevic, Kimmo Palin, Zhan Su, et al. 2012. "Bayesian Refinement of Association Signals for 14 Loci in 3 Common Diseases." *Nature Genetics* 44 (12): 1294–1301.

Yang, Jian, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, et al. 2010. “Common SNPs Explain a Large Proportion of the Heritability for Human Height.” *Nature Genetics* 42 (7): 565–69.

Yang, Jian, Teresa Ferreira, Andrew P. Morris, Sarah E. Medland, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Pamela A. F. Madden, et al. 2012. “Conditional and Joint Multiple-SNP Analysis of GWAS Summary Statistics Identifies Additional Variants Influencing Complex Traits.” *Nature Genetics* 44 (4): 369–75, S1–3.

Zheng, Jie, A. Mesut Erzurumluoglu, Benjamin L. Elsworth, John P. Kemp, Laurence Howe, Philip C. Haycock, Gibran Hemani, et al. 2017. “LD Hub: A Centralized Database and Web Interface to Perform LD Score Regression That Maximizes the Potential of Summary Level GWAS Data for SNP Heritability and Genetic Correlation Analysis.” *Bioinformatics* 33 (2): 272–79.

Zhou, Wei, Jonas B. Nielsen, Lars G. Fritsche, Rounak Dey, Maiken E. Gabrielsen, Brooke N. Wolford, Jonathon LeFaive, et al. 2018. “Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies.” *Nature Genetics* 50 (9): 1335–41.