SUPPLEMENTAL MATERIAL

for the manuscript

COL4A1 **Is Associated With Arterial Stiffness By Genome Wide Association Scan.**

Tarasov KV et al.

Introduction

Founder population

In population genetics, founder effect refers to the establishment of a new colony by a very small number of individuals from a larger population, often many hundreds or thousands of years ago. When such founder groups expand to large modern populations without substantial in-migration, they facilitate detailed genealogical research and reconstruction of extended multigenerational pedigrees; and the relatedness and relative environmental homogeneity make it easier to determine the alleles that contribute to the values of a complex phenotype or to a disease. $1-3$

Methods

Reproducibility of PWV in SardiNIA

Each measurement of PWV was performed by 2 sonographers, with one recording the waveforms at the carotid site, and the other simultaneously recording the waveforms at the femoral site. Because several sonographers participated in the collection of the PWV data, a reproducibility study was performed on 3 subjects who underwent repeated measurements of PWV using random combinations of testers (there were 5 testers at the carotid site and 3 testers at the femoral site). Each subject underwent approximately 45 measurements of PWV. The coefficient of variation for PWV ranged from 7.2 to 10.6%. Both general linear models as well as mixed effects models showed that only inter-subject differences significantly contributed to the overall error in the measurement of PWV, whereas carotid testers, femoral testers, and the interaction between carotid and femoral testers did not.

Imputations

For individuals who were genotyped with the 500k chip, the SNP being tested is coded as 0, 1, or 2, depending on the number of copies carried of an arbitrary reference allele. If this SNP was not included in the 10K chip, than for individuals with missing genotype data (i.e. those genotyped only with the 10k chip), the Lander-Green algorithm was used to estimate the number of copies of this allele that are being carried, which was assigned a score ranging between 0 and $2⁴$. Importantly, this score represents a probabilistic estimate of the number of copies of the allele, and it incorporates information on allele frequency, the genotype of relatives for the SNP of interest, their degree of relatedness, and data on flanking markers. Thus, this score need not be an ordinal value of 0, 1 or 2, but can be a fractional number that ranges between 0 and 2, which enables uncertain genotypes to be accounted for. For computational efficiency, the Lander-Green algorithm was applied to sub-pedigrees, each including no more than 20-25 individuals, resulting in a dataset where the average analysis unit consisted of a family with 12.3 members and 3.2 generations.

Kinship coefficient:

Kinship coefficient in the SardiNIA study was determined on the basis of family structure, not genotype data. Family structure, in turn, was determined on the basis of self-reported information, which allowed us to construct family genealogy trees that were up to five generations deep. However, analytic constraints required us to ignore any relatedness that was not observed within these 5 generations. All the first and second degree relationships were verified based on the genotype data using GRR software⁴, and adjusted if necessary.

Comparison of PWV protocols in SardiNIA and HAPPI Heart:

The device used to measure PWV in HAPI Heart was the Complior[®] SP device (Artech Medical, Pantin, France), which differs than the custom-designed device that was used in SardiNIA. However, both devices have been validated^{5, 6}, and PWV measured with either of these devices has been shown to be an independent predictor of mortality in large epidemiological studies^{7, 8}. Importantly, since the raw PWV data from the 2 studies were not combined, rather the results were combined in the meta-analysis; and since the adjustment for family structure and the data transformation and the statistical analyses in HAPI Heart were performed in a manner that was analogous to the way they were performed in SardiNIA, we don't feel that the results of the genetic analyses could be an artifact of the minor methodological differences that arise from the use of two different devices. In fact, several meta-analyses have recently been performed for many other quantitative traits $3,9-11$, and the

combined results have consistently pointed to genes that were associated with the trait in most of the cohorts in spite of differences in the devices or methodologies used to perform the measurements.

Discussion

Collagen Type 4

Type 4 Collagen had not previously been considered to be involved in regulating arterial stiffness. Unlike Collagen types 1 and 3, which are constituents of the extracellular matrix that are found in the medial layer of the arterial walls where they impart the tensile strength to the arteries, type 4 Collagen is a structural component of basement membranes. Basement membranes surround vascular smooth muscle cells in the media. During the aging process within arteries, smooth muscle cells manufacture, secrete and activate type 2 matrix metalloprotease (MMPII), a collagenase whose substrate is type 4 Collagen. Degradation of type 4 Collagen in the basement membranes permits these cells to invade the internal elastic membrane and enter the subendothelial space. These cells then proliferate and secrete matrix, resulting in a diffusely thickened intima that interferes with endothelial function and also alters the transduction of mechanical forces imparted by flow and pressure of the blood¹². The *COL4A1* polymorphism described in the present study could impact on age-associated arterial stiffening by affecting the binding of MMPII to collagen 4, its substrate. Alternatively, this polymorphism could affect the polymerization of the Collagen 4α1 molecules not only rendering basement membranes more susceptible to metalloprotease

digestion but also differentially affecting formation of the advanced glycation end-products which are known to influence the assembly of type 4 Collagen¹³ and to modulate arterial stiffness.

Beyond its role as a structural component of the basement membrane, type 4 Collagen also provides structural support and anchorage for cells, serves as a ligand for cell surface receptors, modulates endothelial cell proliferation, and regulates angiogenesis and tumor growth. Thus, allelic variants of *COL4A1* could differentially affect the permeability of the basement membrane, or the cell-matrix and/or cell-basement membrane interactions of the endothelial or smooth muscle cells in the arterial wall. In this context, it is noteworthy that the *MAGI1* gene product also plays a role in cell-cell interactions, as a scaffolding protein at cell-cell junctions and as a mediator of vascular endothelial cell adhesions¹⁴.

Finally, allelic variants could differentially influence the type or quantity of growth factors secreted by these cells (e.g. $TGF-\beta$), which, in turn, would influence signaling cascades e.g. SMAD signaling, initiated by TGF-β receptor activation, which leads to increased production of collagen I and III matrix proteins known to directly affect arterial stiffness.

We should note that at the present time there is no functional evidence implicating *COL4A1* as a determinant of PWV. Therefore the foregoing discussion of its putative role in influencing arterial stiffness remains speculative. Furthermore, by using standard prediction algorithms we could not find any evidence that the Glycine to Histidine substitution at position 1334 that is associated with the rs3742207 SNP leads to any significant changes in the 3-dimensional views of protein structure or domain interactions, even though this replacement could lead to changes in the local charge.

Supplementary Table S1.

The top 100 SNPs associated with PWV in the initial GWAS in the SardiNIA cohort.

*- adjusted p-values according to the genomic control method (Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999;55:997-1004

SUPPLEMENTAL FIGURES

Original Phenotype

Transformed Phenotype

Figure S1

Figure S2

Figure S3

Figure Legends

Figure S1: Frequency distribution of PWV values before and after inverse normal transformation. The raw values have a skewed distribution. The transformed values appear normally distributed.

Figure S2: Quantile-quantile plot of SNPs associated with PWV in the SardiNIA study after adjusting the p-values according to the genomic control method¹⁵. Red symbols represent all tested SNPs (N= 362,129) in the GWA scan. The gray area corresponds to the 90% confidence region from a null distribution of p-values (generated from 100 simulations).

Figure S3. High resolution linkage disequilibrium plot and LD blocks in the Sardinian (top panel) and Old Order Amish (bottom panel) populations around SNP rs3742207 (highlighted with a circle) from Haploview¹⁶. LD coloring scheme is the standard D'/LOD scheme, while numbers shown within squares are r^2 values. Genotyped SNPs in the region slightly differ between the two cohorts, but clearly define the length of the second haplotype block, delimited by SNPs rs3742207 and rs11069830, and of identical length in the HapMap CEU population. Squared SNPs are those analyzed by the Framingham study. No other SNP analyzed by Framingham falls in this region.

SUPPLEMENTAL REFERENCES

- **1.** Li S, Sanna S, Maschio A, Busonero F, Usala G, Mulas A, Lai S, Dei M, Orru M, Albai G, Bandinelli S, Schlessinger D, Lakatta E, Scuteri A, Najjar SS, Guralnik J, Naitza S, Crisponi L, Cao A, Abecasis G, Ferrucci L, Uda M, Chen WM, Nagaraja R. The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet.* 2007;3(11):e194.
- **2.** Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orru M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR. Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet.* 2007;3(7):e115.
- **3.** Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* 2008;40(2):161-169.
- **4.** Abecasis GR, Cherny SS, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics.* 2001;17(8):742-743.
- **5.** Sutton-Tyrrell K, Mackey RH, Holubkov R, Vaitkevicius PV, Spurgeon HA, Lakatta EG. Measurement variation of aortic pulse wave velocity in the elderly. *Am J Hypertens.* 2001;14(5 Pt 1):463-468.
- **6.** Asmar R, Benetos A, Topouchian J, Laurent P, Pannier B, Brisac AM, Target R, Levy BI. Assessment of arterial distensibility by automatic pulse wave velocity measurement. Validation and clinical application studies. *Hypertension.* 1995;26(3):485-490.
- **7.** Mattace-Raso FU, van der Cammen TJ, Hofman A, van Popele NM, Bos ML, Schalekamp MA, Asmar R, Reneman RS, Hoeks AP, Breteler MM, Witteman JC. Arterial stiffness and risk of coronary heart disease and stroke: the Rotterdam Study. *Circulation.* 2006;113(5):657-663.
- **8.** Sutton-Tyrrell K, Najjar SS, Boudreau RM, Venkitachalam L, Kupelian V, Simonsick EM, Havlik R, Lakatta EG, Spurgeon H, Kritchevsky S, Pahor M, Bauer D, Newman A. Elevated aortic pulse wave velocity, a marker of arterial stiffness, predicts cardiovascular events in well-functioning older adults. *Circulation.* 2005;111(25):3384-3390.
- **9.** Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, Shen H, Timpson N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L, Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN, Watanabe RM, Uda M, Tuomilehto J, Coresh J, Hirschhorn JN, Shuldiner AR, Schlessinger D, Collins FS, Davey Smith G, Boerwinkle E, Cao A, Boehnke M, Abecasis GR, Mohlke KL. Common variants in

the GDF5-UQCC region are associated with variation in human height. *Nat Genet.* 2008;40(2):198-203.

- **10.** Chen WM, Erdos MR, Jackson AU, Saxena R, Sanna S, Silver KD, Timpson NJ, Hansen T, Orru M, Grazia Piras M, Bonnycastle LL, Willer CJ, Lyssenko V, Shen H, Kuusisto J, Ebrahim S, Sestu N, Duren WL, Spada MC, Stringham HM, Scott LJ, Olla N, Swift AJ, Najjar S, Mitchell BD, Lawlor DA, Smith GD, Ben-Shlomo Y, Andersen G, Borch-Johnsen K, Jorgensen T, Saramies J, Valle TT, Buchanan TA, Shuldiner AR, Lakatta E, Bergman RN, Uda M, Tuomilehto J, Pedersen O, Cao A, Groop L, Mohlke KL, Laakso M, Schlessinger D, Collins FS, Altshuler D, Abecasis GR, Boehnke M, Scuteri A, Watanabe RM. Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. *J Clin Invest.* 2008;118(7):2620-2628.
- **11.** Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, Berndt SI, Jacobs KB, Chanock SJ, Hayes RB, Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS, Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghori J, Groves CJ, Gwilliam R, Hadley D, Hall AS, Hattersley AT, Hebebrand J, Heid IM, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T, Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC, Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L, Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA, Song K, Soranzo N, Speliotes EK, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M, Vogel CI, Wallace C, Waterworth DM, Weedon MN, Willer CJ, Wraight, Yuan X, Zeggini E, Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM, Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso I, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Lyon HN, Voight BF, Ridderstrale M, Groop L, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Jackson AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008;40(6):768-775.
- **12.** Wang M, Lakatta E. Central arterial aging: humans to molecules hypertension in the elderly. In: Safar ME ORM, ed. *Handbook of Hypertension.* Vol 23. Amsterdam, The Netherlands: Elsevier; 2006:137–160.
- **13.** Tsilibary EC, Charonis AS, Reger LA, Wohlhueter RM, Furcht LT. The effect of nonenzymatic glucosylation on the binding of the main noncollagenous NC1 domain to type IV collagen. *J Biol Chem.* 1988;263(9):4302-4308.
- **14.** Sakurai A, Fukuhara S, Yamagishi A, Sako K, Kamioka Y, Masuda M, Nakaoka Y, Mochizuki N. MAGI-1 is required for Rap1 activation upon cell-cell contact and for enhancement of vascular endothelial cadherin-mediated cell adhesion. *Mol Biol Cell.* 2006;17(2):966-976.
- **15.** Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999;55(4):997-1004.
- **16.** Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263-265.