

Polygenic scores for obstructive sleep apnea based on BMI- adjusted  
and -unadjusted genetic associations reveal pathways  
contributing to cardiovascular disease:  
Supplementary Information

Supplementary Figures.....4

    Supplementary Figure 1: Effect estimates from association analysis of PRS-CS based OSA PGSs (BMI-adjusted GWAS) comparing LD reference panel. ....4

    Supplementary Figure 2: Effect estimates from association analysis of PRS-CS based OSA PGSs (BMI-unadjusted GWAS) comparing LD reference panel. ....5

    Supplementary Figure 3: Comparison of ancestry-specific OSA PGSs developed by PRS-CS and LDPred2 (BMI-adjusted analysis).....6

    Supplementary Figure 4: Comparison of ancestry-specific OSA PGSs developed by PRS-CS and LDPred2 (BMI-unadjusted analysis).....7

    Supplementary Figure 5: Comparison of multi-population PGSs (BMI-adjusted GWAS) .....8

    Supplementary Figure 6: Comparison of multi-population PGSs (BMI-unadjusted GWAS) .....9

    Supplementary Figure 7: comparison of estimated OSA associations with sex-specific and sex-combined OSA PGSs (BMI-adjusted). ....10

    Supplementary Figure 8: comparison of estimated OSA associations with sex-specific and sex-combined OSA PGSs (BMI-unadjusted). ....11

    Supplementary Figure 9: Estimated associations of OSA PGSs OSA in strata defined by sex (MyCode) and by sex, age, and obesity (All of Us). ....12

    Supplementary Figure 10: Estimated associations of OSA PGSs with oxygen saturation phenotypes in the Human Phenotypes Project.....13

    Supplementary Figure 11: Estimated associations of OSA PGSs with sleep stage measures in the Human Phenotypes Project.....14

    Supplementary Figure 12: Estimated associations of OSA PGSs with self-reported sleep phenotypes in the Human Phenotypes Project.....15

    Supplementary Figure 13: Sex-stratified associations of OSA PGSs with clinical outcomes in All of Us .....17

    Supplementary Figure 14: BMIunadjOSA PGS associations with DXA measures in the Human Phenotype Project .....18

Supplementary Tables .....19

    Supplementary Table 1: Characteristics of individuals participating in OSA PGS association analyses in TOPMed. ....19

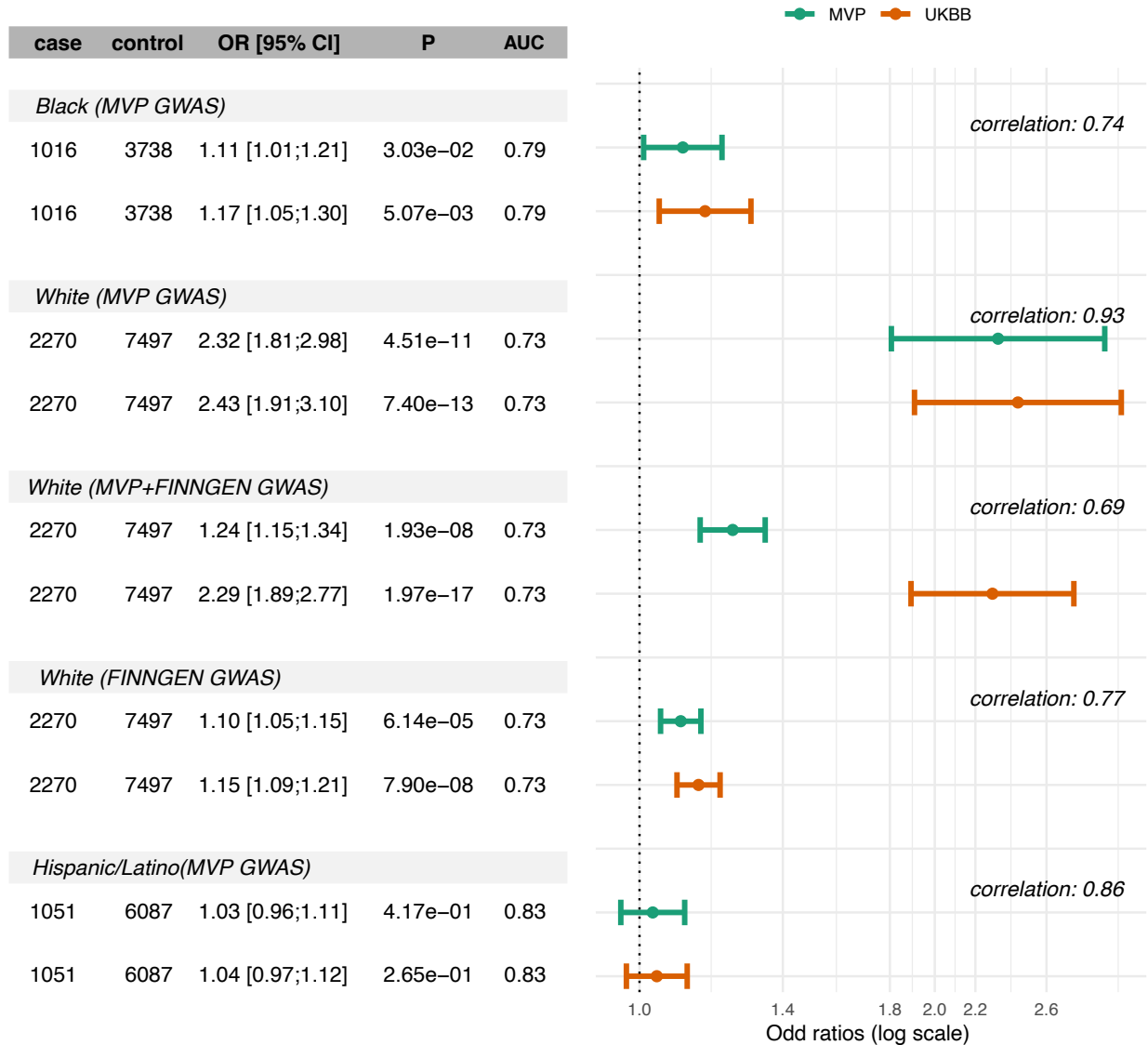
    Supplementary Table 2: Characteristics of individuals participating in OSA analysis in the MGB Biobank.....20

Supplementary Table 3: TOPMed-evaluated mean, SDs, and combination weights, of PRS-CS based PGSs used in the final BMIadjOSA PGS and BMIunadjOSA PGS.....	21
Supplementary Table 4: OSA assessment methods in participating studies.....	22
Supplementary Table 5. Participants characteristics in analyses of OSA-related phenotypes (not REM/NREM stratified). .....	24
Supplementary Table 6. Participants characteristics in analyses of REM/NREM stratified OSA-related phenotypes. ....	25
Supplementary Table 7: Association of FEV1/FVC PGS with OSA continuous phenotypes.....	26
Supplementary Table 8: Geisinger’s MyCode Community Health Initiative (MyCode) Study health system participating patient characteristics. ....	27
Supplementary Table 9: Human Phenotypes Project participant characteristics. ....	28
Supplementary Table 10: Human Phenotypes Project sleep monitoring participant characteristics.....	29
Supplementary Table 11: Characteristics of Human Phenotype Projects participants and their self-reported sleep phenotypes. ....	30
Supplementary Table 12: Characteristics of individuals participating in All of Us association analyses. ....	32
Supplementary Table 13: Comparison of newly developed and existing OSA PGS associations with OSA in All of Us. ....	33
Supplementary Table 14: Characteristics of DXA measures in the Human Phenotype Project	34
Supplementary Table 15: FinnGen participant characteristics.....	35
Supplementary Table 16: TOPMed studies contributing to OSA PGS association with each of the non-OSA sleep phenotypes .....	36
Supplementary Note 1: Choosing optimal OSA PGSs in analysis in TOPMed.....	37
Using UK Biobank as a reference panel is sufficient for developing PGS based on MVP summary statistics.....	37
Identifying best performing single-ancestry PGSs.....	37
Multi-ethnic GWAS PGS versus a multi-PGS combining ancestry-specific PGSs .....	38
Supplementary Note 2: MGB biobank.....	38
Supplementary Note 3: TOPMed parent study descriptions.....	41
ARIC .....	41
CARDIA .....	42
CFS.....	43
CHS .....	44
COPDGene .....	45

FHS.....	46
HCHS/SOL .....	46
JHS .....	48
MESA .....	49
WHI.....	51
Supplementary Note 4: Whole genome sequencing and quality control in TOPMed.....	51
Supplementary Note 5: All of Us phenotypes and acknowledgements .....	52
Supplementary Note 6: FinnGen full ethics statement and acknowledgement.....	54
Supplementary Note 7: TOPMed and CCDG acknowledgments .....	57
Supplementary Note 8: TOPMed consortium members .....	59
Supplementary Note 9: FinnGen consortium members.....	68
References.....	71

## Supplementary Figures

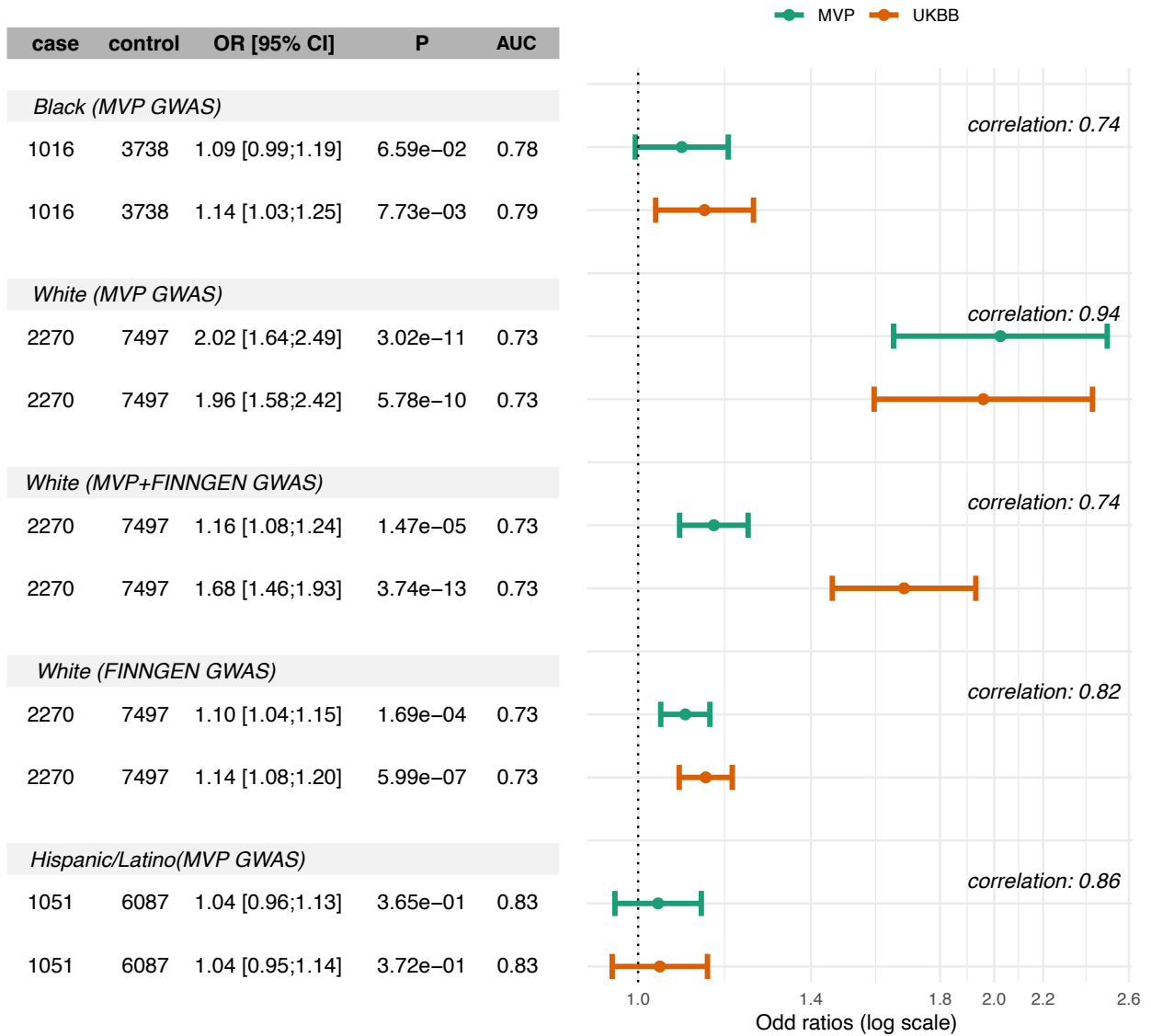
Supplementary Figure 1: Effect estimates from association analysis of PRS-CS based OSA PGSs (BMI-adjusted GWAS) comparing LD reference panel.



OSA was defined as  $AHI/REI \geq 15$ .

AHI: apnea hypopnea index; OSA: obstructive sleep apnea; REI: respiratory event index.

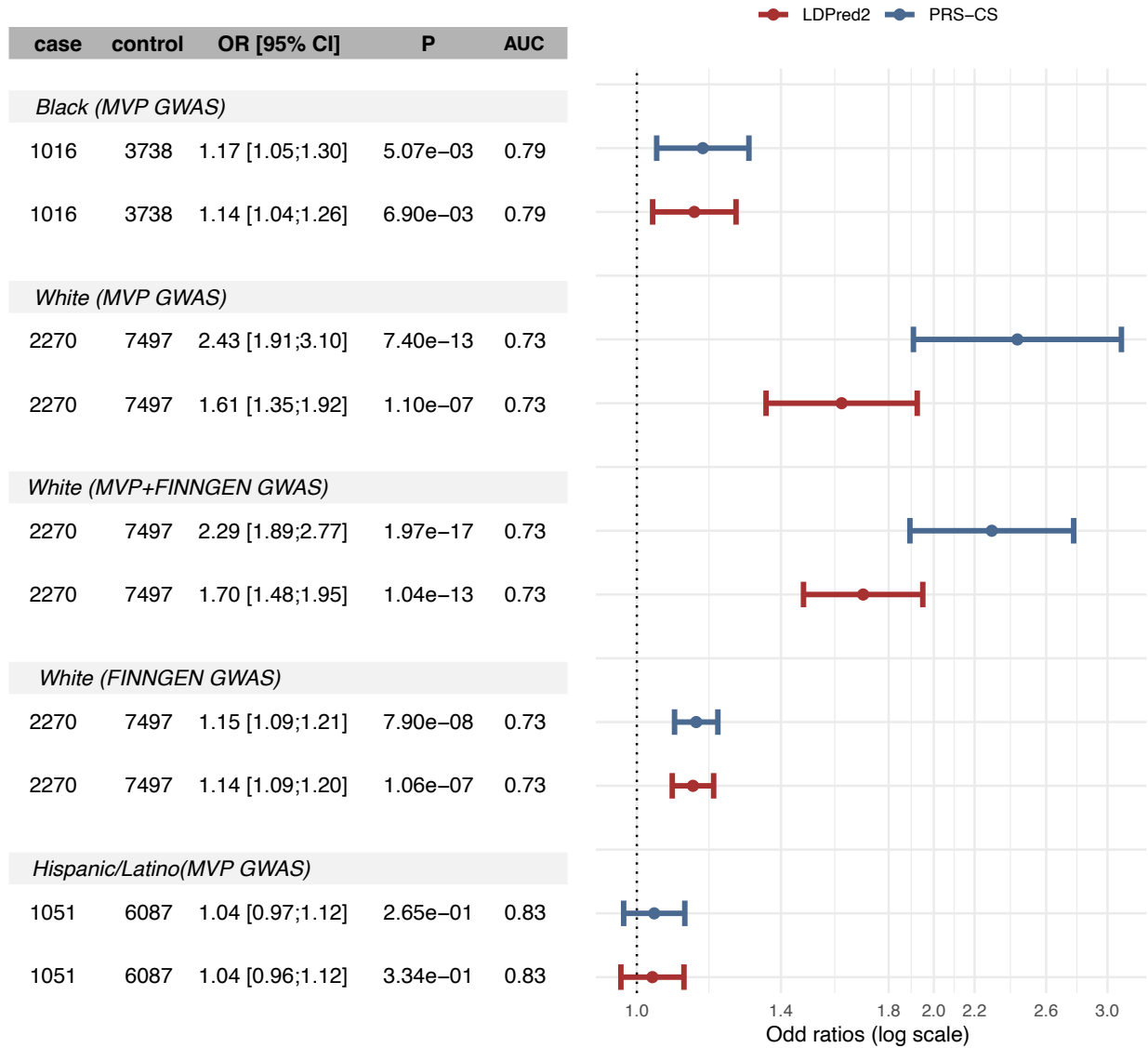
Supplementary Figure 2: Effect estimates from association analysis of PRS-CS based OSA PGSs (BMI-unadjusted GWAS) comparing LD reference panel.



OSA was defined as  $AHI/REI \geq 15$ .

AHI: apnea hypopnea index; OSA: obstructive sleep apnea; REI: respiratory event index.

Supplementary Figure 3: Comparison of ancestry-specific OSA PGSs developed by PRS-CS and LDPreD2 (BMI-adjusted analysis)

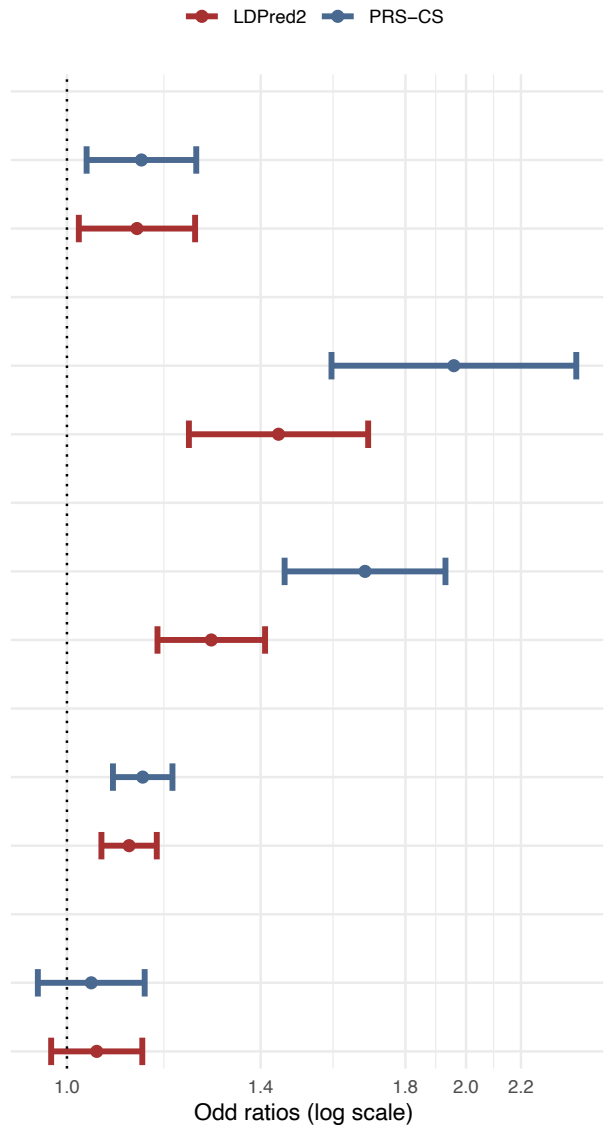


OSA was defined as AHI/REI $\geq$ 15.

AHI: apnea hypopnea index; OSA: obstructive sleep apnea; REI: respiratory event index.

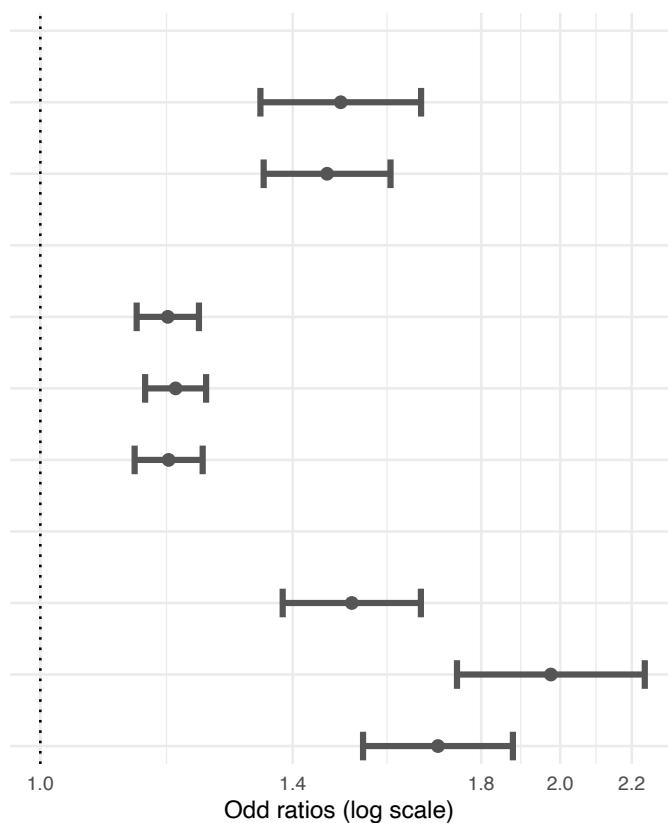
Supplementary Figure 4: Comparison of ancestry-specific OSA PGSs developed by PRS-CS and LDPreD2 (BMI-unadjusted analysis)

case	control	OR [95% CI]	P	AUC
<i>Black (MVP GWAS)</i>				
1016	3738	1.14 [1.03;1.25]	7.73e-03	0.79
1016	3738	1.13 [1.02;1.25]	1.83e-02	0.79
<i>White (MVP GWAS)</i>				
2270	7497	1.96 [1.58;2.42]	5.78e-10	0.73
2270	7497	1.44 [1.24;1.69]	3.70e-06	0.73
<i>White (MVP+FINNGEN GWAS)</i>				
2270	7497	1.68 [1.46;1.93]	3.74e-13	0.73
2270	7497	1.28 [1.17;1.41]	1.45e-07	0.73
<i>White (FINNGEN GWAS)</i>				
2270	7497	1.14 [1.08;1.20]	5.99e-07	0.73
2270	7497	1.11 [1.06;1.17]	1.07e-05	0.73
<i>Hispanic/Latino(MVP GWAS)</i>				
1051	6087	1.04 [0.95;1.14]	3.72e-01	0.83
1051	6087	1.05 [0.97;1.14]	2.00e-01	0.83



Supplementary Figure 5: Comparison of multi-population PGSs (BMI-adjusted GWAS)

PRS method	OR [95% CI]	P	AUC
<i>Meta-Analysis</i>			
<i>LDPred2:MVP</i>	1.49 [1.34;1.66]	2.31e-13	0.764
<i>LDPred2:MVP+ FinnGen</i>	1.47 [1.35;1.59]	7.91e-19	0.765
<i>Unweighted PRSsum</i>			
<i>LDPred2</i>	1.19 [1.14;1.24]	1.03e-15	0.788
<i>PRS-CS</i>	1.20 [1.15;1.25]	3.44e-18	0.788
<i>PRS-CSx</i>	1.19 [1.13;1.24]	1.45e-13	0.788
<i>Weighted PRSsum</i>			
<i>LDPred2</i>	1.51 [1.38;1.66]	1.01e-18	0.790
<i>PRS-CS</i>	1.98 [1.74;2.24]	1.61e-26	0.790
<i>PRS-CSx</i>	1.70 [1.54;1.88]	2.46e-25	0.789

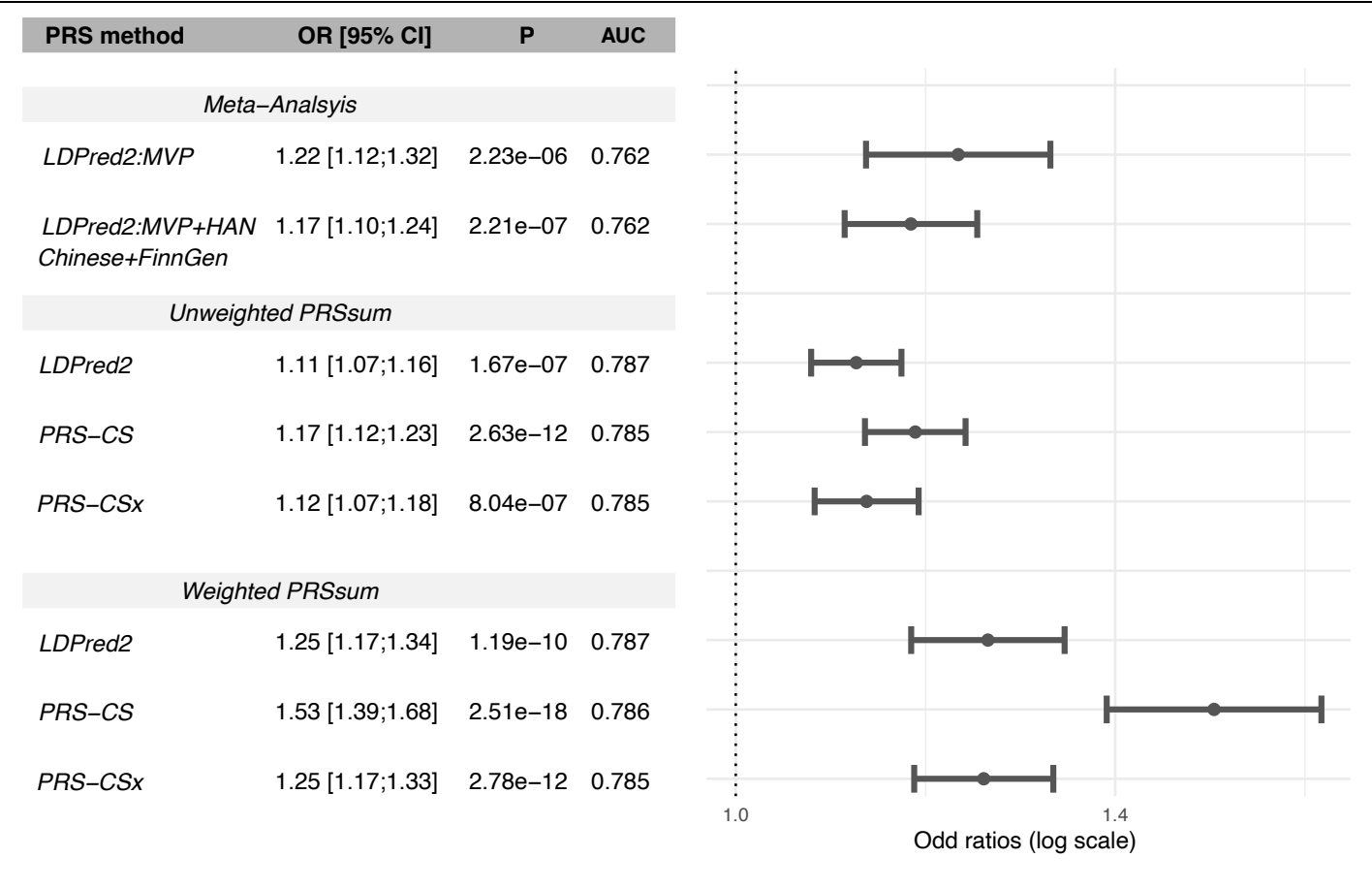


Comparison of PGS associations with OSA in TOPMed. PGS were constructed by different approaches to combine data across ancestry-specific GWAS. In all analyses, N cases=4,458 and N control= 17,466. OSA was defined using AHI/REI>15, or using self-reported doctor diagnosis (only in participants from the COPD gene study). All analyses used both MVP and FinnGen summary statistics, other than LDPred2:MVP over meta-analysis which used only MVP data.

AHI: apnea hypopnea index; OSA: obstructive sleep apnea; REI: respiratory event index.

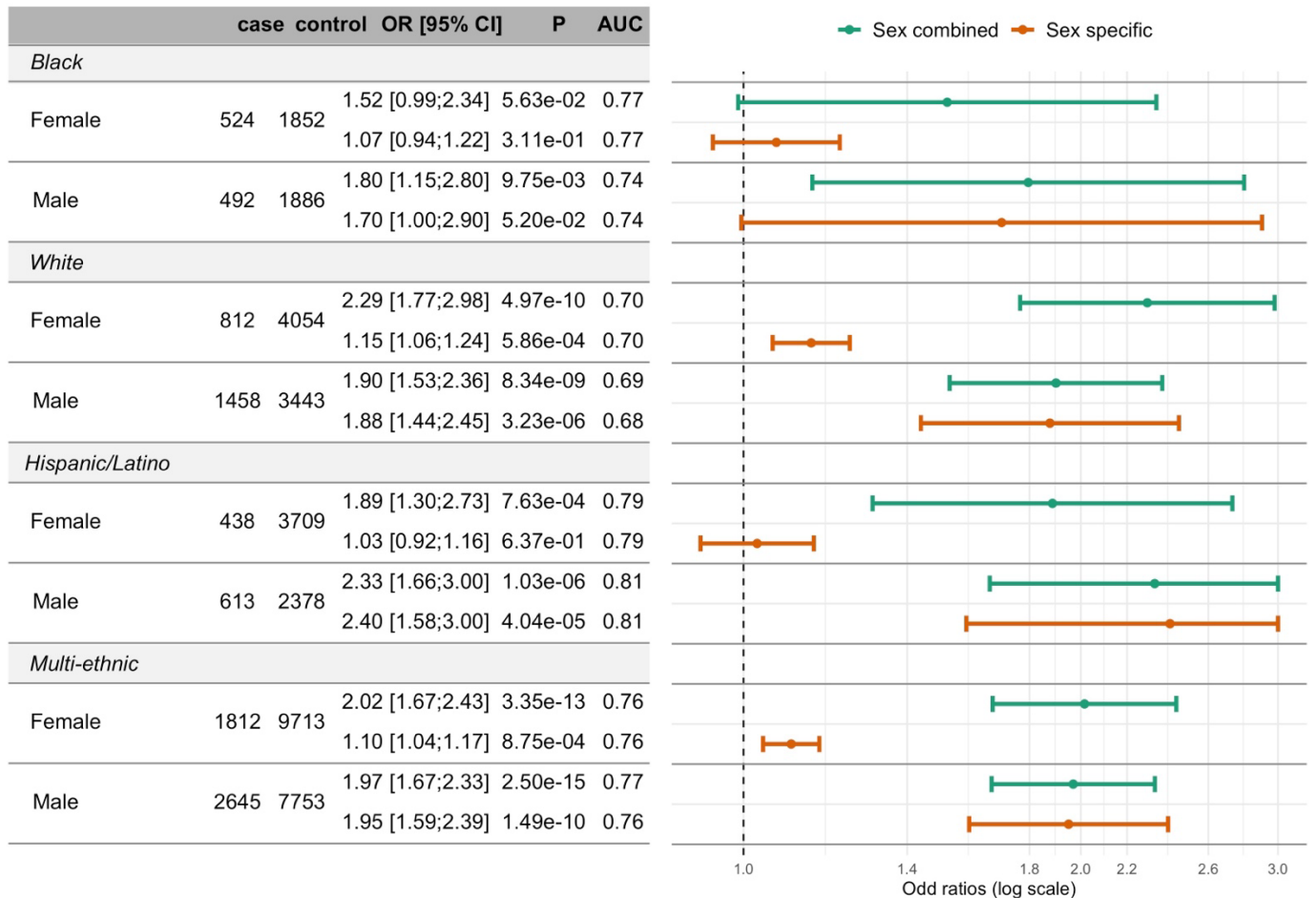


Supplementary Figure 6: Comparison of multi-population PGSs (BMI-unadjusted GWAS)



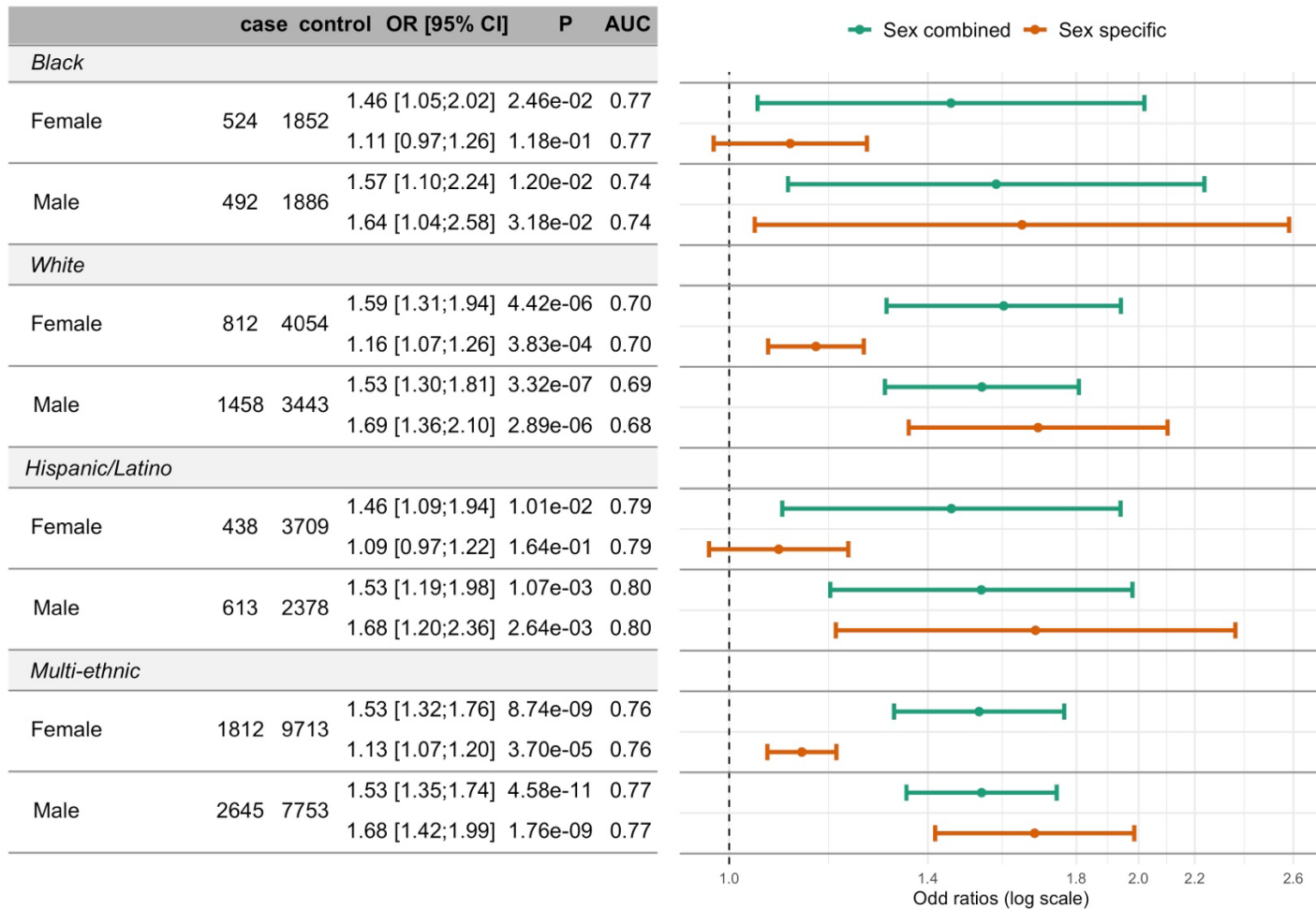
[results here: focusing on BMIunadj GWAS; BMI-adjusted analysis]  
 OSA was defined using AHI/REI>15, or using self-reported doctor diagnosis (only in participants from the COPD gene study). In all analyses N cases=4,458 and N control= 17,466.  
 All analyses used both MVP, FinnGen, and Han Chinese summary statistics, other than LDPred2:MVP over meta-analysis which used only MVP data.  
 AHI: apnea hypopnea index; OSA: obstructive sleep apnea; REI: respiratory event index

Supplementary Figure 7: comparison of estimated OSA associations with sex-specific and sex-combined OSA PGSs (BMI-adjusted).



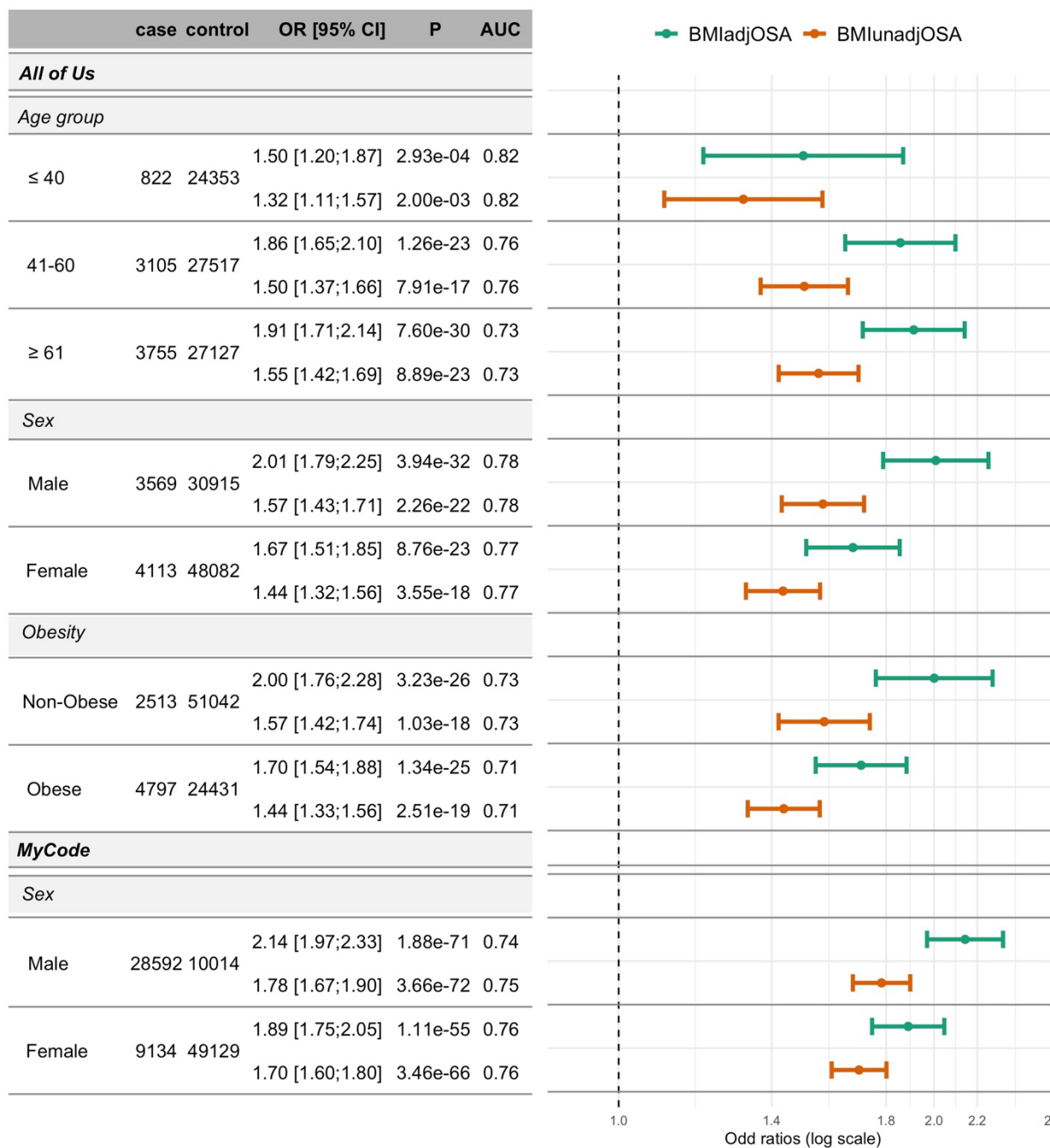
Associations of PGSs constructed using sex-specific and sex-combined analyses within their respective sex groups, in the combined and race/ethnicity-stratified TOPMed dataset. PGSs were constructed based on BMI-adjusted GWAS. Sex-combined PGS is the BMIadjOSA PGS selected as optimal in TOPMed.

Supplementary Figure 8: comparison of estimated OSA associations with sex-specific and sex-combined OSA PGSs (BMI-unadjusted).



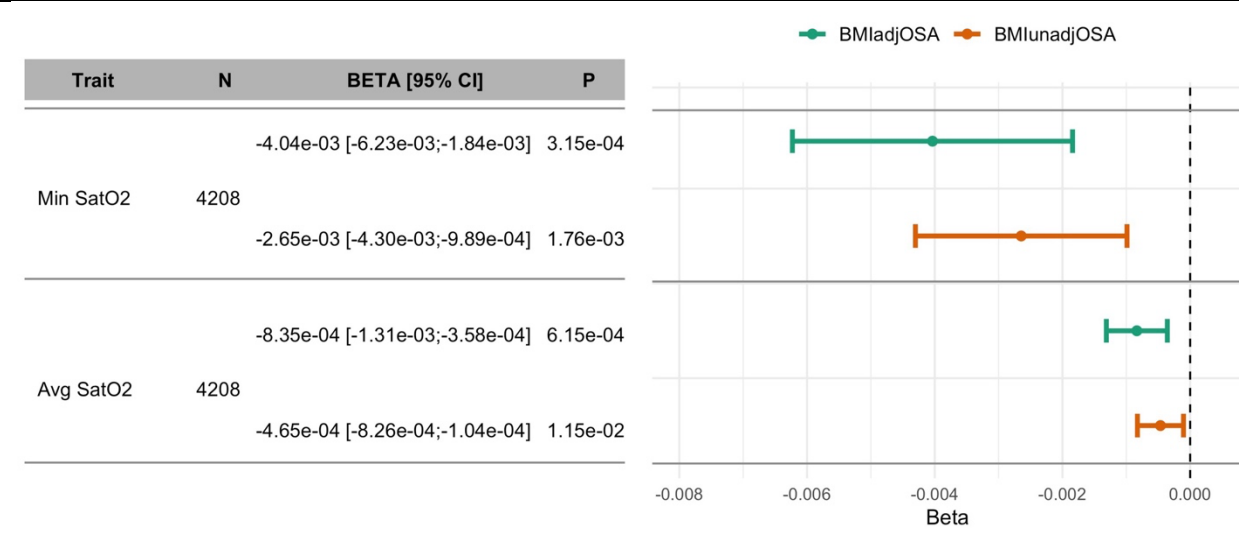
Associations of PGSs constructed using sex-specific and sex-combined analyses within their respective sex groups, in the combined and race/ethnicity-stratified TOPMed dataset. PGSs were constructed based on BMI-unadjusted GWAS. Sex-combined PGS is the BMIunadjOSA PGS selected as optimal in TOPMed.

Supplementary Figure 9: Estimated associations of OSA PGSs OSA in strata defined by sex (MyCode) and by sex, age, and obesity (All of Us).



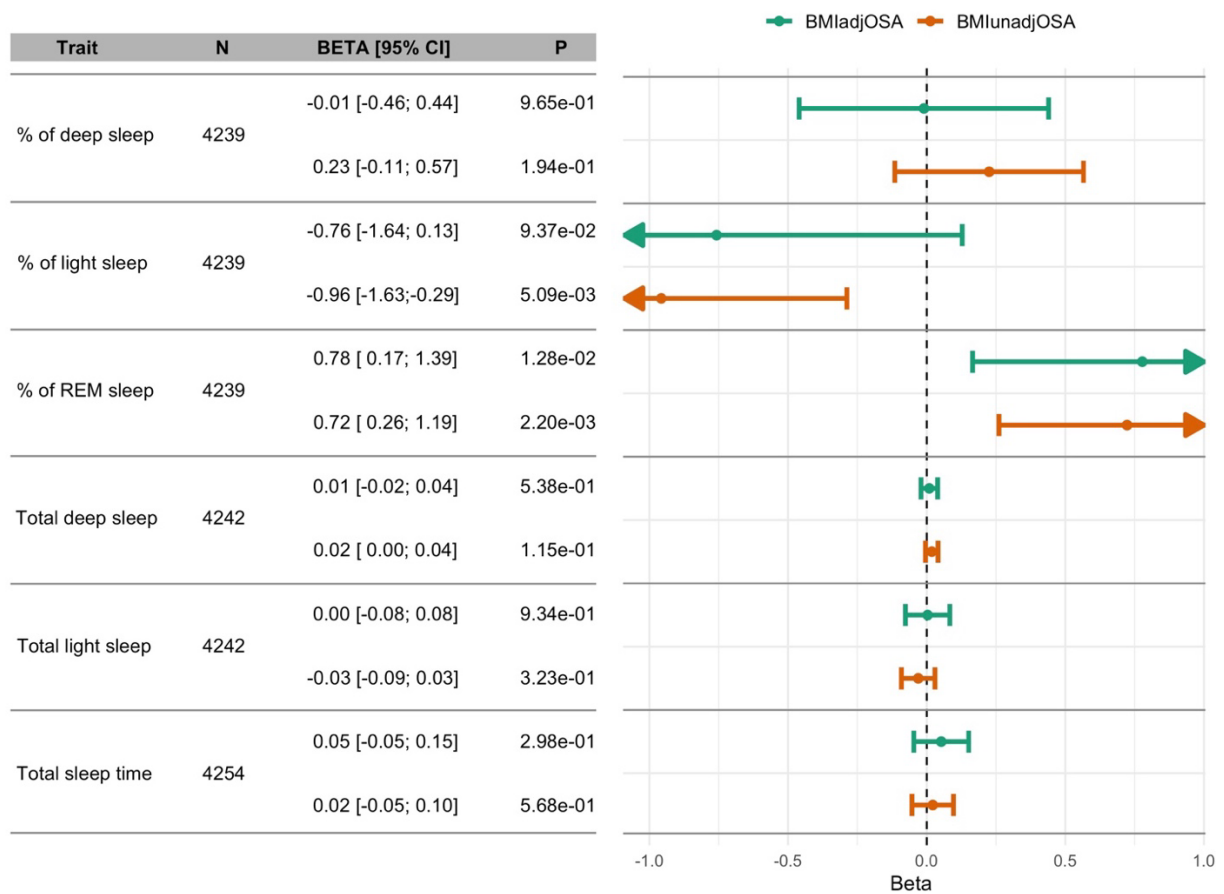
All of Us analyses were adjusted for age, sex, BMI (linear and squared terms) and 10 genetic PCs. MyCode analyses were adjusted for age, BMI (linear and squared terms), sex (in combined analysis), self-reported or EHR-derived race/ethnicity, and the top 20 genetic PCs.

Supplementary Figure 10: Estimated associations of OSA PGSs with oxygen saturation phenotypes in the Human Phenotypes Project.



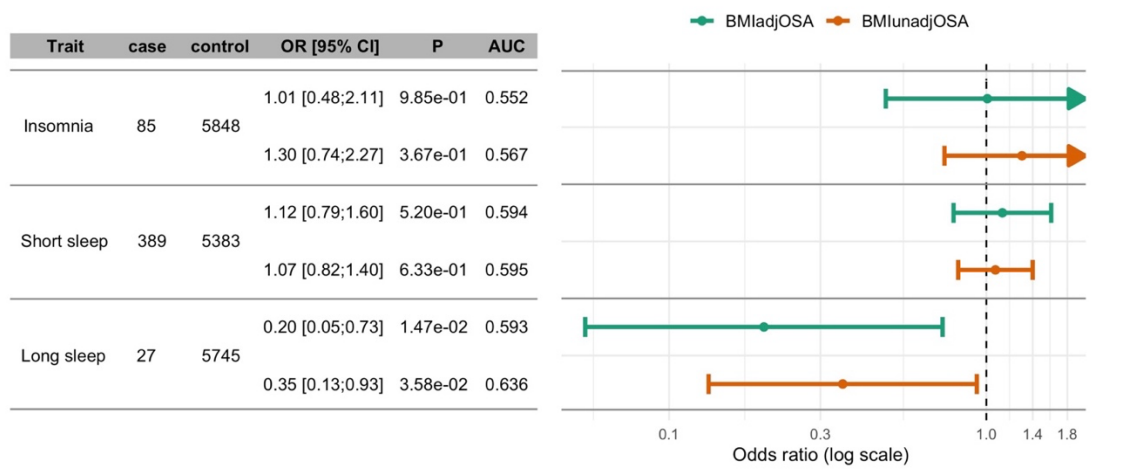
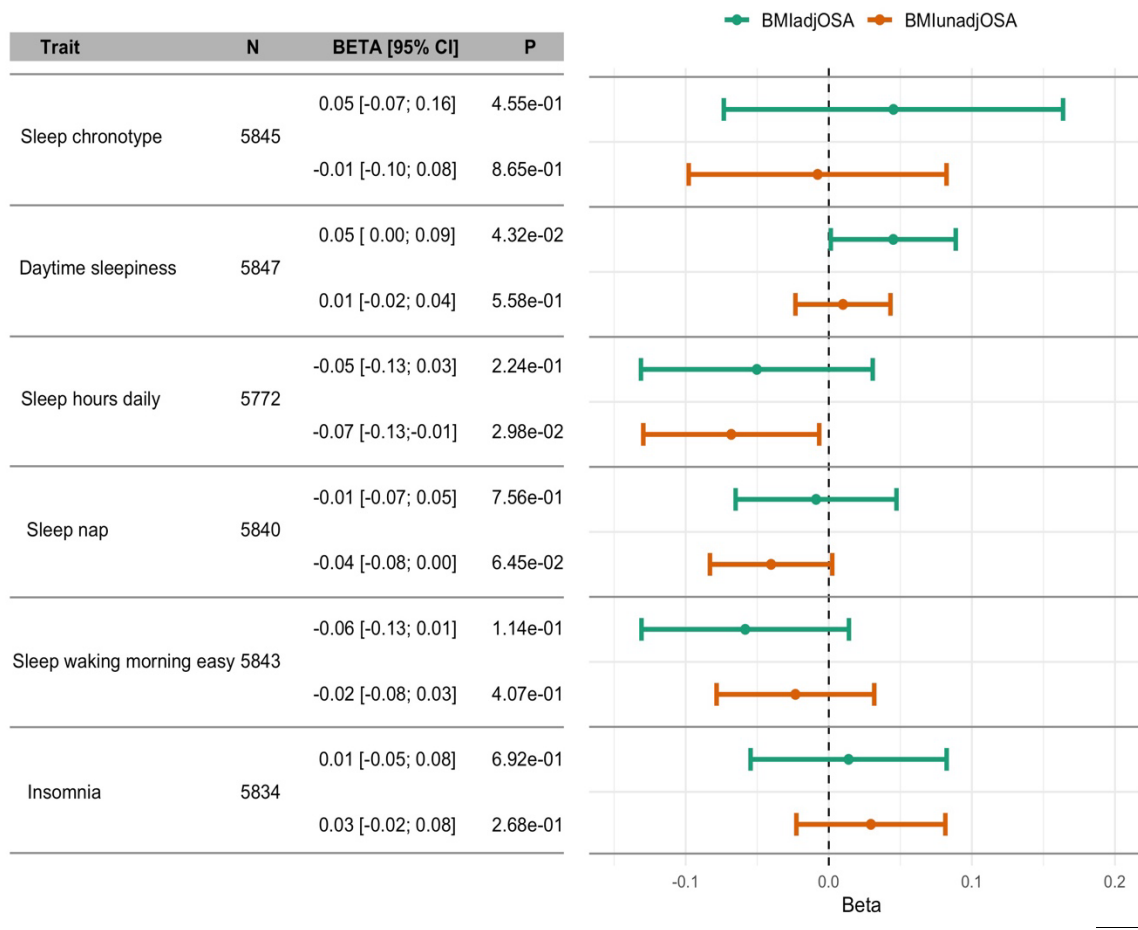
Estimated associations of the two OSA PGSs with log-transformed minimum and average oxyhemoglobin saturation measures during sleep (Min SatO2 and AvgSatO2), where the measures were averaged over the three nights of measurement. All analyses were adjusted for age, sex, BMI (linear and squared terms) and 10 genetic PCs.

Supplementary Figure 11: Estimated associations of OSA PGSs with sleep stage measures in the Human Phenotypes Project.



OSA PGS associations with sleep stage-related measures, averaged over the three nights of measurement. All analyses were adjusted for age, sex, BMI (linear and squared terms) and 10 genetic PCs.

Supplementary Figure 12: Estimated associations of OSA PGSs with self-reported sleep phenotypes in the Human Phenotypes Project.

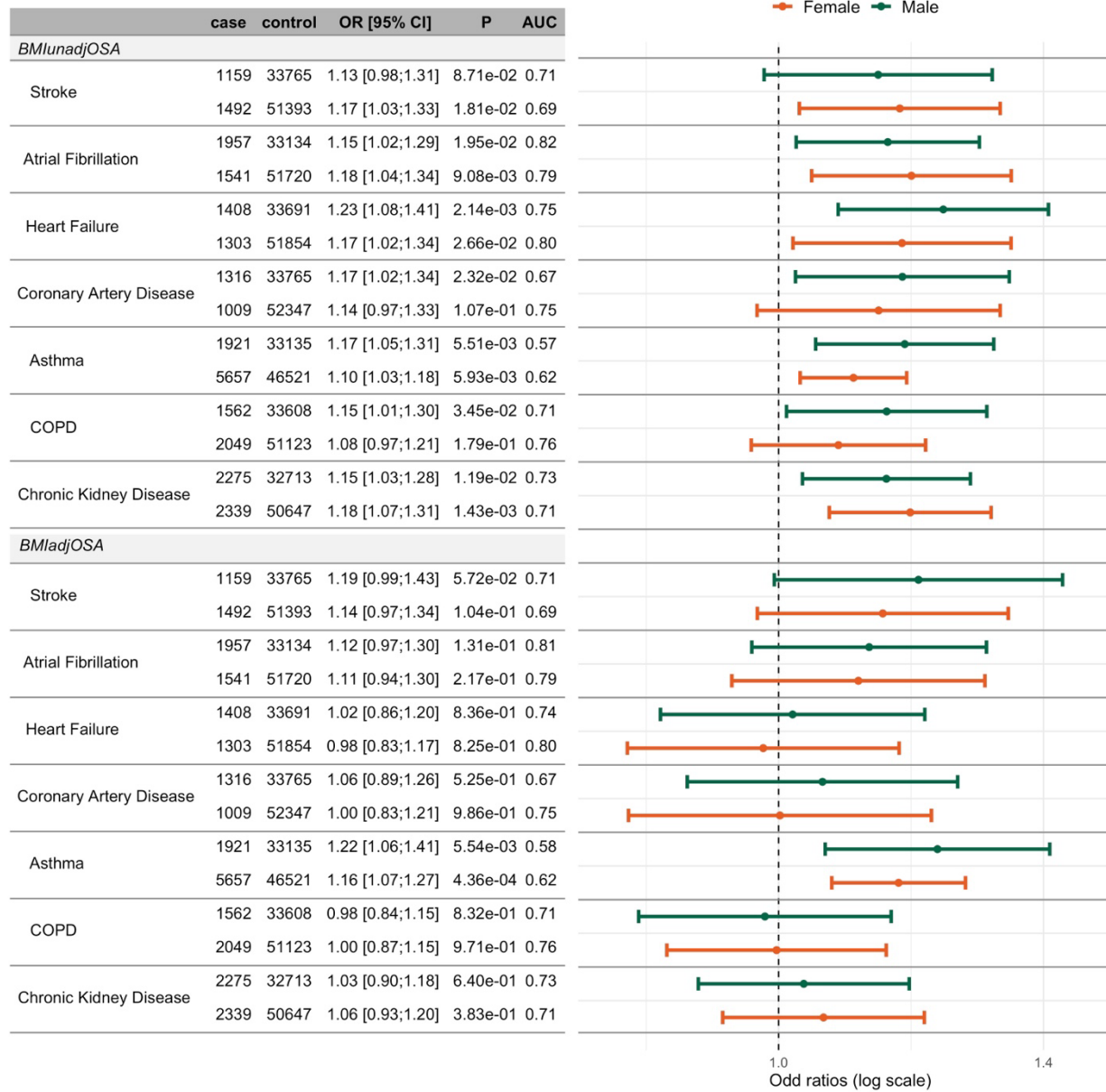


Excessive daytime sleepiness was assessed using the question “How likely are you to doze or fall asleep during day time you don’t mean to? E.g. when working or driving?”. Insomnia/sleeplessness was assessed using the question “Do you have trouble falling sleep at night or do you wake up in the middle of the night?”. Sleep naps were assessed using the question “Do you have nap during the day?”. Easiness of waking up in the morning was

assessed using the question “on average day, how easy do you find getting up in the morning?”. Long sleep was defined as self-reported sleep duration >9, and short sleep was defined as self-reported sleep duration < 6.  
All analyses were adjusted for age, sex, BMI (linear and squared terms) and 10 genetic PCs.

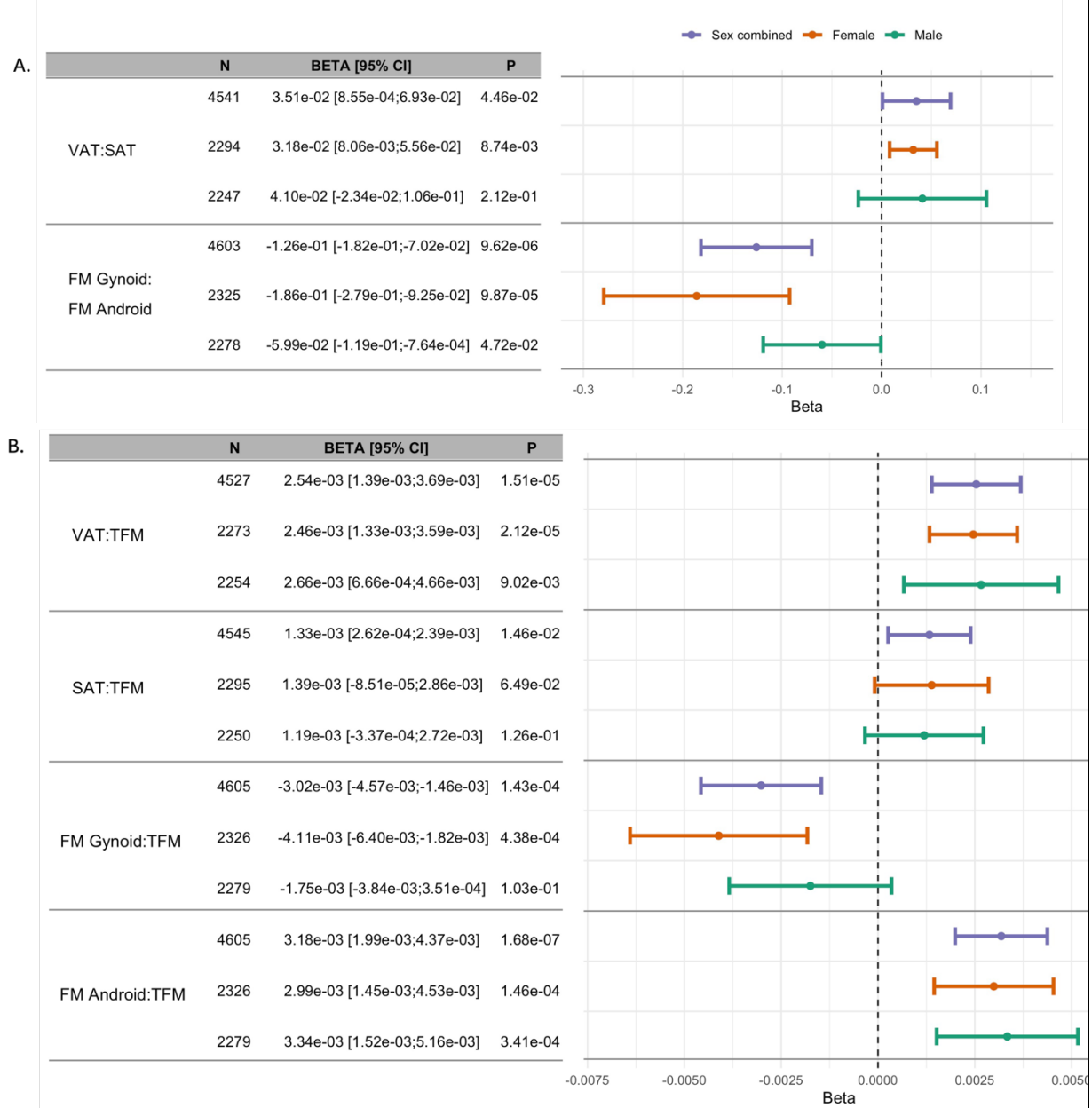


Supplementary Figure 13: Sex-stratified associations of OSA PGSs with clinical outcomes in All of Us



All analyses were adjusted for age, sex assigned at birth, BMI (linear and squared terms), and the first 10 genetic principal components.

Supplementary Figure 14: BMIunadjOSA PGS associations with DXA measures in the Human Phenotype Project



All associations were adjusted for age, sex, and the 10 first genetic principal components. BMI: body mass index; VAT: visceral adipose tissue; SAT: subcutaneous adipose tissue; FM: fat mass; TFM: total scan fat mass.

## Supplementary Tables

Supplementary Table 1: Characteristics of individuals participating in OSA PGS association analyses in TOPMed.

	<b>Black</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>
<b>N</b>	4760	230	9789	7196
<b>Age (mean (SD))</b>	56.38 (12.94)	67.86 (9.11)	62.67 (11.05)	48.05 (14.63)
<b>BMI (mean (SD))</b>	29.90 (7.00)	24.27 (3.30)	28.63 (5.92)	30.07 (6.30)
<b>study (%)</b>				
<b>ARIC</b>	0 ( 0.0)	0 ( 0.0)	1108 (11.3)	0 ( 0.0)
<b>CFS</b>	506 (10.6)	0 ( 0.0)	487 ( 5.0)	0 ( 0.0)
<b>CHS</b>	200 ( 4.2)	0 ( 0.0)	687 ( 7.0)	0 ( 0.0)
<b>COPDGene</b>	2983 (62.7)	0 ( 0.0)	6322 (64.6)	0 ( 0.0)
<b>FHS</b>	0 ( 0.0)	0 ( 0.0)	485 ( 5.0)	0( 0.0)
<b>HCHS/SOL</b>	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)	6734 (93.6)
<b>JHS</b>	579 (12.2)	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)
<b>MESA</b>	492 (10.3)	230 (100.0)	700 ( 7.2)	461 ( 6.4)
<b>Female sex (%)</b>	2379 (50.0)	115 ( 50.0)	4876 (49.8)	4183 (58.1)
<b>Moderate-to-severe OSA (%)</b>	1020 (21.4)	114 ( 49.6)	2271 (23.2)	1058 (14.7)

Moderate-to-severe OSA was defined based on AHI or REI  $\geq 15$ .

Abbreviations: AHI: Apnea Hypopnea Index; BMI: Body Mass Index; OSA: Obstructive Sleep Apnea; REI: Respiratory Event Index.

Supplementary Table 2: Characteristics of individuals participating in OSA analysis in the MGB Biobank.

	<b>Black</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>	<b>Unknown</b>
<b>N</b>	2023	932	34764	1182	1292
<b>Age (mean (SD))</b>	53.16 (16.23)	47.00 (15.67)	60.62 (16.83)	52.29 (16.56)	52.08 (18.42)
<b>Sex F (%)</b>	1267 (62.6)	542 (58.2)	18535 (53.3)	812 (68.7)	687 (53.2)
<b>BMI (mean (SD))</b>	31.38 (7.64)	25.15 (4.69)	28.32 (6.31)	29.80 (6.12)	28.79 (6.33)
<b>OSA Yes (%)</b>	345 (17.1)	66 ( 7.1)	4856 (14.0)	117 ( 9.9)	130 (10.1)

MGB: Mass General Brigham; BMI: body mass index; OSA: obstructive sleep apnea.

Supplementary Table 3: TOPMed-evaluated mean, SDs, and combination weights, of PRS-CS based PGSs used in the final BMIadjOSA PGS and BMIunadjOSA PGS.

Study	Mean	SD	PGS summation weight
<b>PGSs based on BMI-adjusted GWAS</b>			
<b>MVP-AFR</b>	2.66e-07	1.46e-07	0.042
<b>MVP-EUR</b>	6.21e-07	1.95e-07	0.729
<b>FinnGen</b>	-2.41E-08	4.95e-08	0.122
<b>MVP-HIS</b>	-5.62E-07	1.27e-07	0.040
<b>Final PGS sum</b>	1.27e-11	7.32E-01	
<b>PGSs based on BMI-unadjusted GWAS</b>			
MVP-AFR	3.32e-07	1.59e-07	0.024
<b>MVP-ASN + Han Chinese meta</b>	-1.94E-06	2.03e-07	0.043
<b>MVP-EUR</b>	3.96e-07	1.9e-07	0.566
<b>FinnGen</b>	-1.10E-07	5.26e-08	0.120
<b>MVP-HIS</b>	-1.14E-06	1.82e-07	0.037
<b>Final PGS sum</b>	9.14e-11	5.29E-01	

Means and SDs of the PGSs were computed based on the TOPMed dataset used in this study. PGS summation weights of the ancestry-specific PGSs were computed in a logistic regression of OSA over the PGSs, adjusted to age, sex, BMI, in the MGB biobank.

MPV: Million Veteran Program; AFR: MVP African HARE group; EUR: MVP European HARE group; HIS: MVP European HARE; ASN: MVP Asian HARE group; HARE: Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity; PGS: polygenic score; MGB: Mass General Brigham.

Supplementary Table 4: OSA assessment methods in participating studies.

Study	OSA assessment	Study role
<b>MVP</b>	Electronic health records based on ICD codes	GWAS to guide PGS development
<b>FinnGen</b>	Electronic health records-based extraction; subjective symptoms, clinical examination and sleep registration applying AHI $\geq 5$ or REI $\geq 5$ .	
<b>MGB Biobank</b>	Electronic health records (OSA code)	Develop PGS combination weights
<b>ARIC</b>	Part of the Sleep Heart Health Study. The first visit took place between November 1995 and January 1998. At-home polysomnography using a Compumedics P Series system, Abbotsville, Victoria, Australia. OSA defined based on AHI.	TOPMed PGS evaluation
<b>CHS</b>		
<b>FHS</b>		
<b>CFS</b>	In primary analysis: type 3 home sleep apnea test (Edentrace; Eden Prairie, MN). OSA defined based on REI. In REM/NREM stratified analysis, data was used from a different examination (between 2000 and 2006), where sleep apnea was assessed using 14-channel overnight polysomnography (Compumedics E series, Abottsford, AU) obtained in a clinical research unit.	
<b>COPDGene</b>	Self-reported doctor-diagnosed OSA.	
<b>HCHS/SOL</b>	At-home overnight recording using an ARES Unicorder 5.2 (B-Alert, Carlsbad, CA). OSA defined based on REI.	
<b>JHS</b>	Type 3 home sleep apnea device (Embletta-Gold device, Embla, Broomfield, CO). OSA defined based on REI.	
<b>MESA</b>	At home polysomnography, Compumedics Somte Systems, Abbotsville, Australia, AU. OSA defined based on AHI.	
<b>All of Us</b>	Electronic health records (ICD codes)	Independent study validation or OSA PGSs; association analysis with other clinical outcomes
<b>Geisinger's MyCode Study</b>	Electronic health records (ICD codes) with a minimum of three encounter dates with OSA code. Controls had zero encounters. Patients with one or two OSA codes were excluded from analysis.	Independent study validation or OSA PGSs
<b>Human phenotype project</b>	WatchPAT device (itamar medical, Israel) applied over 3 nights, AHI averaged across the nights. OSA defined based on AHI.	Independent study validation or OSA PGSs; association analysis with other sleep measures



Supplementary Table 5. Participants characteristics in analyses of OSA-related phenotypes (not REM/NREM stratified).

	<b>Black</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>
<b>N</b>	1777	230	3467	7195
<b>Age (mean (SD))</b>	59.24 (18.65)	67.86 (9.11)	63.60 (14.17)	48.05 (14.64)
<b>Sex F (%)</b>	1049 (59.0)	115 ( 50.0)	1832 (52.9)	4183 (58.1)
<b>BMI(mean (SD))</b>	31.18 (7.34)	24.27 (3.30)	28.56 (5.76)	30.07 (6.30)
<b>study (%)</b>				
<b>ARIC</b>	0 ( 0.0)	0 ( 0.0)	1108 (32.0)	0 ( 0.0)
<b>CFS</b>	506 (28.5)	0 ( 0.0)	487 (14.0)	0 ( 0.0)
<b>CHS</b>	200 (11.3)	0 ( 0.0)	687 (19.8)	0 ( 0.0)
<b>FHS</b>	0 ( 0.0)	0 ( 0.0)	485 (14.0)	0 ( 0.0)
<b>HCHS/SOL</b>	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)	6734 (93.6)
<b>JHS</b>	579 (32.6)	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)
<b>MESA</b>	492 (27.7)	230 (100.0)	700 (20.2)	461 ( 6.4)
<b>AHI/REI(mean (SD))</b>	17.60 (20.55)	20.37 (18.54)	15.39 (16.85)	7.71 (13.67)
<b>Avg SatO2 (mean (SD))</b>	94.64 (2.70)	94.92 (1.22)	94.23 (2.23)	96.28 (1.15)
<b>Min SatO2 (mean (SD))</b>	84.29 (7.97)	83.23 (7.56)	85.17 (6.90)	86.58 (6.48)
<b>Percent sleep &lt; 90 saturation (mean (SD))</b>	3.75 (10.44)	2.24 (4.45)	3.81 (10.65)	1.09 (3.80)



Supplementary Table 6. Participants characteristics in analyses of REM/NREM stratified OSA-related phenotypes.

	<b>Black</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>
<b>N</b>	1605	228	3213	460
<b>Age (mean (SD))</b>	61.48 (17.13)	67.82 (9.13)	65.30 (12.15)	68.45 (9.24)
<b>Sex F (%)</b>	955 (59.5)	113 ( 49.6)	1716 (53.5)	245 ( 53.3)
<b>BMI (mean (SD))</b>	31.00 (6.91)	24.27 (3.29)	28.39 (5.34)	30.11 (5.46)
<b>Study (%)</b>				
<b>ARIC</b>	0 ( 0.0)	0 ( 0.0)	1108 (34.5)	0 ( 0.0)
<b>CFS</b>	337 (21.0)	0 ( 0.0)	235 ( 7.3)	0 ( 0.0)
<b>CHS</b>	200 (12.5)	0 ( 0.0)	687 (21.4)	0 ( 0.0)
<b>FHS</b>	0 ( 0.0)	0 ( 0.0)	485 (15.1)	0 ( 0.0)
<b>JHS</b>	578 (36.0)	0 ( 0.0)	0 ( 0.0)	0 ( 0.0)
<b>MESA</b>	490 (30.5)	228 (100.0)	698 (21.7)	460 (100.0)
<b>Avg DesatO2 (mean (SD))</b>	3.42 (1.93)	3.72 (1.78)	2.63 (1.41)	3.62 (2.13)
<b>Avg DesatO2 NREM (mean (SD))</b>	3.22 (1.98)	3.50 (1.63)	2.53 (1.35)	3.33 (1.92)
<b>Avg DesatO2 REM (mean (SD))</b>	4.14 (3.19)	5.05 (3.79)	3.12 (2.39)	5.04 (4.34)
<b>Avg SatO2 NREM (mean (SD))</b>	94.99 (2.07)	95.02 (1.17)	94.38 (1.85)	94.49 (1.42)
<b>Avg SatO2 REM (mean (SD))</b>	94.36 (3.07)	94.43 (1.76)	94.07 (2.56)	93.61 (2.63)
<b>Min SatO2 NREM (mean (SD))</b>	86.58 (6.26)	85.40 (6.33)	86.81 (5.51)	84.77 (7.14)
<b>Min SatO2 REM (mean (SD))</b>	85.57 (7.69)	84.79 (7.68)	86.51 (6.45)	82.52 (9.60)
<b>AHI/REI NREM (mean (SD))</b>	14.23 (19.27)	18.38 (18.36)	12.98 (15.87)	18.91 (19.40)
<b>AHI/REI REM (mean (SD))</b>	27.74 (24.46)	26.60 (21.01)	22.43 (19.86)	33.18 (23.11)

Supplementary Table 7: Association of FEV1/FVC PGS with OSA continuous phenotypes

<b>Outcome</b>	<b>N</b>	<b>Estimated association</b>	<b>SE</b>	<b>P-value</b>
<b>AHI/REI</b>	12438	0.01	0.005	0.02
<b>Hypoxic burden</b>	10842	0.003	0.005	0.54
<b>Min SatO2</b>	12363	-0.0006	0.0003	0.06
<b>Avg SatO2</b>	12359	-0.0002	6.53E-05	0.0009

FEV1/FVC PGS weights were downloaded from the PGS catalog (PGP000244). Associations were adjusted for age, sex, BMI (both linear and squared terms), study, self-reported race/ethnicity, and 11 genetic PCs.

Supplementary Table 8: Geisinger’s MyCode Community Health Initiative (MyCode)  
Study health system participating patient characteristics.

	<b>Black/African American</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>	<b>Other/Mixed</b>
<b>N</b>	2,557	663	89,684	3,289	676
<b>Age (mean (SD))</b>	47.95 (15.43)	46.59 (15.89)	59.26 (17.13)	45.47	54.64 (15.88)
<b>Sex F (%)</b>	1,555 (60.81)	459 (69.23)	53,528 (59.69)	2,307 (70.14)	414 (61.24)
<b>BMI (mean (SD))</b>	32.55 (8.55)	26.50 (5.23)	30.59 (7.78)	31.39 (7.61)	30.47 (7.62)
<b>OSA Yes (%)</b>	573 (22.41)	65 (9.80)	17,721 (19.76)	635 (19.31)	154 (22.78)

Supplementary Table 9: Human Phenotypes Project participant characteristics.

	<b>Female</b>	<b>Male</b>
<b>N</b>	3,070	2,880
<b>Age (mean(SD))</b>	52.1(7.7)	51.3(7.7)
<b>BMI (mean(SD))</b>	25.6(4.3)	26.5(3.7)
<b>OSA based on medical diagnosis Yes (%)</b>	23(0.75)	119 (4.1)
<b>Insomnia based on medical diagnosis Yes (%)</b>	53(1.72)	32(1.11)

Supplementary Table 10: Human Phenotypes Project sleep monitoring participant characteristics.

	<b>Female</b>	<b>Male</b>
<b>N</b>	2609	2600
<b>Age (mean(SD))</b>	52.4(7.6)	51.6(7.7)
<b>BMI (mean(SD))</b>	25.7(4.4)	26.8(3.8)
<b>AHI (mean(SD))</b>	8.7(8.3)	13.0(10.8)
<b>AHI-REM (mean(SD))</b>	13.7(12.1)	18.7(13.3)
<b>AHI-NREM (mean(SD))</b>	7.1(7.6)	11.2(10.6)
<b>ODI (mean(SD))</b>	3.3(4.6)	5.7(7.2)
<b>ODI-REM (mean(SD))</b>	5.7(7.7)	8.6(9.5)
<b>ODI-NREM (mean(SD))</b>	2.5(4.0)	4.8(6.9)
<b>RDI (mean(SD))</b>	13.5(8.3)	18.8(10.5)
<b>RDI-REM (mean(SD))</b>	19.3(11.3)	25.4(12.1)
<b>RDI-NREM (mean(SD))</b>	11.5(8.0)	16.7(10.8)
<b>MinSat O2 (mean(SD))</b>	87.8(4.2)	87.1(4.6)
<b>MaxSat O2 (mean(SD))</b>	99.0(0.7)	99.0(0.7)
<b>AvgSat O2 (mean(SD))</b>	95.2(1.2)	94.9(1.1)
<b>% of deep sleep time (mean(SD))</b>	17.8(4.5)	17.4(4.5)
<b>% of light sleep time (mean(SD))</b>	57.8(8.8)	59.3(8.7)
<b>% of rem sleep time (mean(SD))</b>	24.4(5.9)	23.3(6.1)
<b>Total deep sleep time (mean(SD))</b>	1.1(0.3)	1.0(0.3)
<b>Total light sleep time (mean(SD))</b>	3.6(0.8)	3.5(0.8)
<b>Total sleep time (mean(SD))</b>	6.2(1.0)	5.9(1.0)
<b>OSA (AHI &gt; = 15) Yes (%)</b>	449(17.2)	797(30.7)

Supplementary Table 11: Characteristics of Human Phenotype Projects participants and their self-reported sleep phenotypes.

	<b>Female</b>	<b>Male</b>
<b>N</b>	3262	2950
<b>Age (mean(SD))</b>	52.1(7.8)	51.3(7.8)
<b>BMI (mean(SD))</b>	25.7(4.4)	26.5(3.8)
<b>Total sleep duration (mean(SD))</b>	6.9(0.9)	6.6(0.9)
<b>Long sleep Yes(%)</b>	19(0.58)	8(0.28)
<b>Short sleep Yes (%)</b>	182(5.58)	241(8.17)
<b>Insomnia/sleeplessness (%)</b>		
<b>Never / rarely</b>	1263(20.44)	1484(24.01)
<b>Sometime</b>	1193(19.3)	964(15.60)
<b>Usually</b>	781 (12.63)	487(7.89)
<b>Prefer not to answer</b>	6 (0.1)	2(0.03)
<b>Easy to wake up in the morning (%)</b>		
<b>Not at all easy</b>	209(3.38)	85(1.37)
<b>Not very easy</b>	714(11.54)	503(8.12)
<b>Fairly easy</b>	1468(23.72)	1439(23.25)
<b>Very easy</b>	844(13.64)	907(14.66)
<b>Prefer not to answer</b>	3(0.05)	2(0.03)
<b>Do not know</b>	11(0.18)	4(0.06)
<b>Sleep chronotype %</b>		
<b>Definitely morning person</b>	975(15.75)	718(11.60)
<b>More a morning than evening person</b>	1024(16.54)	885(14.30)
<b>More an evening than a morning person</b>	632(10.21)	660(10.66)
<b>Definitely evening person</b>	325(5.25)	336(5.42)
<b>Prefer not to answer</b>	5(0.08)	3(0.05)
<b>Do not know</b>	289(4.67)	338(5.42)
<b>Sleep nap (%)</b>		
<b>Never / rarely</b>	2103(34.00)	1956(31.62)
<b>Sometime</b>	851(13.76)	783(12.65)
<b>Usually</b>	289((4.67)	200(3.23)
<b>Prefer not to answer</b>	3(0.04)	1(0.01)
<b>Excessive daytime sleepiness (%)</b>		
<b>Never/rarely</b>	2673(43.16)	2300(37.14)
<b>Sometime</b>	461(7.44)	524(8.46)
<b>Often</b>	98(1.58)	104(1.68)
<b>Prefer not to answer</b>	1(0.01)	1(0.01)
<b>Do not know</b>	18(0.3)	12(0.21)

Excessive daytime sleepiness was assessed using the question “How likely are you to doze or fall asleep during day time you don’t mean to? E.g. when working or driving?”. Insomnia/sleeplessness was assessed using the question “Do you have trouble falling sleep at night or do you wake up in the middle of the night?”. Sleep naps were assessed using the question “Do you have nap during the day?”. Easiness of waking up in the morning was assessed using the question “on average day, how easy do you find getting up in the morning?”. Long sleep was defined as self-reported sleep duration >9, and short sleep was defined as self-reported sleep duration < 6.

Supplementary Table 12: Characteristics of individuals participating in All of Us association analyses.

	<b>Black</b>	<b>Asian</b>	<b>White</b>	<b>Hispanic/Latino</b>	<b>Other</b>	<b>Unknown</b>
<b>N</b>	18894	2843	46903	17555	552	3858
<b>Age (mean (SD))</b>	49.60 (14.72)	44.45 (17.03)	56.33 (16.89)	45.86 (15.84)	45.95 (17.04)	50.38 (17.36)
<b>BMI (mean (SD))</b>	30.32 (7.35)	25.04 (4.63)	28.53 (6.29)	30.08 (6.29)	28.21 (6.12)	28.88 (6.58)
<b>Sex at birth F(%)</b>	10366 (54.9)	1694 (59.6)	27855 (59.4)	11961 (68.1)	284 (51.4)	2236 (58.0)
<b>OSA Yes(%)</b>	1757 ( 9.3)	175 ( 6.2)	7228 (15.4)	1430 ( 8.1)	53 ( 9.6)	466 (12.1)
<b>T2D Yes(%)</b>	2856 (15.1)	260 ( 9.1)	5791 (12.3)	2675 (15.2)	57 (10.3)	532 (13.8)
<b>AFib Yes(%)</b>	525 ( 2.8)	53 ( 1.9)	3493 ( 7.4)	548 ( 3.1)	25 ( 4.5)	206 ( 5.3)
<b>CAD Yes(%)</b>	751 ( 4.0)	53 ( 1.9)	2032 ( 4.3)	510 ( 2.9)	22 ( 4.0)	169 ( 4.4)
<b>Stroke Yes(%)</b>	818 ( 4.3)	66 ( 2.3)	2684 ( 5.7)	699 ( 4.0)	20 ( 2.7)	200 ( 5.2)
<b>Asthma Yes(%)</b>	1963 (10.4)	170 ( 6.0)	5530 (11.8)	1927 (11.0)	44 ( 8.0)	515 (13.3)
<b>CKD Yes(%)</b>	1415 ( 7.5)	80 ( 2.8)	3469 ( 7.4)	1033 ( 5.9)	22 ( 4.0)	329 ( 8.5)
<b>HF Yes(%)</b>	1043 ( 5.5)	51 ( 1.8)	2288 ( 4.9)	591 ( 3.4)	17 ( 3.1)	205 ( 5.3)
<b>HTN Yes(%)</b>	13044 (69.5)	1358 (48.2)	30174 (64.9)	10112 (57.9)	0	0
<b>COPD Yes(%)</b>	1126 ( 6.0)	30 ( 1.1)	3018 ( 6.4)	548 ( 3.1)	19 ( 3.4)	249 ( 6.5)



Supplementary Table 13: Comparison of newly developed and existing OSA PGS associations with OSA in All of Us.

Study	GWAS adjustment	Method	PMID	OR and 95% CI	P-value	AUC
<b>BMIunadjOSA</b>	BMI unadjusted	Weighted PGS summation of PRS-CS generated PGS	Current manuscript	1.49 [1.40;1.58]	2.72e-38	0.76
<b>BMIadjOSA</b>	BMI adjusted	Weighted PGS summation of PRS-CS generated PGS	Current manuscript	1.82 [1.68;1.96]	1.12e-52	0.76
<b>Sofer &amp; Kurniansyah et al.</b>	BMI unadjusted	Genome-wide significant SNPs	36989840	1.61 [1.27; 2.06]	9.03e-05	0.76
<b>Sofer &amp; Kurniansyah et al.</b>	BMI adjusted	Genome-wide significant SNPs	36989840	1.21 [0.76; 1.90]	0.41	0.76
<b>Zhang et al</b>	BMI unadjusted	LDPred2	36174398	1.26 [1.18;1.36]	1.78e-11	0.76

N cases: 7,682, N controls: 78,997.

All association models were adjusted for sex assigned at birth, age, BMI (linear and squared terms), the first 10 PCs of genetic ancestry.

Abbreviations. AIC: area under the (receiver operating) curve; BMI: body mass index; CI: confidence interval; OR: odds ratio; OSA: obstructive sleep apnea; PGS: polygenic score.

AUC was the same for all PGSs when rounding to the second digit.

Supplementary Table 14: Characteristics of DXA measures in the Human Phenotype Project

	<b>Female</b>	<b>Male</b>
<b>Age (mean(SD))</b>	52.2 (7.5)	51.7 (7.7)
<b>BMI (mean(SD))</b>	25.6 (4.3)	26.8 (3.8)
<b>VAT fat mass ((mean(SD)))</b>	0.63 (0.44)	1.22 (0.74)
<b>SAT fat mass (mean(SD))</b>	1.59 (0.75)	1.41 (0.70)
<b>Android fat mass (mean(SD))</b>	2.05 (0.94)	2.34 (1.11)
<b>Gynoid fat mass (mean(SD))</b>	4.73 (1.42)	3.49 (1.20)
<b>Total fat mass (mean(SD))</b>	25.83 (8.47)	23.32 (8.13)
<b>VAT: Total fat mass (mean(SD))</b>	0.02 (0.01)	0.05 (0.02)
<b>SAT: Total fat mass (mean(SD))</b>	0.06 (0.01)	0.06 (0.01)
<b>FM Gynoid: Total fat mass (mean(SD))</b>	0.19 (0.02)	0.15 (0.02)
<b>FM Android: Total fat mass (mean(SD))</b>	0.08 (0.02)	0.10(0.02)
<b>VAT: SAT (mean(SD))</b>	0.40 (0.25)	0.93 (0.62)
<b>FM Gynoid: FM Android (mean(SD))</b>	2.62 (0.93)	1.67 (0.59)

Fat masses are in kilograms.

BMI: body mass index; VAT: visceral adipose tissue; SAT: subcutaneous adipose tissue; FM: fat mass.

Supplementary Table 15: FinnGen participant characteristics.

	<b>All N=412,181</b>	<b>OSA N=43,901</b>	<b>Non-OSA N=368,280</b>
<b>Sex (male)</b>	181,871	26,199	155,672
<b>Age (SD)</b>	60.3 (18.9)	64.4 (12.6)	59.9 (18.4)
<b>BMI (SD)</b>	27.3 (5.5)	31.7 (6.6)	26.8 (5.1)

Baseline characteristics of the individuals with and without obstructive sleep apnea (OSA) diagnosis in FinnGen.

Supplementary Table 16: TOPMed studies contributing to OSA PGS association with each of the non-OSA sleep phenotypes

<b>Sleep phenotype</b>	<b>Participating studies</b>
<b>Phenotypes measured using sleep apnea device</b>	
<b>AHI/REI*</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA During REM and NREM: ARIC, CHS, FHS, MESA, CFS (subset)
<b>Hypoxic burden</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA
<b>Minimum, average oxyhemoglobin saturation during sleep</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA
<b>Average oxyhemoglobin desaturation during respiratory events</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA
<b>Percent sleep time with oxyhemoglobin saturation &lt; 90%</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA
<b>Self-reported phenotypes</b>	
<b>Sleep duration, long, short sleep</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA, WHI
<b>Epworth sleepiness scale, excessive daytime sleepiness</b>	ARIC, CHS, FHS, CFS, HCHS/SOL, JHS, MESA
<b>Women Health Initiative insomnia rating scale, insomnia</b>	CHS, HCHS/SOL, JHS, MESA, WHI

\*HCHS/SOL, measured REI while other studies measured AHI.

AHI: Apnea Hypopnea Index; REI: Respiratory Event Index.

## Supplementary Note 1: Choosing optimal OSA PGSs in analysis in TOPMed.

### Using UK Biobank as a reference panel is sufficient for developing PGS based on MVP summary statistics

We studied whether using UK Biobank reference panel to develop PGSs based on summary statistics from MVP (and potentially other studies, via meta-analysis) is sufficient.

Supplementary Figures 1 and 2 visualizes the associations of ancestry-specific PGS developed using PRS-CS with OSA in TOPMed individuals of similar ancestry as the GWAS dataset used for PGS derivation. It compares the estimated associations when using MVP and when using UKBB - matched reference panels, i.e., EUR HARE group (MVP) or EUR participants (UKBB) for European ancestries PGS, AFR HARE group (MVP) or AFR participants (UKBB) for African ancestries PGS, and HIS HARE group (MVP) or AMR participants (UKBB) for Hispanic/Latino PGS. Note that Asian-specific PGS was not derived due to the low sample size of participants of East Asian ancestries. However, it was later derived as part of PRS-CSx multi-ancestry derivation as explained later. One can see that PGS that were derived using UKBB participants for LD reference panels performed well and sometimes better than PGSs that used MVP participants for LD reference panels. This is likely related to the imputation quality of the SNPs. Thus, when using the PRS-CS and the PRS-CSx software in multi-PGS methods next, we only focus on PGSs developed using the UKBB reference panels.

### Identifying best performing single-ancestry PGSs

Supplementary Figures 3 and 4 compares single-ancestry PGS constructed using PRS-CS (UKBB reference panel) and using LDpred2 (MVP reference panel) in association with OSA in their

respective ancestry group in the TOPMed dataset. The results are similar between LDpred2 and PRS-CS. We therefore move forward with PRS-CS for multi-PGS methods.

### Multi-ethnic GWAS PGS versus a multi-PGS combining ancestry-specific PGSs

Supplementary Figure 5 compares the OSA associations of a PGS constructed based on a multi-ethnic GWAS, where we meta-analyzed MVP and FinnGen, to multi-PGS, i.e. PGS summations combining PGSs developed using PRS-CSx (in a joint model of a few ancestries), using LDpred2 (each genetic ancestry group PGS was developed separately), and using PRS-CS (each genetic ancestry group PGS was developed separately). Associations are provided in the combined and in separate race/ethnic groups. The weighted sum of PRS-CS constructed PGSs had the strongest associations with OSA. This is true also when constructing PGS based on BMI-unadjusted analysis (Supplementary Figure 6). We proceed with these PGSs as the best performing OSA PGSs, and we refer to them as BMIadjOSA-PGS and BMIunadjOSA-PGS.

### Supplementary Note 2: MGB biobank

Samples, genomic data, and health information were obtained from the Mass General Brigham Biobank, a biorepository of consented patient samples at Mass General Brigham. Phenotypic data was extracted from the MGB Biobank on November 14, 2021. OSA status was extracted as “Obstructive sleep apnea OSA existence” yes/no, and BMI was the median BMI value available in the records.

## **DNA samples**

DNA samples are processed from whole blood that was collected as a dedicated research draw or as a clinical discard. Dedicated research samples are aimed to be processed within four hours of collection. Clinical discards are processed 24+ hours after collection. Whole blood is spun to buffy coat with a centrifuge and the buffy coat is stored in a freezer up to several months. The buffy coat is then extracted to DNA. The DNA is then placed in an ultralow freezer (-80°C). Each DNA aliquot contains a minimum of 2 ug of DNA. The concentration varies.

## **Genotyping**

Samples have been genotyped using three versions of the biobank SNP array offered by Illumina that is designed to capture the diversity of genetic backgrounds across the globe. The first batch of data was generated on the Multi-Ethnic Genotyping Array (MEGA) array, the first release of this SNP array. The second, third, and fourth batches were generated on the Expanded Multi-Ethnic Genotyping Array (MEGA Ex) array. All remaining data were generated on the Multi-Ethnic Global (MEG) BeadChip.

## **Imputation**

Prior to performing imputation, files were converted to VCF format, separated by chromosomes. When multiple probes measured the same genotypes, they were checked for concordance and were set to a missing value if the genotypes did not match. Files were uploaded to the Michigan Imputation Server, and Genotypes were imputed using TOPMed reference panel. Genomic coordinates are provided in GRCh38.

## **Quality control**

We performed quality control using PLINK (v2.0). We filtered SNPs with low-quality imputation ( $r < 0.5$ ), with missing call rates  $> 0.1$ , HWE p-value less than  $1 \times 10^{-6}$  and MAF  $< 1\%$ .

We computed principal component (PC) using PLINK: we pruned the genotype data using a window size of 1000 variants, sliding across the genome with a step size of 250 variants at a time, filtering out any SNPs with LD  $R^2 > 0.1$ . We used unrelated individuals (3rd degree, identified using PLINK) to compute the loadings for the first 10 PCs.

## **PGS construction**

We constructed all PGS using PRSice 2 (1), using the same SNPs as those selected in the various developed PGS, i.e. without further clumping. PGSs were scaled using the same mean and SD as the TOPMed dataset, so that PGS summation weights will be appropriate for TOPMed.

## **Ethics statement**

All Biobank subjects have provided their consent to join the Partners Biobank, which includes agreeing to provide a blood sample linked to the electronic medical record. Subjects also agree to be recontacted by the Partners Biobank staff as needed.

## **Acknowledgements**



We thank Mass General Brigham Biobank for providing samples, genomic data, and health information data.

## Supplementary Note 3: TOPMed parent study descriptions

### ARIC

The Atherosclerosis Risk in Communities study (dbGaP accession phs000090) is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations, a fifth exam in 2011-2013, a sixth exam in 2016-2017, a seventh exam in 2018-2019, an eighth exam in 2020, a ninth exam in 2021-2022, a tenth exam in 2023, and the 11th exam started in 2024. The ARIC study has been described in detail previously (2,3).

### **Ethics statement:**

The ARIC study has been approved by a single Institutional Review Board (sIRB) at Johns Hopkins School of Medicine and Institutional Review Boards (IRB) at all participating institutions: University of North Carolina at Chapel Hill IRB, Johns Hopkins University School of Public Health IRB, University of Minnesota IRB, Wake Forest University Health Sciences IRB, and University of Mississippi Medical Center IRB. Study participants provided written informed consent at all study visits.

### **ARIC acknowledgements:**

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers 75N92022D00001,

75N92022D00002, 75N92022D00003, 75N92022D00004, 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions.

WGS for “NHLBI TOPMed: Atherosclerosis Risk in Communities (ARIC)” (phs001211) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201500015C and 3U54HG003273-12S2) and the Broad Institute for MIT and Harvard (3R01HL092577- 06S1). The Genome Sequencing Program (GSP) was funded by the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross program scientific initiatives and provided logistical and general study coordination. The Centers for Common Disease Genomics (CCDG) program was supported by NHGRI and NHLBI, and whole genome sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (UM1 HG008898).

## CARDIA

The Coronary Artery Risk Development in Young Adults study (dbGaP accession phs000285) is a prospective multicenter study with 5,115 adults Caucasian and African American participants of the age group 18-30 years at baseline, recruited from four centers at the baseline examination in 1985-1986 (4). The recruitment was done from the total community in Birmingham, AL, from selected census tracts in Chicago, IL and Minneapolis, MN; and from the Kaiser Permanente health plan membership in Oakland, CA. Nine examinations have been completed in the years 0, 2, 5, 7, 10, 15, 20, 25 and 30, with high retention rates (91%, 86%, 81%, 79%, 74%, 72%, 72%, and 71%, respectively) and written informed consent was obtained in each visit.

### **Ethics statement:**

All CARDIA participants provided informed consent, and the study was approved by the Institutional Review Boards of the University of Alabama at Birmingham and the University of Texas Health Science Center at Houston.

**CARDIA acknowledgements:**

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (75N92023D00002 & 75N92023D00005), Northwestern University (75N92023D00004), University of Minnesota (75N92023D00006), and Kaiser Foundation Research Institute (75N92023D00003). CARDIA was also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005).

**CFS**

The Cleveland Family Study (CFS) was designed to examine the genetic basis of sleep apnea in 2,534 African-American and European-American individuals from 356 families. Index probands with confirmed sleep apnea were recruited from sleep centers in northern Ohio, supplemented with additional family members and neighborhood control families (5). Four visits occurred between 1990 and 2006; in the first 3, data were collected in participants' homes while the last occurred in a clinical research center (2000 - 2006). Measurements included sleep apnea monitoring, blood pressure, anthropometry, spirometry and other related phenotypes. Blood samples (overnight fasting, before bed and following an oral glucose tolerance test), nasal and oral ultrasound, and ECG were also obtained during the 4th exam. Institutional Review Board approval and signed informed consent was obtained for all participants.

**Ethics statement:**

Cleveland Family Study was approved by the Institutional Review Board (IRB) of Case Western Reserve University and Mass General Brigham (formerly Partners HealthCare). Written informed consent was obtained from all participants.

## CHS

The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults  $\geq 65$  years conducted across four field centers (6). The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists; subsequently, an additional predominantly African-American cohort of 687 persons was enrolled for a total sample of 5,888. Participants underwent annual extensive clinical examinations through 1999 with additional phone contact at six month intervals between clinic visits and every six months after clinic visits ended to ascertain hospitalizations and health status. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples.

### **Ethics statement:**

CHS was approved by institutional review committees at each field center and individuals in the present analysis had available DNA and gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

### **CHS acknowledgements:**

Cardiovascular Health Study: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006, and grants R01HL120393, U01HL080295, U01HL130114, and HL105756 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## COPDGene

COPDGene (7) is a cohort study for respiratory disease research, recruiting more than 10,000 subjects between the ages of 45 and 80 who had at least 10 pack-years of smoking during January 2008 - June 2011 at 21 clinical centers. Participants were characterized using spirometry, six-minute walk, inspiratory and expiratory chest CT scans, respiratory symptoms, medical history, medication history and 36-Item short form health survey. In the current analysis, we only used COPDGene control participants (meaning, individuals without COPD).

### **OSA phenotype in COPDGene:**

In their respiratory disease questionnaire, participants were asked: "Have you ever had sleep apnea?", and if the response was "yes", they were asked: "Was it diagnosed by a doctor or other health professional?". We used OSA definition based on a positive ("yes") response to diagnosis by a doctor or other health professional.

### **Ethics statement:**

All COPDGene participants provided written informed consent, and the study was approved by the Institutional Review Boards of the participating clinical centers.

### **COPDGene acknowledgements:**

The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: <http://www.copdgene.org/directory>

## FHS

The Framingham Heart Study (dbGaP accession phs000007) began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort (mean age 37 years; 52% women) consisted of 5,124 children and spouses of children of the original cohort. A third-generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled in 2002-2005 and are seen every 4 to 8 years. Details of study designs for the three cohorts are summarized elsewhere (8–10). At each clinic visit, a medical history was obtained, and participants underwent a physical examination. Only study participants consented for genetic and non-genetic data are included. FHS has been approved by the Boston University IRB.

### **Ethics statement:**

The Framingham Heart Study was approved by the Institutional Review Board of the Boston University Medical Center. All study participants provided written informed consent.

### **FHS acknowledgements:**

The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

## HCHS/SOL

The Hispanic Community Health Study/Study of Latinos (dbGaP accession phs000810) is a community-based longitudinal cohort study of 16,415 self-identified Hispanic/Latino persons aged 18–74 years and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and

national origin or background (11–13). The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto Rican (n = 2,728), and South American (n = 1,068). The HCHS/SOL baseline clinical examination occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. Visit 2 took place between 2014 and 2017, which re-examined 11,623 participants from the baseline sample. Visit 3 has started in 2020 and will last 4 years, ending on January 2024. In addition to clinic visit, participants are contacted annually to assess clinical outcomes. The study was approved by the Institutional Review Boards at each participating institution and written informed consent was obtained from all participants.

**Ethics statement:**

This study was approved by the institutional review boards (IRBs) at each field center, where all participants gave written informed consent, and by the Non-Biomedical IRB at the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the study are: Non-Biomedical IRB at the University of North Carolina at Chapel Hill. Chapel Hill, NC; Einstein IRB at the Albert Einstein College of Medicine of Yeshiva University. Bronx, NY; IRB at Office for the Protection of Research Subjects (OPRS), University of Illinois at Chicago. Chicago, IL; Human Subject Research Office, University of Miami. Miami, FL; Institutional Review Board of San Diego State University. San Diego, CA.

**HCHS/SOL acknowledgements:**

The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC- 65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago – HHSN268201300003I / N01- HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following

Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

## JHS

The Jackson Heart Study (dbGaP accession phs000286) is a longitudinal investigation of genetic and environmental risk factors associated with the disproportionate burden of cardiovascular disease in African Americans (14,15). At baseline, the JHS recruited 5306 African American residents of the Jackson, Mississippi Metropolitan Statistical Area, which included approximately 6.6% of all African American adults aged 35-84 residing in the area. Participants were recruited via random sampling (17% of participants), volunteers (30%), prior participants in the Atherosclerosis Risk in Communities (ARIC) study (31%), and secondary family members (22%). Among these participants, approximately 3400 gave consent that allows genetic research. JHS participants received three back-to-back clinical examinations (Exam 1, 2000-2004; Exam 2, 2005-2008; and Exam 3, 2009-2013), and a fourth clinical examination started in 2020. Participants are also contacted annually by telephone to update personal and health information including vital status, interim medical events, hospitalizations, functional status and sociocultural information.

### **Ethics statement:**

The Institutional Review Boards at Jackson State University, Tougaloo College, and the University of Mississippi Medical Center approved the study, and all participants provided written informed consent.

### **JHS acknowledgements:**



The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staffs and participants of the JHS.

Genome sequencing (dbGap accession phs000964) was performed at the Northwest Genomics Center (HHSN268201100037C). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (R01HL117626; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL120393; U01HL120393; contract HHSN268201800001I).

## MESA

The Multi-Ethnic Study of Atherosclerosis (dbGaP accession phs000209) is a study of the characteristics of subclinical cardiovascular disease (disease detected non-invasively before it has produced clinical signs and symptoms) and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease (16). MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were White, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and

congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by five examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

**Ethics statements:**

All MESA participants provided written informed consent, and the study was approved by the Institutional Review Boards at The Lundquist Institute (formerly Los Angeles BioMedical Research Institute) at Harbor-UCLA Medical Center, University of Washington, Wake Forest School of Medicine, Northwestern University, University of Minnesota, Columbia University, and Johns Hopkins University.

**MESA acknowledgements:**

The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, R01AG08598, and R01HL105756.

Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-64278.

Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found at <http://www.mesa-nhlbi.org>.

## WHI

The Women's Health Initiative (WHI) is a prospective national health study focused on identifying optimal strategies for preventing chronic diseases that are the major causes of death and disability in postmenopausal women. The WHI initially recruited 161,808 women between 1993 and 1997 with the goal of including a socio-demographically diverse population with diversity background groups proportionate to the total minority population of US women aged 50-79 years. The WHI consists of two major parts: a set of randomized Clinical Trials and an Observational Study. The WHI Clinical Trials (CT; N=68,132) includes three overlapping components, each a randomized controlled comparison: the Hormone Therapy Trials (HT), Dietary Modification Trial, and Calcium and Vitamin D Trial. A parallel prospective observational study (OS; N = 93,676) examined biomarkers and risk factors associated with various chronic diseases. While the HT trials ended in the mid-2000s, active follow-up of the WHI-CT and WHI-OS cohorts has continued for over 25 years, with the accumulation of large numbers of diverse clinical outcomes, risk factor measurements, medication use, and many other types of data.

### **Ethics statement:**

All WHI participants provided informed consent and the study was approved by the Institutional Review Board (IRB) of the Fred Hutchinson Cancer Research Center.

### **WHI acknowledgements:**

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

## Supplementary Note 4: Whole genome sequencing and quality control in TOPMed

We used genetic data from whole genome sequencing via the TOPMed program

(freeze 8 release) (17). The TOPMed Data Coordinating Center (DCC) created a kinship matrix to

estimate recent genetic relatedness and provided genetic principal components (PCs) using the PC-Relate algorithm. Detailed information on genome sequencing, allele calling, and quality control in TOPMed can be found at <https://www.nhlbiwgs.org/topmed-whole-genome-sequencing-methods-freeze-8>. Only variants passing TOPMed QC with a minor allele frequency (MAF) > 0.01 were included. We further identified unrelated individuals based on the kinship matrix, ensuring they were degree-3 unrelated using the ppairPartition function from the GENESIS R package version v2.16.1

## Supplementary Note 5: All of Us phenotypes and acknowledgements

### **Clinical outcomes in All of Us:**

We identified previously-published manuscript that reported use of ICD codes to identify patients with a specific disease outcomes, with the exception of hypertension, for which we used blood pressure measures from the baseline All of Us exam. Details and references are below.

To define the condition of OSA, we used the following ICD-9 and ICD-10 codes: 780.53, 327.20, G47.39, G47.3, 780.51, 327.29, 780.57, 327.23, and G47.33. Snoring (ICD-10: R06.83) was excluded from the control group (18).

Hypertension was defined according to AHA guidelines as SBP  $\geq$  130, DBP  $\geq$  80, or use of antihypertensive medication. We extracted SBP and DBP measurements from physical exam data for all participants. The earliest measurements were used. Antihypertensive medication usage was obtained from either electronic health records (EHR) or physical exam data, using

the closest date to the BP measurement. Antihypertensive medications included peripheral vasodilators, agents acting on the renin-angiotensin system, beta-blocking agents, antihypertensives, calcium channel blockers, and diuretics. SBP and DBP values were increased by 15 mmHg and 10 mmHg, respectively (19).

For stroke, we used the following codes: hemorrhagic stroke (ICD-9: 431; ICD-10: I61.XX), ischemic stroke (ICD-9: 433.XX-434.XX, 436; ICD-10: I63.XXX), and transient ischemic attack (TIA) (ICD-9: 435.X; ICD-10: G45.9) (20).

Type 2 diabetes was identified using the ICD-9 code E11.X (21).

Atrial fibrillation was identified using ICD-9 code 427.3 and ICD-10 code I48.X (21).

For asthma, we used the following ICD-10 codes: J45.2, J45.3, J45.4, J45.5, J45.9, J45.990, J45.991, and J45.998 (22).

For chronic obstructive pulmonary disease (COPD), we used ICD-9 codes 491.X, 492.X, 496, and ICD-10 codes J41.X, J42.X, J43.X, and J44.X (23).

Chronic kidney disease (CKD) was identified using the following codes: chronic renal failure (ICD-9: 403.11, 404.12, 404.13, 404.92, 585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9, 586, 587), systemic disease causing CKD (ICD-9: 203.0, 277.3, 287.0, 446.0, 446.2, 446.4, 446.6, 710.0, 710.1), and ICD-10 codes for CKD (N18.1, N18.2, N18.3, N18.4, N18.5, N18.6, N18.9) (24).

For coronary artery disease (CAD), we used ICD-9 codes 410.X, 411.0, 412.X, 429.79, and ICD-10 codes I21.X, I22.X, I23.X, I24.1, I25.2 (21).

Heart failure was identified using ICD-10 codes I50.2, I50.3, and ICD-9 code 428.X (25,26).

**All of Us acknowledgements:**

The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Quali- fied Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. The All of Us Research Program would not be possible without the partnership of its participants.

## Supplementary Note 6: FinnGen full ethics statement and acknowledgement

**Ethics statement:**

Study subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The

Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4<sup>th</sup> July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 10 include: THL Biobank BB2017\_55, BB2017\_111, BB2018\_19, BB\_2018\_34, BB\_2018\_67, BB2018\_71, BB2019\_7, BB2019\_8, BB2019\_26, BB2020\_1, BB2021\_65, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/430/2021 §28, §29, HUS/150/2022 §12, §13, §14, §15, §16, §17, §18, §23, §58 and §59, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and amendments BB\_2021-0140, BB\_2021-0156 (August 26 2021, Feb 2 2022), BB\_2021-0169, BB\_2021-0179, BB\_2021-0161, AB20-5926 and amendment #1 (April 23 2020) and its modification (Sep 22 2021), BB\_2022-0262, BB\_2022-0256, Biobank Borealis of Northern Finland\_2017\_1013, 2021\_5010, 2021\_5018, 2021\_5015, 2021\_5015 Amendment, 2021\_5023, 2021\_5023 Amendment, 2021\_5017,

2022\_6001, 2022\_6006 Amendment, BB22-0067, 2022\_0262, Biobank of Eastern Finland 1186/2018 and amendment 22§/2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, 27§/2022, 28§/2022, 29§/2022, 33§/2022, 35§/2022, 36§/2022, 37§/2022, 39§/2022, 7§/2023, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), 8§/2021, 9§/2021, 9§/2022, §10/2022, §12/2022, 13§/2022, §20/2022, §21/2022, §22/2022, §23/2022, 28§/2022, 29§/2022, 30§/2022, 31§/2022, 32§/2022, 38§/2022, 40§/2022, 42§/2022, 1§/2023, Central Finland Biobank 1-2017, BB\_2021-0161, BB\_2021-0169, BB\_2021-0179, BB\_2021-0170, BB\_2022-0256, and Terveystalo Biobank STB 2018001 and amendment 2<sup>5</sup>th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 1<sup>8</sup>th June 2021, Arctic biobank P0844: ARC\_2021\_1001.

#### **FinnGen acknowledgements:**

We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis Pharma AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern Finland (<https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere ([www.tays.fi/en-](http://www.tays.fi/en-)



[US/Research and development/Finnish Clinical Biobank Tampere](#)), Biobank of Eastern Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.ksshp.fi/fi-FI/Potilaalle/Biopankki](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank ([www.veripalvelu.fi/verenluovutus/biopankkitoiminta](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)), Terveystalo Biobank ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)) and Arctic Biobank (<https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank>). All Finnish Biobanks are members of BBMRI.fi infrastructure ([www.bbMRI.fi](http://www.bbMRI.fi)). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB.

## Supplementary Note 7: TOPMed and CCDG acknowledgments

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Genome sequencing for “NHLBI TOPMed: Genome sequencing for “NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study” (phs000974.v4.p3) was performed at the Broad Institute Genomics Platform (3R01HL092577-06S1, 3U54HG003067-12S2). Genome sequencing for “NHLBI TOPMed: The Jackson Heart Study” (phs000964.v1.p1) was performed at the Northwest Genomics Center (HHSN268201100037C). Genome sequencing for the “NHLBI TOPMed: The Atherosclerosis Risk in Communities Study” (phs001211.v3.p2) was performed at the Broad Institute Genomics Platform (3R01HL092577-06S1) and the Baylor College of Medicine Human

Genome Sequencing Center (HHSN268201500015C, 3U54HG003273-12S2). Genome sequencing for “NHLBI TOPMed: Cleveland Family Study” (phs000954.v3.p2) was performed at the Northwest Genomics Center (3R01HL098433-05S1, HHSN268201600032I). Genome sequencing for “NHLBI TOPMed: Genetic Epidemiology of COPD (COPDGene) in the TOPMed Program” (phs000951) was performed at the University of Washington Northwest Genomics Center (3R01 HL089856-08S1) and the Broad Institute of MIT and Harvard (HHSN268201500014C). Genomics sequencing for “NHLBI TOPMed: Cardiovascular Health Study” (phs001368.v2.p1) was performed at the Baylor College of Medicine Human Genome Sequencing Center (3U54HG003273-12S2, HHSN268201500015C, HHSN268201600033I). Genome sequencing for “NHLBI TOPMed: Hispanic Community Health Study/Study of Latinos” (phs001395.v1.p1) was performed at the Baylor College of Medicine Human Genome Sequencing Center (HHSN268201600033I). Genome sequencing for “NHLBI TOPMed: Women’s Health Initiative (WHI)” (phs001237.v2.p1) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C). Genome sequencing for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis” (phs001416.v2.p1) was performed at Broad Institute Genomics Platform (HHSN268201500014C, 3U54HG003067-13S1. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL- 120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data

for TOPMed. The Genome Sequencing Program (GSP) was funded by the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), and the National Eye Institute (NEI). The GSP Coordinating Center (U24 HG008956) contributed to cross-program scientific initiatives and provided logistical and general study coordination. The Centers for Common Disease Genomics (CCDG) program was supported by NHGRI and NHLBI, and whole genome sequencing was performed at the Baylor College of Medicine Human Genome Sequencing Center (UM1 HG008898 and R01HL059367).

## Supplementary Note 8: TOPMed consortium members

Namiko Abe<sup>1</sup>, Gonçalo Abecasis<sup>2</sup>, Francois Aguet<sup>3</sup>, Christine Albert<sup>4</sup>, Laura Almasy<sup>5</sup>, Alvaro Alonso<sup>6</sup>, Seth Ament<sup>7</sup>, Peter Anderson<sup>8</sup>, Pramod Anugu<sup>9</sup>, Deborah Applebaum-Bowden<sup>10</sup>, Kristin Ardlie<sup>3</sup>, Dan Arking<sup>11</sup>, Donna K Arnett<sup>12</sup>, Allison Ashley-Koch<sup>13</sup>, Stella Aslibekyan<sup>14</sup>, Tim Assimes<sup>15</sup>, Paul Auer<sup>16</sup>, Dimitrios Avramopoulos<sup>11</sup>, Najib Ayas<sup>17</sup>, Adithya Balasubramanian<sup>18</sup>, John Barnard<sup>19</sup>, Kathleen Barnes<sup>20</sup>, R. Graham Barr<sup>21</sup>, Emily Barron-Casella<sup>11</sup>, Lucas Barwick<sup>22</sup>, Terri Beaty<sup>11</sup>, Gerald Beck<sup>23</sup>, Diane Becker<sup>24</sup>, Lewis Becker<sup>11</sup>, Rebecca Beer<sup>25</sup>, Amber Beitelshes<sup>7</sup>, Emelia Benjamin<sup>26</sup>, Takis Benos<sup>27</sup>, Marcos Bezerra<sup>28</sup>, Larry Bielak<sup>2</sup>, Joshua Bis<sup>29</sup>, Thomas Blackwell<sup>2</sup>, John Blangero<sup>30</sup>, Eric Boerwinkle<sup>31</sup>, Donald W. Bowden<sup>32</sup>, Russell Bowler<sup>33</sup>, Jennifer Brody<sup>8</sup>, Ulrich Broeckel<sup>34</sup>, Jai Broome<sup>8</sup>, Deborah Brown<sup>35</sup>, Karen Bunting<sup>1</sup>, Esteban Burchard<sup>36</sup>, Carlos Bustamante<sup>37</sup>, Erin Buth<sup>38</sup>, Brian Cade<sup>39</sup>, Jonathan Cardwell<sup>40</sup>, Vincent Carey<sup>41</sup>, Julie Carrier<sup>42</sup>, April Carson<sup>43</sup>, Cara Carty<sup>44</sup>, Richard Casaburi<sup>45</sup>, Juan P Casas Romero<sup>46</sup>, James Casella<sup>11</sup>, Peter Castaldi<sup>47</sup>, Mark Chaffin<sup>3</sup>, Christy Chang<sup>7</sup>, Yi-Cheng Chang<sup>48</sup>, Daniel

Chasman<sup>49</sup>, Sameer Chavan<sup>40</sup>, Bo-Juen Chen<sup>1</sup>, Wei-Min Chen<sup>50</sup>, Yii-Der Ida Chen<sup>51</sup>, Michael Cho<sup>41</sup>, Seung Hoan Choi<sup>3</sup>, Lee-Ming Chuang<sup>52</sup>, Mina Chung<sup>53</sup>, Ren-Hua Chung<sup>54</sup>, Clary Clish<sup>55</sup>, Suzy Comhair<sup>56</sup>, Matthew Conomos<sup>38</sup>, Elaine Cornell<sup>57</sup>, Adolfo Correa<sup>58</sup>, Carolyn Crandall<sup>45</sup>, James Crapo<sup>59</sup>, L. Adrienne Cupples<sup>60</sup>, Joanne Curran<sup>61</sup>, Jeffrey Curtis<sup>62</sup>, Brian Custer<sup>63</sup>, Coleen Damcott<sup>7</sup>, Dawood Darbar<sup>64</sup>, Sean David<sup>65</sup>, Colleen Davis<sup>8</sup>, Michelle Daya<sup>40</sup>, Mariza de Andrade<sup>66</sup>, Lisa de las Fuentes<sup>67</sup>, Paul de Vries<sup>68</sup>, Michael DeBaun<sup>69</sup>, Ranjan Deka<sup>70</sup>, Dawn DeMeo<sup>41</sup>, Scott Devine<sup>7</sup>, Huyen Dinh<sup>18</sup>, Harsha Doddapaneni<sup>18</sup>, Qing Duan<sup>71</sup>, Shannon Dugan-Perez<sup>18</sup>, Ravi Duggirala<sup>72</sup>, Jon Peter Durda<sup>57</sup>, Susan K. Dutcher<sup>73</sup>, Charles Eaton<sup>74</sup>, Lynette Ekunwe<sup>9</sup>, Adel El Boueiz<sup>75</sup>, Patrick Ellinor<sup>76</sup>, Leslie Emery<sup>8</sup>, Serpil Erzurum<sup>19</sup>, Charles Farber<sup>50</sup>, Jesse Farek<sup>18</sup>, Tasha Fingerlin<sup>77</sup>, Matthew Flickinger<sup>2</sup>, Myriam Fornage<sup>31</sup>, Nora Franceschini<sup>78</sup>, Chris Frazar<sup>8</sup>, Mao Fu<sup>7</sup>, Stephanie M. Fullerton<sup>8</sup>, Lucinda Fulton<sup>79</sup>, Stacey Gabriel<sup>3</sup>, Weiniu Gan<sup>25</sup>, Shanshan Gao<sup>40</sup>, Yan Gao<sup>9</sup>, Margery Gass<sup>80</sup>, Heather Geiger<sup>81</sup>, Bruce Gelb<sup>82</sup>, Mark Geraci<sup>27</sup>, Soren Germer<sup>1</sup>, Robert Gerszten<sup>83</sup>, Auyon Ghosh<sup>41</sup>, Richard Gibbs<sup>18</sup>, Chris Gignoux<sup>15</sup>, Mark Gladwin<sup>27</sup>, David Glahn<sup>84</sup>, Stephanie Gogarten<sup>8</sup>, Da-Wei Gong<sup>7</sup>, Harald Goring<sup>85</sup>, Sharon Graw<sup>86</sup>, Kathryn J. Gray<sup>87</sup>, Daniel Grine<sup>40</sup>, Colin Gross<sup>2</sup>, C. Charles Gu<sup>79</sup>, Yue Guan<sup>7</sup>, Xiuqing Guo<sup>51</sup>, Namrata Gupta<sup>3</sup>, David M. Haas<sup>88</sup>, Jeff Haessler<sup>80</sup>, Michael Hall<sup>89</sup>, Yi Han<sup>18</sup>, Patrick Hanly<sup>90</sup>, Daniel Harris<sup>91</sup>, Nicola L. Hawley<sup>92</sup>, Jiang He<sup>93</sup>, Ben Heavner<sup>38</sup>, Susan Heckbert<sup>94</sup>, Ryan Hernandez<sup>36</sup>, David Herrington<sup>95</sup>, Craig Hersh<sup>96</sup>, Bertha Hidalgo<sup>14</sup>, James Hixson<sup>31</sup>, Brian Hobbs<sup>41</sup>, John Hokanson<sup>40</sup>, Elliott Hong<sup>7</sup>, Karin Hoth<sup>97</sup>, Chao (Agnes) Hsiung<sup>98</sup>, Jianhong Hu<sup>18</sup>, Yi-Jen Hung<sup>99</sup>, Haley Huston<sup>100</sup>, Chii Min Hwu<sup>101</sup>, Marguerite Ryan Irvin<sup>14</sup>, Rebecca Jackson<sup>102</sup>, Deepti Jain<sup>8</sup>, Cashell Jaquish<sup>103</sup>, Jill Johnsen<sup>104</sup>, Andrew Johnson<sup>25</sup>, Craig Johnson<sup>8</sup>, Rich Johnston<sup>6</sup>, Kimberly Jones<sup>11</sup>, Hyun Min Kang<sup>105</sup>, Robert Kaplan<sup>106</sup>, Sharon Kardia<sup>2</sup>, Shannon

Kelly<sup>36</sup>, Eimear Kenny<sup>82</sup>, Michael Kessler<sup>7</sup>, Alyna Khan<sup>8</sup>, Ziad Khan<sup>18</sup>, Wonji Kim<sup>107</sup>, John Kimoff<sup>108</sup>, Greg Kinney<sup>109</sup>, Barbara Konkle<sup>110</sup>, Charles Kooperberg<sup>80</sup>, Holly Kramer<sup>111</sup>, Christoph Lange<sup>112</sup>, Ethan Lange<sup>40</sup>, Leslie Lange<sup>113</sup>, Cathy Laurie<sup>8</sup>, Cecelia Laurie<sup>8</sup>, Meryl LeBoff<sup>41</sup>, Jiwon Lee<sup>41</sup>, Sandra Lee<sup>18</sup>, Wen-Jane Lee<sup>101</sup>, Jonathon LeFaive<sup>2</sup>, David Levine<sup>8</sup>, Dan Levy<sup>25</sup>, Joshua Lewis<sup>7</sup>, Xiaohui Li<sup>51</sup>, Yun Li<sup>71</sup>, Henry Lin<sup>51</sup>, Honghuang Lin<sup>114</sup>, Xihong Lin<sup>115</sup>, Simin Liu<sup>116</sup>, Yongmei Liu<sup>117</sup>, Yu Liu<sup>118</sup>, Ruth J.F. Loos<sup>119</sup>, Steven Lubitz<sup>76</sup>, Kathryn Lunetta<sup>114</sup>, James Luo<sup>25</sup>, Ulysses Magalang<sup>120</sup>, Michael Mahaney<sup>61</sup>, Barry Make<sup>11</sup>, Ani Manichaikul<sup>50</sup>, Alisa Manning<sup>121</sup>, JoAnn Manson<sup>41</sup>, Lisa Martin<sup>122</sup>, Melissa Marton<sup>81</sup>, Susan Mathai<sup>40</sup>, Rasika Mathias<sup>11</sup>, Susanne May<sup>38</sup>, Patrick McArdle<sup>7</sup>, Merry-Lynn McDonald<sup>123</sup>, Sean McFarland<sup>107</sup>, Stephen McGarvey<sup>124</sup>, Daniel McGoldrick<sup>125</sup>, Caitlin McHugh<sup>38</sup>, Becky McNeil<sup>126</sup>, Hao Mei<sup>9</sup>, James Meigs<sup>127</sup>, Vipin Menon<sup>18</sup>, Luisa Mestroni<sup>86</sup>, Ginger Metcalf<sup>18</sup>, Deborah A Meyers<sup>128</sup>, Emmanuel Mignot<sup>129</sup>, Julie Mikulla<sup>25</sup>, Nancy Min<sup>9</sup>, Mollie Minear<sup>130</sup>, Ryan L Minster<sup>27</sup>, Braxton D. Mitchell<sup>7</sup>, Matt Moll<sup>47</sup>, Zeineen Momin<sup>18</sup>, May E. Montasser<sup>7</sup>, Courtney Montgomery<sup>131</sup>, Donna Muzny<sup>18</sup>, Josyf C Mychaleckyj<sup>50</sup>, Girish Nadkarni<sup>82</sup>, Rakhi Naik<sup>11</sup>, Take Naseri<sup>132</sup>, Pradeep Natarajan<sup>3</sup>, Sergei Nekhai<sup>133</sup>, Sarah C. Nelson<sup>38</sup>, Bonnie Neltner<sup>40</sup>, Caitlin Nessner<sup>18</sup>, Deborah Nickerson<sup>134</sup>, Osuji Nkechinyere<sup>18</sup>, Kari North<sup>71</sup>, Jeff O'Connell<sup>135</sup>, Tim O'Connor<sup>7</sup>, Heather Ochs-Balcom<sup>136</sup>, Geoffrey Okwuonu<sup>18</sup>, Allan Pack<sup>137</sup>, David T. Paik<sup>138</sup>, Nicholette Palmer<sup>139</sup>, James Pankow<sup>140</sup>, George Papanicolaou<sup>25</sup>, Cora Parker<sup>141</sup>, Gina Peloso<sup>142</sup>, Juan Manuel Peralta<sup>72</sup>, Marco Perez<sup>15</sup>, James Perry<sup>7</sup>, Ulrike Peters<sup>143</sup>, Patricia Peyser<sup>2</sup>, Lawrence S Phillips<sup>6</sup>, Jacob Pleiness<sup>2</sup>, Toni Pollin<sup>7</sup>, Wendy Post<sup>144</sup>, Julia Powers Becker<sup>145</sup>, Meher Preethi Boorgula<sup>40</sup>, Michael Preuss<sup>82</sup>, Bruce Psaty<sup>8</sup>, Pankaj Qasba<sup>25</sup>, Dandi Qiao<sup>41</sup>, Zhaohui Qin<sup>6</sup>, Nicholas Rafaels<sup>146</sup>, Laura Raffield<sup>147</sup>, Mahitha Rajendran<sup>18</sup>, Vasam S. Ramachandran<sup>114</sup>, D.C. Rao<sup>79</sup>, Laura Rasmussen-Torvik<sup>148</sup>, Aakrosh Ratan<sup>50</sup>, Susan Redline<sup>47</sup>,

Robert Reed<sup>7</sup>, Catherine Reeves<sup>149</sup>, Elizabeth Regan<sup>59</sup>, Alex Reiner<sup>150</sup>, Muagututi‘a Sefuiva Reupena<sup>151</sup>, Ken Rice<sup>8</sup>, Stephen Rich<sup>50</sup>, Rebecca Robillard<sup>152</sup>, Nicolas Robine<sup>81</sup>, Dan Roden<sup>153</sup>, Carolina Roselli<sup>3</sup>, Jerome Rotter<sup>154</sup>, Ingo Ruczinski<sup>11</sup>, Alexi Runnels<sup>81</sup>, Pamela Russell<sup>40</sup>, Sarah Ruuska<sup>100</sup>, Kathleen Ryan<sup>7</sup>, Ester Cerdeira Sabino<sup>155</sup>, Danish Saleheen<sup>21</sup>, Shabnam Salimi<sup>156</sup>, Sejal Salvi<sup>18</sup>, Steven Salzberg<sup>11</sup>, Kevin Sandow<sup>157</sup>, Vijay G. Sankaran<sup>158</sup>, Jireh Santibanez<sup>18</sup>, Karen Schwander<sup>79</sup>, David Schwartz<sup>40</sup>, Frank Scurba<sup>27</sup>, Christine Seidman<sup>159</sup>, Jonathan Seidman<sup>160</sup>, Frédéric Sériès<sup>161</sup>, Vivien Sheehan<sup>162</sup>, Stephanie L. Sherman<sup>163</sup>, Amol Shetty<sup>7</sup>, Aniket Shetty<sup>40</sup>, Wayne Hui- Heng Sheu<sup>101</sup>, M. Benjamin Shoemaker<sup>164</sup>, Brian Silver<sup>165</sup>, Edwin Silverman<sup>41</sup>, Robert Skomro<sup>166</sup>, Albert Vernon Smith<sup>167</sup>, Jennifer Smith<sup>2</sup>, Josh Smith<sup>8</sup>, Nicholas Smith<sup>94</sup>, Tanja Smith<sup>1</sup>, Sylvia Smoller<sup>106</sup>, Beverly Snively<sup>168</sup>, Michael Snyder<sup>15</sup>, Tamar Sofer<sup>41</sup>, Nona Sotoodehnia<sup>8</sup>, Adrienne M. Stilp<sup>8</sup>, Garrett Storm<sup>169</sup>, Elizabeth Streeten<sup>7</sup>, Jessica Lasky Su<sup>170</sup>, Yun Ju Sung<sup>79</sup>, Jody Sylvia<sup>41</sup>, Adam Szpiro<sup>8</sup>, Daniel Taliun<sup>2</sup>, Hua Tang<sup>171</sup>, Margaret Taub<sup>11</sup>, Kent D. Taylor<sup>172</sup>, Matthew Taylor<sup>86</sup>, Simeon Taylor<sup>7</sup>, Marilyn Telen<sup>13</sup>, Timothy A. Thornton<sup>8</sup>, Machiko Threlkeld<sup>173</sup>, Lesley Tinker<sup>174</sup>, David Tirschwell<sup>8</sup>, Sarah Tishkoff<sup>175</sup>, Hemant Tiwari<sup>176</sup>, Catherine Tong<sup>177</sup>, Russell Tracy<sup>178</sup>, Michael Tsai<sup>140</sup>, Dhananjay Vaidya<sup>11</sup>, David Van Den Berg<sup>179</sup>, Peter VandeHaar<sup>2</sup>, Scott Vrieze<sup>140</sup>, Tarik Walker<sup>40</sup>, Robert Wallace<sup>97</sup>, Avram Walts<sup>40</sup>, Fei Fei Wang<sup>8</sup>, Heming Wang<sup>180</sup>, Jiongming Wang<sup>167</sup>, Karol Watson<sup>45</sup>, Jennifer Watt<sup>18</sup>, Daniel E. Weeks<sup>27</sup>, Joshua Weinstock<sup>105</sup>, Bruce Weir<sup>8</sup>, Scott T Weiss<sup>181</sup>, Lu-Chen Weng<sup>76</sup>, Jennifer Wessel<sup>182</sup>, Cristen Willer<sup>62</sup>, Kayleen Williams<sup>38</sup>, L. Keoki Williams<sup>183</sup>, Carla Wilson<sup>41</sup>, James Wilson<sup>184</sup>, Lara Winterkorn<sup>81</sup>, Quenna Wong<sup>8</sup>, Joseph Wu<sup>138</sup>, Huichun Xu<sup>7</sup>, Lisa Yanek<sup>11</sup>, Ivana Yang<sup>40</sup>, Ketian Yu<sup>2</sup>, Seyedeh Maryam Zekavat<sup>3</sup>, Yingze Zhang<sup>185</sup>, Snow Xueyan Zhao<sup>59</sup>, Wei Zhao<sup>186</sup>, Xiaofeng Zhu<sup>187</sup>, Michael Zody<sup>1</sup>, Sebastian Zoellner<sup>2</sup>

1 - New York Genome Center, New York, New York; 2 - University of Michigan, Ann Arbor, Michigan; 3 - Broad Institute, Cambridge, Massachusetts; 4 - Cedars Sinai, Boston, Massachusetts; 5 - Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania; 6 - Emory University, Atlanta, Georgia; 7 - University of Maryland, Baltimore, Maryland; 8 - University of Washington, Seattle, Washington; 9 - University of Mississippi, Jackson, Mississippi; 10 - National Institutes of Health, Bethesda, Maryland; 11 - Johns Hopkins University, Baltimore, Maryland; 12 - University of Kentucky, Lexington, Kentucky; 13 - Duke University, Durham, North Carolina; 14 - University of Alabama, Birmingham, Alabama; 15 - Stanford University, Stanford, California; 16 - Medical College of Wisconsin, Milwaukee, Wisconsin; 17 - Medicine, Providence Health Care, Vancouver; 18 - Baylor College of Medicine Human Genome Sequencing Center, Houston, Texas; 19 - Cleveland Clinic, Cleveland, Ohio; 20 - Tempus, University of Colorado Anschutz Medical Campus, Aurora, Colorado; 21 - Columbia University, New York, New York; 22 - LTRC, The Emmes Corporation, Rockville, Maryland; 23 - Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio; 24 - Medicine, Johns Hopkins University, Baltimore, Maryland; 25 - National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland; 26 - Boston University School of Medicine, Boston University, Massachusetts General Hospital, Boston, Massachusetts; 27 - University of Pittsburgh, Pittsburgh, Pennsylvania; 28 - Fundação de Hematologia e Hemoterapia de Pernambuco - Hemope, Recife; 29 - Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington; 30 - Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, Texas; 31 - University of Texas Health at Houston, Houston, Texas; 32 - Department of Biochemistry, Wake Forest Baptist Health, Winston-Salem, North Carolina; 33 - National Jewish Health, National Jewish Health, Denver, Colorado; 34 - Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin; 35 - Pediatrics, University of Texas Health at Houston, Houston, Texas; 36 - University of California, San Francisco, San Francisco, California; 37 – Biomedical Data Science, Stanford University, Stanford, California; 38 - Biostatistics, University of Washington, Seattle, Washington; 39 - Brigham and Women's Hospital, Brigham & Women's Hospital, Boston, Massachusetts; 40 -

University of Colorado at Denver, Denver, Colorado; 41 - Brigham & Women's Hospital, Boston, Massachusetts; 42 - University of Montreal; 43 - Medicine, University of Mississippi, Jackson, Mississippi; 44 - Washington State University, Pullman, Washington; 45 - University of California, Los Angeles, Los Angeles, California; 46 - Brigham & Women's Hospital; 47 - Medicine, Brigham & Women's Hospital, Boston, Massachusetts; 48 - National Taiwan University, Taipei; 49 - Division of Preventive Medicine, Brigham & Women's Hospital, Boston, Massachusetts; 50 - University of Virginia, Charlottesville, Virginia; 51 - Lundquist Institute, Torrance, California; 52 - National Taiwan University Hospital, National Taiwan University, Taipei; 53 - Cleveland Clinic, Cleveland Clinic, Cleveland, Ohio; 54 - National Health Research Institute Taiwan, Miaoli County; 55 - Metabolomics Platform, Broad Institute, Cambridge, Massachusetts; 56 - Immunity and Immunology, Cleveland Clinic, Cleveland, Ohio; 57 - University of Vermont, Burlington, Vermont; 58 - Population Health Science, University of Mississippi, Jackson, Mississippi; 59 - National Jewish Health, Denver, Colorado; 60 - Biostatistics, Boston University, Boston, Massachusetts; 61 - University of Texas Rio Grande Valley School of Medicine, Brownsville, Texas; 62 - Internal Medicine, University of Michigan, Ann Arbor, Michigan; 63 - Vitalant Research Institute, San Francisco, California; 64 - University of Illinois at Chicago, Chicago, Illinois; 65 - University of Chicago, Chicago, Illinois; 66 - Health Quantitative Sciences Research, Mayo Clinic, Rochester, Minnesota; 67 - Department of Medicine, Cardiovascular Division, Washington University in St Louis, St. Louis, Missouri; 68 - Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health at Houston, Houston, Texas; 69 - Vanderbilt University, Nashville, Tennessee; 70 - University of Cincinnati, Cincinnati, Ohio; 71 - University of North Carolina, Chapel Hill, North Carolina; 72 - University of Texas Rio Grande Valley School of Medicine, Edinburg, Texas; 73 - Genetics, Washington University in St Louis, St Louis, Missouri; 74 - Brown University, Providence, Rhode Island; 75 - Channing Division of Network Medicine, Harvard University, Cambridge, Massachusetts; 76 - Massachusetts General Hospital, Boston, Massachusetts; 77 - Center for Genes, Environment and Health, National Jewish Health, Denver, Colorado; 78 - Epidemiology, University of North Carolina, Chapel Hill, North Carolina; 79 - Washington University in St Louis, St Louis, Missouri; 80 - Fred Hutchinson Cancer Research



Center, Seattle, Washington; 81 - New York Genome Center, New York City, New York; 82 - Icahn School of Medicine at Mount Sinai, New York, New York; 83 - Beth Israel Deaconess Medical Center, Boston, Massachusetts; 84 - Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts; 85 - University of Texas Rio Grande Valley School of Medicine, San Antonio, Texas; 86 - University of Colorado Anschutz Medical Campus, Aurora, Colorado; 87 - Obstetrics and Gynecology, Mass General Brigham, Boston, Massachusetts; 88 - OB/GYN, Indiana University, Indianapolis, Indiana; 89 - Cardiology, University of Mississippi, Jackson, Mississippi; 90 - Medicine, University of Calgary, Calgary; 91 - Genetics, University of Maryland, Philadelphia, Pennsylvania; 92 - Department of Chronic Disease Epidemiology, Yale University, New Haven, Connecticut; 93 - Tulane University, New Orleans, Louisiana; 94 - Epidemiology, University of Washington, Seattle, Washington; 95 - Wake Forest Baptist Health, Winston-Salem, North Carolina; 96 - Channing Division of Network Medicine, Brigham & Women's Hospital, Boston, Massachusetts; 97 - University of Iowa, Iowa City, Iowa; 98 - Institute of Population Health Sciences, NHRI, National Health Research Institute Taiwan, Miaoli County; 99 - Tri-Service General Hospital National Defense Medical Center; 100 - Blood Works Northwest, Seattle, Washington; 101 - Taichung Veterans General Hospital Taiwan, Taichung City; 102 - Internal Medicine, Division of Endocrinology, Diabetes and Metabolism, Oklahoma State University Medical Center, Columbus, Ohio; 103 - NHLBI, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland; 104 - Research Institute, Blood Works Northwest, Seattle, Washington; 105 - Biostatistics, University of Michigan, Ann Arbor, Michigan; 106 - Albert Einstein College of Medicine, New York, New York; 107 - Harvard University, Cambridge, Massachusetts; 108 - McGill University, Montreal; 109 - Epidemiology, University of Colorado at Denver, Aurora, Colorado; 110 - Medicine, Blood Works Northwest, Seattle, Washington; 111 - Public Health Sciences, Loyola University, Maywood, Illinois; 112 - Biostats, Harvard School of Public Health, Boston, Massachusetts; 113 - Medicine, University of Colorado at Denver, Aurora, Colorado; 114 - Boston University, Boston, Massachusetts; 115 - Harvard School of Public Health, Boston, Massachusetts; 116 - Epidemiology and Medicine, Brown University, Providence, Rhode Island; 117 - Cardiology, Duke University, Durham, North Carolina; 118 - Cardiovascular Institute, Stanford University,

Stanford, California; 119 - The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York; 120 - Division of Pulmonary, Critical Care and Sleep Medicine, Ohio State University, Columbus, Ohio; 121 - Broad Institute, Harvard University, Massachusetts General Hospital; 122 - cardiology, George Washington University, Washington, District of Columbia; 123 - University of Alabama at Birmingham, University of Alabama, Birmingham, Alabama; 124 - Epidemiology, Brown University, Providence, Rhode Island; 125 - Genome Sciences, University of Washington, Seattle, Washington; 126 - RTI International; 127 - Medicine, Massachusetts General Hospital, Boston, Massachusetts; 128 - University of Arizona, Tucson, Arizona; 129 - Center For Sleep Sciences and Medicine, Stanford University, Palo Alto, California; 130 - National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland; 131 - Genes and Human Disease, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma; 132 - Ministry of Health, Government of Samoa, Apia; 133 - Howard University, Washington, District of Columbia; 134 - Department of Genome Sciences, University of Washington, Seattle, Washington; 135 - University of Maryland, Baltimore, Maryland; 136 - University at Buffalo, Buffalo, New York; 137 - Division of Sleep Medicine/Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 138 - Stanford Cardiovascular Institute, Stanford University, Stanford, California; 139 - Biochemistry, Wake Forest Baptist Health, Winston-Salem, North Carolina; 140 - University of Minnesota, Minneapolis, Minnesota; 141 - Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, North Carolina; 142 - Department of Biostatistics, Boston University, Boston, Massachusetts; 143 - Fred Hutch and UW, Fred Hutchinson Cancer Research Center, Seattle, Washington; 144 - Cardiology/Medicine, Johns Hopkins University, Baltimore, Maryland; 145 - Medicine, University of Colorado at Denver, Denver, Colorado; 146 - CCPM, University of Colorado at Denver, Denver, Colorado; 147 - Genetics, University of North Carolina, Chapel Hill, North Carolina; 148 - Northwestern University, Chicago, Illinois; 149 - New York Genome Center, New York Genome Center, New York City, New York; 150 - Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington; 151 - Lutia I Puava Ae Mapu I Fagalele, Apia; 152 - Sleep Research Unit, University of Ottawa Institute for Mental Health Research, University

of Ottawa, Ottawa; 153 - Medicine, Pharmacology, Biomedical Informatics, Vanderbilt University, Nashville, Tennessee; 154 - Pediatrics, Lundquist Institute, Torrance, California; 155 - Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo; 156 - Pathology, University of Maryland, Seattle, Washington; 157 - TGPS, Lundquist Institute, Torrance, California; 158 - Division of Hematology/Oncology, Harvard University, Boston, Massachusetts; 159 - Genetics, Harvard Medical School, Boston, Massachusetts; 160 - Harvard Medical School, Boston, Massachusetts; 161 - Université Laval, Quebec City; 162 - Pediatrics, Emory University, Atlanta, Georgia; 163 - Human Genetics, Emory University, Atlanta, Georgia; 164 - Medicine/Cardiology, Vanderbilt University, Nashville, Tennessee; 165 - UMass Memorial Medical Center, Worcester, Massachusetts; 166 - University of Saskatchewan, Saskatoon; 167 - University of Michigan; 168 - Biostatistical Sciences, Wake Forest Baptist Health, Winston-Salem, North Carolina; 169 - Genomic Cardiology, University of Colorado at Denver, Aurora, Colorado; 170 - Channing Department of Medicine, Brigham & Women's Hospital, Boston, Massachusetts; 171 - Genetics, Stanford University, Stanford, California; 172 - Institute for Translational Genomics and Populations Sciences, Lundquist Institute, Torrance, California; 173 - University of Washington, Department of Genome Sciences, University of Washington, Seattle, Washington; 174 - Cancer Prevention Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington; 175 - Genetics, University of Pennsylvania, Philadelphia, Pennsylvania; 176 - Biostatistics, University of Alabama, Birmingham, Alabama; 177 - Department of Biostatistics, University of Washington, Seattle, Washington; 178 - Pathology & Laboratory Medicine, University of Vermont, Burlington, Vermont; 179 - USC Methylation Characterization Center, University of Southern California, University of Southern California, California; 180 - Brigham & Women's Hospital, Mass General Brigham, Boston, Massachusetts; 181 - Channing Division of Network Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, Massachusetts; 182 - Epidemiology, Indiana University, Indianapolis, Indiana; 183 - Henry Ford Health System, Detroit, Michigan; 184 - Cardiology, Beth Israel Deaconess Medical Center, Cambridge, Massachusetts; 185 - Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; 186 - Department of Epidemiology, University of Michigan, Ann Arbor, Michigan; 187 -

Department of Population and Quantitative Health Sciences, Case Western Reserve University,  
Cleveland, Ohio

## Supplementary Note 9: FinnGen consortium members

Aarno Palotie<sup>1,2</sup>, Mark Daly<sup>1</sup>, Bridget Riley-Gills<sup>3</sup>, Howard Jacob<sup>3</sup>, Coralie Viollet<sup>4,5</sup>, Slavé Petrovski<sup>4</sup>, Chia-Yen Chen<sup>6</sup>, Sally John<sup>6</sup>, George Okafo<sup>7</sup>, Robert Plenge<sup>8</sup>, Joseph Maranville<sup>8</sup>, Mark McCarthy<sup>9</sup>, Rion Pendergrass<sup>9</sup>, Jonathan Davitte<sup>10</sup>, Kirsi Auro<sup>11</sup>, Simonne Longerich<sup>12</sup>, Anders Mälarstig<sup>13</sup>, Anna Vlahiotis<sup>13</sup>, Katherine Klinger<sup>14</sup>, Clement Chatelain<sup>14</sup>, Jorg Blankenstein<sup>14</sup>, Karol Estrada<sup>15</sup>, Robert Graham<sup>15</sup>, Dawn Waterworth<sup>16</sup>, Chris O'Connell<sup>17</sup>, Nicole Renaud<sup>17</sup>, Tomi P. Mäkelä<sup>18</sup>, Jaakko Kaprio<sup>2</sup>, Minna Ruddock<sup>19</sup>, Petri Virolainen<sup>20</sup>, Antti Hakanen<sup>20</sup>, Terhi Kilpi<sup>21</sup>, Markus Perola<sup>21</sup>, Jukka Partanen<sup>22</sup>, Taneli Raivio<sup>23</sup>, Jani Tikkanen<sup>24</sup>, Raisa Serpi<sup>24</sup>, Kati Kristiansson<sup>21,25</sup>, Veli-Matti Kosma<sup>26</sup>, Jari Laukkanen<sup>27,28</sup>, Marco Hautalahti<sup>29,30</sup>, Outi Tuovila<sup>31</sup>, Jeffrey Waring<sup>3</sup>, Bridget Riley-Gillis<sup>3</sup>, Fedik Rahimov<sup>3</sup>, Ioanna Tachmazidou<sup>4</sup>, Zhihao Ding<sup>7</sup>, Marc Jung<sup>7</sup>, Hanati Tuoken<sup>7</sup>, Shameek Biswas<sup>8</sup>, Neha Raghavan<sup>12</sup>, Adriana Huertas-Vazquez<sup>12</sup>, Jae-Hoon Sul<sup>12</sup>, Xinli Hu<sup>13</sup>, Åsa Hedman<sup>13</sup>, Ma'en Obeidat<sup>17</sup>, Jonathan Chung<sup>17</sup>, Jonas Zierer<sup>17</sup>, Mari Niemi<sup>17</sup>, Samuli Ripatti<sup>2</sup>, Johanna Schleutker<sup>32,33</sup>, Mikko Arvas<sup>22</sup>, Olli Carpén<sup>23,34</sup>, Reetta Hinttala<sup>24</sup>, Johannes Kettunen<sup>24,35</sup>, Arto Mannermaa<sup>26</sup>, Katriina Aalto-Setälä<sup>36</sup>, Mika Kähönen<sup>25,37</sup>, Johanna Mäkelä<sup>29,30</sup>, Reetta Kälviäinen<sup>38</sup>, Valtteri Julkunen<sup>38</sup>, Hilikka Soininen<sup>38</sup>, Anne Remes<sup>35</sup>, Mikko Hiltunen<sup>39</sup>, Jukka Peltola<sup>37</sup>, Minna Raivio<sup>34</sup>, Pentti Tienari<sup>34</sup>, Juha Rinne<sup>33</sup>, Roosa Kallionpää<sup>33</sup>, Juulia Partanen<sup>40</sup>, Adam Ziemann<sup>3</sup>, Nizar Smaoui<sup>3</sup>, Anne Lehtonen<sup>3</sup>, Susan Eaton<sup>6</sup>, Heiko Runz<sup>6</sup>, Sanni Lahdenperä<sup>6</sup>, Natalie Bowers<sup>9</sup>, Edmond Teng<sup>9</sup>, Fanli Xu<sup>41</sup>, David Pulford<sup>42</sup>, Laura Addis<sup>41</sup>, John Eicher<sup>41</sup>, Qingqin S Li<sup>43</sup>, Karen He<sup>16</sup>, Ekaterina Khramtsova<sup>16</sup>, Martti Färkkilä<sup>34</sup>, Jukka Koskela<sup>34</sup>, Sampsa Pikkarainen<sup>34</sup>, Airi Jussila<sup>37</sup>, Katri Kaukinen<sup>37</sup>, Timo Blomster<sup>35</sup>, Mikko Kiviniemi<sup>38</sup>, Markku Voutilainen<sup>33</sup>, Tim Lu<sup>9</sup>, Linda McCarthy<sup>41</sup>, Amy Hart<sup>16</sup>, Meijian Guan<sup>16</sup>, Jason Miller<sup>12</sup>, Kirsi Kalpala<sup>13</sup>, Melissa Miller<sup>13</sup>, Kari Eklund<sup>34</sup>, Antti Palomäki<sup>33</sup>, Pia Isomäki<sup>37</sup>, Laura Pirilä<sup>33</sup>, Oili Kaipainen-Seppänen<sup>38</sup>, Johanna Huhtakangas<sup>35</sup>, Nina Mars<sup>2</sup>, Apinya Lertratanakul<sup>3</sup>, Marla Hochfeld<sup>8</sup>, Jorge Esparza Gordillo<sup>41</sup>, Fabiana Farias<sup>12</sup>, Nan Bing<sup>13</sup>, Tarja Laitinen<sup>2,37</sup>, Margit Pelkonen<sup>38</sup>, Paula Kauppi<sup>34</sup>, Hannu Kankaanranta<sup>44</sup>, Terttu Harju<sup>35</sup>, Riitta Lahesmaa<sup>33</sup>, Hubert Chen<sup>9</sup>, Joanna Betts<sup>41</sup>, Rajashree Mishra<sup>41</sup>, Majd Mouded<sup>45</sup>, Debby Ngo<sup>45</sup>, Teemu Niiranen<sup>46</sup>, Felix Vaura<sup>46</sup>, Veikko Salomaa<sup>46</sup>, Kaj Metsärinne<sup>33</sup>, Jenni Aittokallio<sup>33</sup>, Jussi Hernesniemi<sup>37</sup>, Daniel Gordin<sup>34</sup>, Juha Sinisalo<sup>34</sup>, Marja-Riitta Taskinen<sup>34</sup>, Tiinamaija Tuomi<sup>34</sup>, Timo Hiltunen<sup>34</sup>, Amanda Elliott<sup>47</sup>, Mary Pat Reeve<sup>2</sup>, Sanni Ruotsalainen<sup>2</sup>, Dirk Paul<sup>4</sup>, Audrey Chu<sup>41</sup>, Dermot Reilly<sup>48</sup>, Mike Mendelson<sup>49</sup>, Jaakko Parkkinen<sup>13</sup>, Tuomo Meretoja<sup>50</sup>, Heikki Joensuu<sup>51</sup>, Johanna Mattson<sup>34</sup>, Eveliina Salminen<sup>34</sup>, Annika Auranen<sup>52</sup>, Peeter Karihtala<sup>51</sup>, Päivi Auvinen<sup>38</sup>,

Klaus Elenius<sup>33</sup>, Esa PitkÄännen<sup>2</sup>, Relja Popovic<sup>3</sup>, Margarete Fabre<sup>5</sup>, Jennifer Schutzman<sup>9</sup>, Diptee Kulkarni<sup>41</sup>, Alessandro Porello<sup>16</sup>, Andrey Loboda<sup>12</sup>, Heli Lehtonen<sup>13</sup>, Stefan McDonough<sup>13</sup>, Sauli Vuoti<sup>53</sup>, Kai Kaarniranta<sup>54</sup>, Joni A Turunen<sup>55</sup>, Terhi Ollila<sup>34</sup>, Hannu Uusitalo<sup>37</sup>, Juha Karjalainen<sup>2</sup>, Mengzhen Liu<sup>3</sup>, Stephanie Loomis<sup>6</sup>, Erich Strauss<sup>9</sup>, Hao Chen<sup>9</sup>, Kaisa Tasanen<sup>35</sup>, Laura Huilaja<sup>35</sup>, Katariina Hannula-Jouppi<sup>34</sup>, Teea Salmi<sup>37</sup>, Sirkku Peltonen<sup>33</sup>, Leena Koulu<sup>33</sup>, David Choy<sup>9</sup>, Ying Wu<sup>13</sup>, Pirkko Pussinen<sup>34</sup>, Aino Salminen<sup>34</sup>, Tuula Salo<sup>34</sup>, David Rice<sup>34</sup>, Pekka Nieminen<sup>34</sup>, Ulla Palotie<sup>34</sup>, Maria Siponen<sup>38</sup>, Liisa Suominen<sup>38</sup>, Päivi Mäntylä<sup>38</sup>, Ulvi Gursoy<sup>33</sup>, Vuokko Anttonen<sup>35</sup>, Kirsi Sipilä<sup>56</sup>, Hannele Laivuori<sup>2</sup>, Venla Kurra<sup>37</sup>, Laura Kotaniemi-Talonen<sup>37</sup>, Oskari Heikinheimo<sup>34</sup>, Ilkka Kalliala<sup>34</sup>, Lauri Aaltonen<sup>34</sup>, Varpu Jokimaa<sup>33</sup>, Marja VäärÄäsmäki<sup>35</sup>, Outi Uimari<sup>35</sup>, Laure Morin-Papunen<sup>35</sup>, Maarit Maarit<sup>35</sup>, Terhi Pilttonen<sup>35</sup>, Katja Kivinen<sup>2</sup>, Elisabeth Widen<sup>2</sup>, Taru Tukiainen<sup>2</sup>, Niko Välimäki<sup>57</sup>, Eija Laakkonen<sup>58</sup>, Jaakko Tyrmi<sup>59</sup>, Heidi Silven<sup>60</sup>, Eeva Sliz<sup>60</sup>, Riikka Arffman<sup>60</sup>, Susanna Savukoski<sup>60</sup>, Triin Laisk<sup>61</sup>, Natalia Pujol<sup>61</sup>, Janet Kumar<sup>10</sup>, Iiris Hovatta<sup>62</sup>, Erkki Isometsä<sup>34</sup>, Hanna Ollila<sup>2</sup>, Jaana Suvisaari<sup>46</sup>, Antti Mäkitie<sup>63</sup>, Argyro Bizaki-Vallaskangas<sup>37</sup>, Sanna Toppila-Salmi<sup>64</sup>, Tytti Willberg<sup>33</sup>, Elmo Saarentaus<sup>2</sup>, Antti Aarnisalo<sup>34</sup>, Elisa Rahikkala<sup>35</sup>, Kristiina Aittomäki<sup>65</sup>, Fredrik Åberg<sup>66</sup>, Mitja Kurki<sup>67</sup>, Aki Havulinna<sup>68</sup>, Juha Mehtonen<sup>2</sup>, Priit Palta<sup>2</sup>, Shabbeer Hassan<sup>2</sup>, Pietro Della Briotta Parolo<sup>2</sup>, Wei Zhou<sup>69</sup>, Mutaamba Maasha<sup>69</sup>, Susanna Lemmelä<sup>2</sup>, Manuel Rivas<sup>70</sup>, Aoxing Liu<sup>2</sup>, Arto Lehisto<sup>2</sup>, Andrea Ganna<sup>2</sup>, Vincent Llorens<sup>2</sup>, Henrike Heyne<sup>2</sup>, Joel Rämö<sup>2</sup>, Satu Strausz<sup>2</sup>, Tuula Palotie<sup>71</sup>, Kimmo Palin<sup>57</sup>, Javier Garcia-Tabuenca<sup>72</sup>, Harri Siirtola<sup>72</sup>, Tuomo Kiiskinen<sup>2</sup>, Jiwoo Lee<sup>67</sup>, Kristin Tsuo<sup>67</sup>, Kati HyvÄärinen<sup>73</sup>, Jarmo Ritari<sup>73</sup>, Katri Pylkäs<sup>60</sup>, Minna Karjalainen<sup>60</sup>, Tuomo Mantere<sup>24</sup>, Eeva Kangasniemi<sup>25</sup>, Sami Heikkinen<sup>39</sup>, Nina Pitkänen<sup>20</sup>, Samuel Lessard<sup>14</sup>, Clément Chatelain<sup>14</sup>, Lila Kallio<sup>20</sup>, Tiina Wahlfors<sup>21</sup>, Eero Punkka<sup>23</sup>, Sanna Siltanen<sup>25</sup>, Tiina Jokela<sup>27</sup>, Anu Jalanko<sup>2</sup>, Auli Toivola<sup>2,21</sup>, Hwei-Yi Shen<sup>2</sup>, Risto Kajanne<sup>2</sup>, Rodos Rodosthenous<sup>2</sup>, Mervi Aavikko<sup>2</sup>, Helen Cooper<sup>2</sup>, Denise Öller<sup>2</sup>, Rasko Leinonen<sup>74</sup>, Henna Palin<sup>25</sup>, Malla-Maria Linna<sup>23</sup>, Masahiro Kanai<sup>69</sup>, Zhili Zheng<sup>69</sup>, L. Elisa Lahtela<sup>2</sup>, Mari Kaunisto<sup>2</sup>, Elina Kilpeläinen<sup>2</sup>, Tianduanyi Wang<sup>2</sup>, Timo P. Sipilä<sup>2</sup>, Oluwaseun Alexander Dada<sup>2</sup>, Awaisa Ghazal<sup>2</sup>, Anastasia Kytölä<sup>2</sup>, Rigbe Weldatsadik<sup>2</sup>, Jaska Uimonen<sup>2</sup>, Kati Donner<sup>2</sup>, Anu Loukola<sup>23</sup>, PÄäxivi Laiho<sup>21</sup>, Tuuli Sistonen<sup>21</sup>, Essi Kaiharju<sup>21</sup>, Markku Laukkanen<sup>21</sup>, Elina Järvensivu<sup>21</sup>, Sini Lähteenmäki<sup>21</sup>, Lotta Männikkö<sup>21</sup>, Regis Wong<sup>21</sup>, Minna Brunfeldt<sup>21</sup>, Sami Koskelainen<sup>21</sup>, Tero Hiekkalinna<sup>21</sup>, Teemu Paajanen<sup>21</sup>, Shuang Luo<sup>2</sup>, Shanmukha Sampath Padmanabhuni<sup>2</sup>, Marianna Niemi<sup>72</sup>, Javier Gracia-Tabuenca<sup>72</sup>, Mika Helminen<sup>72</sup>, Tiina Luukkaala<sup>72</sup>, Iida Vähätalo<sup>72</sup>, Iina Laak<sup>2</sup>, Saija Haapa-Paananen<sup>30</sup>, Sarah Smith<sup>30</sup>, Tom Southerington<sup>30</sup>, Meri Lähteenmäki<sup>30</sup>

1 - Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; Broad Institute of MIT and Harvard; Massachusetts General Hospital; 2 - Institute for Molecular

Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; 3 - Abbvie, Chicago, IL, United States; 4 - Astra Zeneca, Cambridge, United Kingdom; 5 - AstraZeneca, Cambridge, United Kingdom; 6 - Biogen, Cambridge, MA, United States; 7 - Boehringer Ingelheim, Ingelheim am Rhein, Germany; 8 - Bristol Myers Squibb, New York, NY, United States; 9 - Genentech, San Francisco, CA, United States; 10 - GlaxoSmithKline, Collegeville, PA, United States; 11 - GlaxoSmithKline, Espoo, Finland; 12 - Merck, Kenilworth, NJ, United States; 13 - Pfizer, New York, NY, United States; 14 - Translational Sciences, Sanofi R&D, Framingham, MA, USA; 15 - Maze Therapeutics, San Francisco, CA, United States; 16 - Janssen Research & Development, LLC, Spring House, PA, United States; 17 - Novartis Institutes for BioMedical Research, Cambridge, MA, United States; 18 - HiLIFE, University of Helsinki, Finland, Finland; 19 - Arctic biobank / University of Oulu; 20 - Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland; 21 - THL Biobank / Finnish Institute for Health and Welfare (THL), Helsinki, Finland; 22 - Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland; 23 - Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki; 24 - Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland; 25 - Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, Tampere, Finland; 26 - Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland; 27 - Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland; 28 - Central Finland Health Care District, Jyväskylä, Finland; 29 - FINBB - Finnish biobank cooperative; 30 - Finnish Biobank Cooperative - FINBB; 31 - Business Finland, Helsinki, Finland; 32 - Auria Biobank / Univ. of Turku / Hospital District of Southwest Finland, Turku, Finland; 33 - Hospital District of Southwest Finland, Turku, Finland; 34 - Hospital District of Helsinki and Uusimaa, Helsinki, Finland; 35 - Northern Ostrobothnia Hospital District, Oulu, Finland; 36 - Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; 37 - Pirkanmaa Hospital District, Tampere, Finland; 38 - Northern Savo Hospital District, Kuopio, Finland; 39 - University of Eastern Finland, Kuopio, Finland; 40 - Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland; 41 - GlaxoSmithKline, Brentford, United Kingdom; 42 - GlaxoSmithKline, Stevenage, United Kingdom; 43 - Janssen Research & Development, LLC, Titusville, NJ 08560, United States; 44 - University of Gothenburg, Gothenburg, Sweden/ Seinäjoki Central Hospital, Seinäjoki, Finland/ Tampere University, Tampere, Finland; 45 - Novartis, Basel, Switzerland; 46 - Finnish Institute for Health and Welfare (THL), Helsinki, Finland; 47 - Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; Broad Institute, Cambridge, MA, USA and Massachusetts General Hospital, Boston, MA, USA; 48 - Janssen Research & Development, LLC, Boston, MA, United

States; 49 - Novartis, Boston, MA, United States; 50 - Department of Breast Surgery, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland; 51 - Department of Oncology, Helsinki University Hospital Comprehensive Cancer Center and University of Helsinki, Helsinki, Finland; 52 - Pirkanmaa Hospital District, Tampere, Finland; 53 - Janssen-Cilag Oy, Espoo, Finland; 54 - Northern Savo Hospital District, Kuopio, Finland; Department of Molecular Genetics, University of Lodz, Lodz, Poland; 55 - Helsinki University Hospital and University of Helsinki, Helsinki, Finland; Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland; 56 - Research Unit of Oral Health Sciences Faculty of Medicine, University of Oulu, Oulu, Finland; Medical Research Center, Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland; 57 - University of Helsinki, Helsinki, Finland; 58 - University of Jyväskylä, Jyväskylä, Finland; 59 - University of Oulu, Oulu, Finland / University of Tampere, Tampere, Finland; 60 - University of Oulu, Oulu, Finland; 61 - Estonian biobank, Tartu, Estonia; 62 - University of Helsinki, Helsinki, Finland; 63 - Department of Otorhinolaryngology - Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 64 - University of Eastern Finland and Kuopio University Hospital, Department of Otorhinolaryngology, Kuopio, Finland and Department of Allergy, Helsinki University Hospital and University of Helsinki, Helsinki, Finland; 65 - Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 66 - Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki University, Helsinki, Finland; 67 - Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; Broad Institute, Cambridge, MA, United States; 68 - Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; Finnish Institute for Health and Welfare (THL), Helsinki, Finland; 69 - Broad Institute, Cambridge, MA, United States; 70 - University of Stanford, Stanford, CA, United States; 71 - University of Helsinki and Hospital District of Helsinki and Uusimaa, Helsinki, Finland; 72 - University of Tampere, Tampere, Finland; 73 - Finnish Red Cross Blood Service, Helsinki, Finland; 74 - Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland; European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK

## References

1. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019 Jul 1;8(7).

2. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, et al. The ARIC (atherosclerosis risk in communities) study: JACC focus seminar 3/8. *J Am Coll Cardiol*. 2021 Jun 15;77(23):2939–59.
3. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol*. 1989 Apr;129(4):687–702.
4. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105–16.
5. Redline S, Tishler PV, Tosteson TD, Williamson J, Kump K, Browner I, et al. The familial aggregation of obstructive sleep apnea. *Am J Respir Crit Care Med*. 1995 Mar;151(3 Pt 1):682–7.
6. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991 Feb;1(3):263–76.
7. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010 Feb;7(1):32–43.
8. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci*. 1963 May 22;107:539–56.
9. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007 Jun 1;165(11):1328–35.
10. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979 Sep;110(3):281–90.
11. Lavange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010 Aug;20(8):642–9.
12. Sorlie PD, Allison MA, Avilés-Santa ML, Cai J, Daviglius ML, Howard AG, et al. Prevalence of hypertension, awareness, treatment, and control in the Hispanic Community Health Study/Study of Latinos. *Am J Hypertens*. 2014 Jun;27(6):793–800.



13. Pirzada A, Cai J, Heiss G, Sotres-Alvarez D, Gallo LC, Youngblood ME, et al. Evolving science on cardiovascular disease among hispanic/latino adults: JACC international. *J Am Coll Cardiol*. 2023 Apr 18;81(15):1505–20.
14. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH, et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. *Ethn Dis*. 2003;13(4):438–55.
15. Taylor HA, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. 2005;15(4 Suppl 6):S6-4.
16. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. 2002 Nov 1;156(9):871–81.
17. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290–9.
18. Keenan BT, Kirchner HL, Veatch OJ, Borthwick KM, Davenport VA, Feemster JC, et al. Multisite validation of a simple electronic health record algorithm for identifying diagnosed obstructive sleep apnea. *J Clin Sleep Med*. 2020 Feb 15;16(2):175–83.
19. Kurniansyah N, Goodman MO, Khan AT, Wang J, Feofanova E, Bis JC, et al. Evaluating the use of blood pressure polygenic risk scores across race/ethnic background groups. *Nat Commun*. 2023 Jun 2;14(1):3202.
20. Kogan E, Twyman K, Heap J, Milentijevic D, Lin JH, Alberts M. Assessing stroke severity using electronic health record data: a machine learning approach. *BMC Med Inform Decis Mak*. 2020 Jan 8;20(1):8.
21. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018 Sep;50(9):1219–24.
22. Ni J, Friedman H, Boyd BC, McGurn A, Babinski P, Markossian T, et al. Early antibiotic exposure and development of asthma and allergic rhinitis in childhood. *BMC Pediatr*. 2019 Jul 5;19(1):225.
23. Gothe H, Rajsic S, Vukicevic D, Schoenfelder T, Jahn B, Geiger-Gritsch S, et al. Algorithms to identify COPD in health systems with and without access to ICD coding: a systematic review. *BMC Health Serv Res*. 2019 Oct 22;19(1):737.

24. Jalal K, Anand EJ, Venuto R, Eberle J, Arora P. Can billing codes accurately identify rapidly progressing stage 3 and stage 4 chronic kidney disease patients: a diagnostic test study. *BMC Nephrol.* 2019 Jul 12;20(1):260.
25. Goyal P, Bose B, Creber RM, Krishnan U, Yang M, Brady J, et al. Performance of electronic health record diagnosis codes for ambulatory heart failure encounters. *J Card Fail.* 2020 Dec;26(12):1060–6.
26. Zhang Z, Cao L, Chen R, Zhao Y, Lv L, Xu Z, et al. Electronic healthcare records and external outcome data for hospitalized patients with heart failure. *Sci Data.* 2021 Feb 5;8(1):46.