

1 **An integrated multi-omics analysis of sleep-disordered breathing traits across multiple**
2 **blood cell types**

3
4 Nuzulul Kurniansyah¹, Danielle A Wallace¹, Ying Zhang¹, Bing Yu¹², Brian Cade,^{1,7,20} Heming
5 Wang^{1,7,20} Heather M. Ochs-Balcom⁴, Alexander P Reiner^{5,6}, Alberto R Ramos¹⁰, Joshua D
6 Smith¹¹, Jianwen Cai¹⁴, Martha Daviglus¹⁵, Phyllis C Zee¹⁶, Robert Kaplan^{21,5}, Charles
7 Kooperberg⁵, Stephen S Rich¹⁸, Jerome I Rotter⁹, Sina A. Gharib¹⁹, Susan Redline,^{1,20} Tamar
8 Sofer^{1,2,*}

9
10 ¹Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA
11 ²Departments of Medicine and of Biostatistics, Harvard University, Boston, MA, USA
12 ⁴Department of Epidemiology and Environmental Health, School of Public Health and Health
13 Professions, University at Buffalo, The State University of New York, Buffalo, NY, USA
14 ⁵Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, Washington, USA
15 ⁶Department of Epidemiology, University of Washington, Seattle, Washington, USA
16 ⁷The Broad Institute of MIT and Harvard, Cambridge, MA, USA
17 ⁸Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of
18 Vermont, Burlington, VT, USA
19 ⁹The Institute for Translational Genomics and Population Sciences, Department of Pediatrics,
20 The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance,
21 CA USA
22 ¹⁰Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
23 ¹¹Northwest Genomic Center, University of Washington, Seattle, WA, USA
24 ¹² Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Pub-
25 lic Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
26 ¹³Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
27 ¹⁴Department of Biostatistics, University of North Carolina, at Chapel Hill, NC, USA
28 ¹⁵Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, USA
29 ¹⁶Division of Sleep Medicine, Department of Neurology, Northwestern University, Chicago, IL,
30 USA.
31 ¹⁷Divisions of Cardiology and Neurology, Duke University Medical Center, Durham, NC, 27710,
32 USA
33 ¹⁸Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville,
34 VA, USA
35 ¹⁹Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Depart-
36 ment of Medicine, University of Washington, Seattle, WA, USA
37 ²⁰Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA
38 ²¹Department of Epidemiology & Population Health, Department of Pediatrics, Albert Einstein
39 College of Medicine, Bronx, NY, USA

40
41 *Correspondence:
42 Tamar Sofer

43 tsofer@bwh.harvard.edu

44 Suite 225C

45 221 Longwood Ave

46 Boston, MA, 02115

47

48

49

50 **ABSTRACT**

51 **Background:** Sleep Disordered Breathing (SDB) is characterized by repeated breathing
52 reductions or cessations during sleep, often accompanied by oxyhemoglobin desaturation. How
53 SDB affects the molecular environment is still poorly understood.

54 **Methods:** We studied the association of three SDB measures: the Apnea Hypopnea Index
55 (AHI), average and minimum oxyhemoglobin saturation during sleep (AvgO2 and MinO2) with
56 gene expression measured using RNA-seq in peripheral blood mononuclear cells (PBMCs),
57 monocytes, and T-cells, in ~500 individuals from the Multi-Ethnic Study of Atherosclerosis
58 (MESA). We developed genetic instrumental variables (IVs) for the associated transcripts as
59 polygenic risk scores (tPRS), then generalized and validated the tPRS in the Women's Health
60 Initiative (WHI). Next, we constructed the tPRS and studied their association with SDB
61 measures (to identify potential reverse causal associations) and with serum metabolites (to
62 identify downstream effects) in ~12,000 and ~4,000 participants, respectively, from the Hispanic
63 Community Health Study/Study of Latinos (HCHS/SOL). Finally, we estimated the association of
64 these SDB measures with transcript IV-associated metabolites in HCHS/SOL, to verify complete
65 association pathways linking SDB, gene expression, and metabolites.

66 **Results:** Across the three leukocyte cell types, 96 gene transcripts were associated with at
67 least one SDB exposure (False Discovery Rate (FDR) p-value <0.1). Across cell populations,
68 estimated log-fold expression changes were similar between AHI and MinO2 (Spearman
69 correlations>0.90), and less similar between AvgO2 and the other exposures. Eight and four

70 associations had FDR p -value <0.05 when the analysis was not adjusted and adjusted to BMI,
71 respectively. Associations include known genes that respond to (*PDGFC*) and regulate
72 response to (*AJUBA*) hypoxia. We identified a complete “chain” linking AvgO₂, *P2RX4*, and
73 butyrylcarnitine (C4), suggesting that increased expression of the purinergic receptor *P2RX4*
74 may improve average oxyhemoglobin saturation and decrease butyrylcarnitine (C4) levels.
75 **Conclusions:** Our results support a mechanistic role for purinergic signaling and hypoxic
76 signaling, among others, in SDB. These findings show differential gene expression by blood cell
77 type in relation to SDB traits and link *P2XR4* expression to influencing AvgO₂ and
78 butyrylcarnitine (C4) levels. Overall, we employed novel methods for integrating multi-omic data
79 to evaluate biological mechanisms underlying multiple SDB traits.

80

81 *Keywords*

82 Average oxyhemoglobin saturation, Apnea hypopnea index, Minimum oxyhemoglobin satura-
83 tion, RNA-seq, Peripheral blood mononuclear cells, Monocyte, T-cell, Instrumental variables,
84 Sleep disordered breathing.

85 **BACKGROUND**

86 Sleep-disordered breathing (SDB) is a common disorder, affecting an estimated 24% of male
87 and 9% of female adults in the U.S. (1). SDB is characterized by episodic periods of breathing
88 cessations and reductions during sleep, often accompanied by oxyhemoglobin desaturation
89 (2,3), and is associated with cardiometabolic, vascular, and cognitive outcomes (4–7). SDB is
90 also strongly associated with inflammation (8,9). While obesity is a strong risk factor for SDB,
91 SDB is also heritable independent of body mass index (BMI) (10,11). The underlying molecular
92 processes by which SDB affects health outcomes are still being studied (12), with interest in un-
93 derstanding SDB-related hypoxia during sleep on cardiometabolic and vascular measures in
94 humans and in animal models (13–15).

95
96 In investigating the molecular changes caused by SDB, previous studies showed changes in
97 distributions and activation of white blood cells (16–18) and inflammatory cytokines (19) in
98 individuals with obstructive sleep apnea (OSA). Other studies reported changes in gene
99 expression in white blood cells following treatment using continuous positive airway pressure
100 (CPAP), or following CPAP withdrawal (20–23), supporting a causal role between SDB-related
101 physiological stressors (such as hypoxia) and immune cell gene expression. Some studies,
102 including those from our group, also reported cross-sectional transcriptomic association with
103 SDB measures from observational studies (23,24). However, these studies focused on a single
104 cell population, and it is unknown whether and how transcriptional effects of SDB differ among
105 circulating leukocyte subpopulations. Likewise, it is yet unknown how SDB-alterations in gene
106 expression translate to metabolic changes. A few previous studies reported associations of
107 blood metabolites with SDB phenotypes, independently of transcriptomics. For example, one
108 study reported change in serum metabolite levels, evaluated on an untargeted platform that
109 surveyed a few hundred metabolites, after multi-level sleep surgery (25). Most other studies
110 considered specific, targeted metabolite changes in sleep disorders (see reviews in (26)).

111
112 Large, untargeted, omics surveys are now becoming available in cohort studies, providing an
113 opportunity to study the association of SDB with well-defined, genetically-regulated molecular
114 measures. We deploy a systems biology approach integrating genomic, transcriptomic, and
115 metabolomic data to identify potential pathways in tissue-specific mechanisms driving SDB-
116 related morbidity.

117
118 Utilizing data from the Multi-Ethnic Study of Atherosclerosis (MESA) and the Hispanic
119 Community Health Study/Study of Latinos (HCHS/SOL), we examined multi-omics data to
120 investigate signaling mechanisms underlying SDB traits. First, we used transcriptomics data
121 measured in peripheral blood mononuclear cells (PBMCs), T-cells and monocytes, assayed by
122 the Trans-Omics for Precision Medicine (TOPMed) program, to perform transcriptome-wide
123 association study of SDB-related phenotypes (measured via overnight polysomnography) in
124 MESA. We compared the results across different peripheral blood cell populations. With these
125 data, we constructed transcript polygenic risk scores (tPRS) predicting transcript expression
126 using genetic data. Next, we built these tPRS in the Women's Health Initiative (WHI) and tested
127 them for association and generalization with their transcripts in whole blood. We calculated the
128 tPRS that generalized in HCHS/SOL. Finally, we applied these tPRS to SDB traits and
129 metabolites in HCHS/SOL to investigate how SDB phenotypes potentially propagate via
130 transcription to metabolic changes in serum, and on the other direction, to assess potential
131 reverse association by which transcript expression causes changes in SDB phenotypes.

132

133 **METHODS**

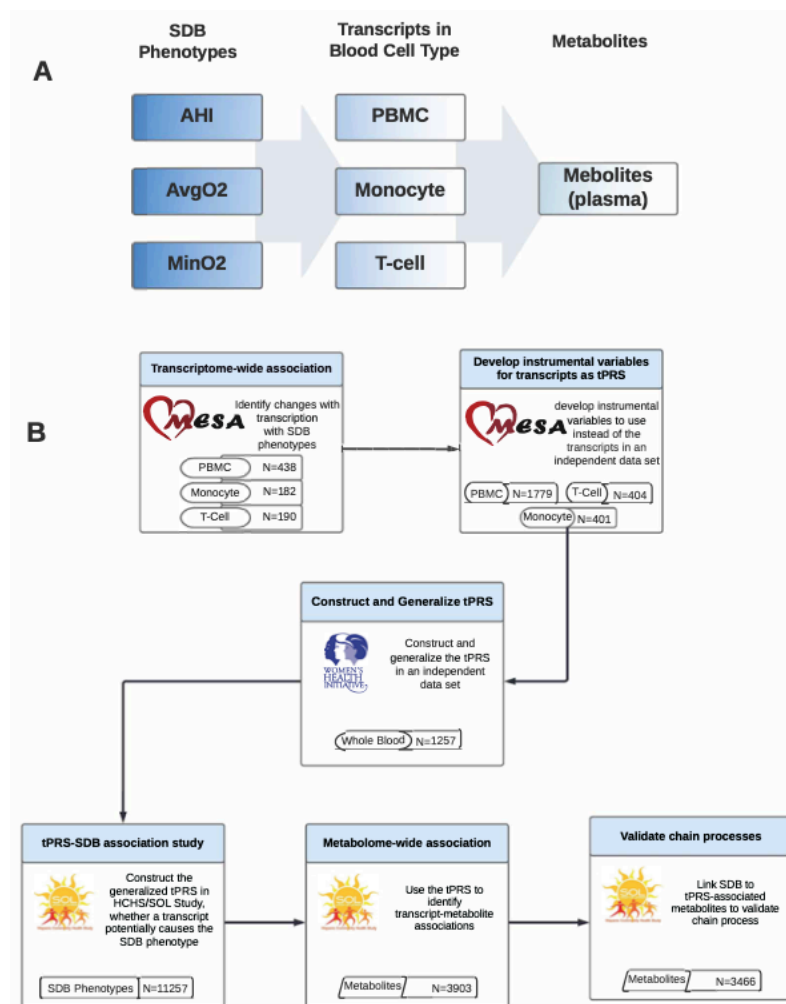
134

135 **Overall study design and purpose**

136 The overall purpose of the study was to investigate the multi-omics signaling mechanisms
137 underlying SDB traits to better understand possible drivers of morbidity in SDB. The study
138 design and purpose of each analysis component is illustrated in **Figure 1**. Briefly, panel A
139 demonstrates the set of associations investigated: SDB phenotypes lead to transcriptional
140 changes which in turn lead to metabolic changes; panel B describes the analysis steps taken to
141 study the potential chain of associations, and the goal of each of these steps. To optimize the
142 available sample size and leverage the fact that transcription is, to some extent, genetically
143 determined, we utilized two separate cohorts to identify the biological components associated
144 sleep exposure to metabolomic changes. **Figure 2** further illustrates potential causal
145 relationships underlying a set of measures, and the assumptions that we used to interrogate
146 some of them. Thus, we first performed transcriptome-wide association studies for SDB
147 phenotypes in MESA. For each transcript associated with a SDB trait (FDR p-value <0.1), we
148 used genetic data to construct a transcript Polygenic Risk Score (tPRS) to serve as a predictor
149 of the transcript. Next, to reduce false positive associations in subsequent analyses, we
150 constructed these tPRS in the WHI and tested their association and generalization with their
151 transcripts in whole blood. We proceeded with tPRS results that generalized (p-value <0.05),
152 and constructed and tested them latter for association with SDB phenotypes in HCHS/SOL. If a
153 tPRS was associated with the SDB phenotype in HCHS/SOL, it was interpreted as evidence of
154 reverse association, i.e., the transcript may contribute to SDB. We then calculated the
155 association of the tPRS with metabolites in HCHS/SOL. Lastly, we used another analytic step to
156 support the existence of an association chain linking a sleep exposure, a transcript, and a
157 metabolite: we required evidence of association between the sleep exposure and the metabolite
158 in HCHS/SOL (i.e. any of the potential pathways in column A of **Figure 2**). If the tPRS was
159 associated with the sleep exposure in HCHS/SOL, it lent support to association chains where
160 the transcript affects both SDB and metabolite levels.

161

Figure 1: Overall study design of the reported analysis.



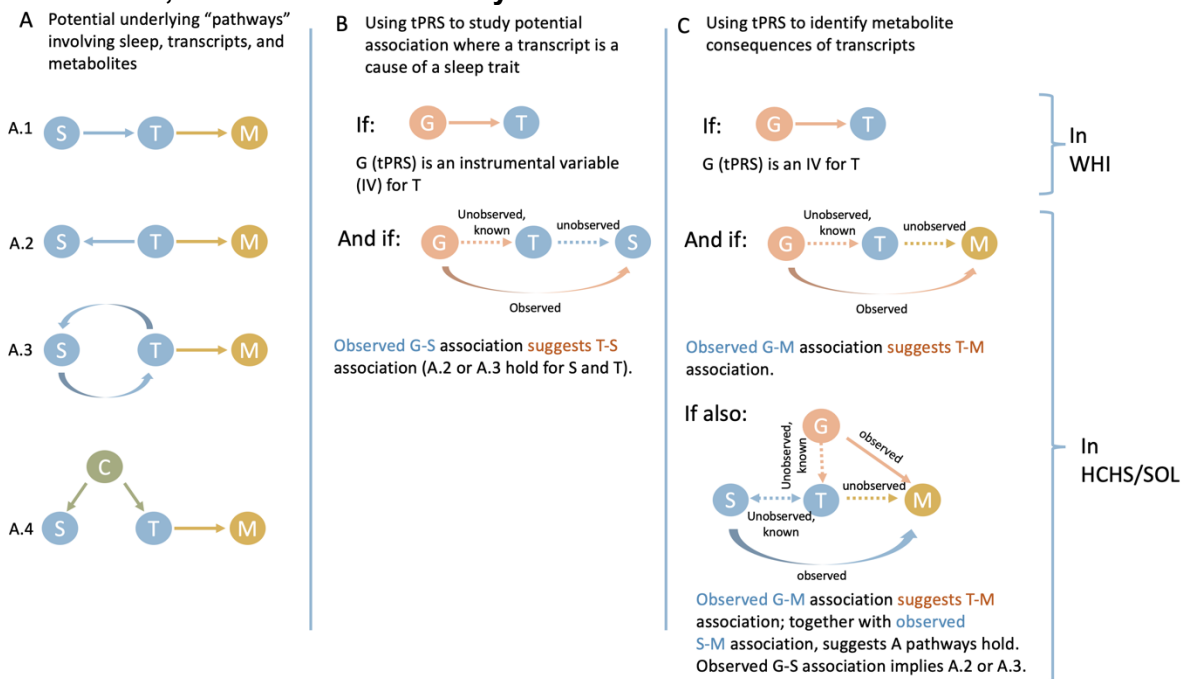
Flow charts illustrating the methodology and purpose of the analysis. The chart portrays the three SDB phenotypes evaluated in the analysis, the three blood tissues with transcript expression measurement, and demonstrates each step of the analysis with associated goals, cohorts, and reasons why step was performed. Panel A: conceptual linking between SDB measures, transcript expression, and metabolites. Panel B: analytic steps supporting the study of the conceptual links. PBMC tPRS analysis sample sizes in MESA correspond to data from two visits (some individuals were used twice, appropriately accounted for by mixed models).

162

163

164

Figure 2: Potential association “chains” of SDB traits, transcript expression, and metabolites, addressed in this study.



The figure illustrates association chains, or pathways, potentially linking an SDB trait, a transcript, and a metabolite. Here we assume that an association between a sleep trait and a transcript was detected in MESA and is assumed “known” for follow-up analysis in WHI in HCHS/SOL. Column A demonstrates potential forms of causal associations between the sleep trait and the transcript, including (A.4) the settings where an association exists due to a common cause, e.g. BMI. Our metabolomics analysis may only detect transcript-metabolite associations, i.e. any sleep-metabolite link is via transcript levels. Column B demonstrate a potential conclusion from an association between a tPRS, validated in WHI and used as an instrumental variable (IV) of a transcript, and a sleep trait: if an association is detected, it provides evidence that changes in transcript levels are upstream (a cause of) changes in sleep trait levels. Column C demonstrates potential conclusions from analyses linking tPRS and a sleep trait to a metabolite. A tPRS is used to link a transcript to a metabolite, and an association, if exists, is likely causal. A sleep-metabolite association should exist if the sleep-transcript and tPRS-transcript associations holds, and therefore observing such an association validates the existence of any of A pathways. Further association between the tPRS and the sleep trait narrows down the potential association chains to A.2 or A.3.

165

166 Participating studies

167 As described in **Figure 1**, our analysis included three studies: MESA, WHI, and HCHS/SOL

168 each contributing to different analytical steps. The three studies are described in the

169 Supplementary Information. In brief, MESA, our primary study used for discovery of SDB-

170 transcript associations, is a longitudinal cohort study (27). The 1st and 5th MESA exams took
171 place between 2000-2002, and 2010-2012, respectively, and whole blood was drawn from
172 participants in both exams. For about 1,000 participants, blood was used later for RNA
173 extraction in at least one of the exams. In addition, a sleep study ancillary to MESA occurred
174 shortly after MESA exam 5 during 2010-2013. Sleep study participants underwent single night
175 in-home polysomnography, as previously described (28). The number of individuals with each
176 type of data and at each time point (exam 1 and exam 5) varies. **Figure S1** in the
177 Supplementary Information visualizes the data flow and overlaps across the various measures
178 used in this study: whole-genome genotyping, RNA-seq, and sleep. All MESA participants
179 provided written informed consent, and the study was approved by the Institutional Review
180 Boards at The Lundquist Institute (formerly Los Angeles BioMedical Research Institute) at
181 Harbor-UCLA Medical Center, University of Washington, Wake Forest School of Medicine,
182 Northwestern University, University of Minnesota, Columbia University, and Johns Hopkins
183 University.

184

185 The WHI was here used to identify tPRS that could be confidently used as IVs for their traits. It
186 is a prospective national health study focused on identifying optimal strategies for preventing
187 chronic diseases that are the major causes of death and disability in postmenopausal women
188 (29). 11,071 WHI participants have whole-genome sequencing data via TOPMed, and 1,274 of
189 these participants have RNA-seq measured in venous blood via TOPMed. All WHI participants
190 provided informed consent and the study was approved by the IRB of the Fred Hutchinson
191 Cancer Research Center.

192

193 The HCHS/SOL was used to establish association chains that include an SDB trait, a transcript,
194 and a metabolite. It is a longitudinal cohort study of U.S. Hispanics/Latinos (30,31). The
195 HCHS/SOL baseline exam occurred on 2008-2011, where 16,415 participants were enrolled.

196 HCHS/SOL individuals who consented further participated in an in-home sleep study, using a
197 validated type 3 home sleep apnea test recording airflow (via nasal pressure), oximetry,
198 position, and snoring. Genetic data were measured and imputed to the TOPMed freeze 5b
199 reference panel, as previously described, for individuals who consented at baseline (32,33).
200 Metabolomic data were also measured for n~4,000 individuals selected at random out of those
201 with genetic data (34). **Figure S2** in the Supplementary Information provides the data flow in
202 HCHS/SOL, focusing on individuals with genetic data and wide consent for genetic data
203 sharing. The HCHS/SOL was approved by the institutional review boards (IRBs) at each field
204 center, where all participants gave written informed consent, and by the Non-Biomedical IRB at
205 the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All
206 IRBs approving the study are: Non-Biomedical IRB at the University of North Carolina at Chapel
207 Hill, Chapel Hill, NC; Einstein IRB at the Albert Einstein College of Medicine of Yeshiva
208 University, Bronx, NY; IRB at Office for the Protection of Research Subjects (OPRS), University
209 of Illinois at Chicago, Chicago, IL; Human Subject Research Office, University of Miami, Miami,
210 FL; Institutional Review Board of San Diego State University, San Diego, CA.

211

212 **RNA sequencing**

213 For both MESA and WHI, RNA-seq was performed via the Trans-Omics in Precision Medicine
214 (TOPMed) program. In MESA, RNA-seq was generated from three blood cell types: peripheral
215 blood mononuclear cells (PBMCs; ~n=1,200 measured in blood from visits 1 and 5), and
216 specific components: T-cells, and monocytes (referred to as T-cell and Mono, for both: n=416),
217 measured in blood from visit 5. Samples were sequenced at the Broad Institute and at the North
218 West Genomics Center (NWGC). Both centers used harmonized protocols. RNA samples
219 quality was assessed using RNA Integrity Number (RIN, Agilent Bioanalyzer) prior to shipment
220 to sequencing centers. QC was re-performed at sequencing centers by RIN analysis at the
221 NWGC and by RNA Quality Score analysis (RQS, Caliper) at the Broad Institute. A minimum of

222 250ng RNA sample was required as input for library construction, performed using the Illumina
223 TruSeq™ Stranded mRNA Sample Preparation Kit. RNA was sequenced as 2x101bp paired-
224 end reads on the Illumina HiSeq 4000 according to the manufacturer's protocols. Target
225 coverage was of ≥40M reads. Comprehensive information about the RNA-seq pipeline used for
226 TOPMed can be found in [https://github.com/broadinstitute/gtex-](https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md)
227 [pipeline/blob/master/TOPMed_RNAseq_pipeline.md](https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md) under MESA RNA-seq pilot commit
228 725a2bc. Here we used gene-level expected counts quantified using RSEM v1.3.0 (35). RNA
229 sequencing for WHI (whole blood) was performed at the Broad Institute using the unified
230 TOPMed protocols. More information about RNA-seq in WHI is provided in the Supplemental
231 Information.

232

233 **Metabolomics data in HCHS/SOL**

234 Metabolomics profiling using fasting blood samples was conducted at Metabolon (Durham, NC)
235 with Discovery HD4 platform in 2017. Serum metabolites were quantified with untargeted, liquid
236 chromatography-mass spectrometry (LC-MS)-based quantification protocol (36,37). The plat-
237 form captured a total of 1,136 metabolites, including 782 known and 354 unknown (unidentified)
238 metabolites. Detailed methodologic information is provided elsewhere (34).

239

240 **Phenotypic measures of sleep-disordered breathing (SDB)**

241 We used three SDB measures, as measured by overnight sleep studies in MESA and
242 HCHS/SOL (methods above): (1) the Apnea-Hypopnea Index (AHI), defined in MESA as the
243 number of apneas (breathing cessation) and hypopneas (at least 30% reduction of breath
244 volume, accompanied by 3% or higher reduction of oxyhemoglobin saturation) per hour of
245 sleep, and in HCHS/SOL, due to differences in the recording montage compared to MESA, as
246 the number of apnea or hypopnea events with 3% desaturation per hour of sleep; (2) minimum

247 oxyhemoglobin saturation during sleep (MinO2), and (3) average oxyhemoglobin saturation
248 during sleep (AvgO2).

249

250 **Testing the association between SDB and blood cell-specific transcriptome-wide gene** 251 **expression**

252 We used the Olivia R package (38) to perform association analyses of gene expression in
253 PBMCs, monocytes, and T-cells with each of the three SDB measures, separately and in a joint
254 analysis in MESA. SDB phenotypes were treated as the exposures. We followed the
255 recommended Olivia pipeline. Briefly, we performed median normalization, and then filtered
256 lowly expressed gene transcripts defined by removing transcripts with proportion of zero higher
257 than 0.5, median value lower than 1, maximum expression range value lower than 5, and
258 maximum expression value lower than 10. Transcript counts were log transformed after counts
259 of zero were replaced with half the minimum of the observed transcript count in the sample. The
260 analyses were adjusted for age, sex, study center, race/ethnic group, and batch variables:
261 plates, shipment batch, and study site. Because BMI is a strong risk factor for SDB and is
262 assumed to be part of the causal chain, we conducted additional analyses adjusting for BMI
263 (BMIadj). We computed empirical p-values to account for the highly skewed distribution of SDB
264 phenotypes, which may lead to false negative associations if ignored. Finally, we accounted for
265 multiple testing by applying False-Discovery Rate (FDR) correction to each of the association
266 analyses using the Benjamini-Hochberg (BH) procedure (39). We carried forward transcript
267 associations with FDR p-value<0.1 for additional analyses and visualized their association with
268 SDB phenotypes via a hierarchically-clustered heatmap.

269

270 **Transcript polygenic risk scores (tPRS) construction and validation**

271 To develop tPRS, we first performed a genome-wide association study (GWAS) for each SDB-
272 associated transcript using the MESA TOPMed WGS dataset; each GWAS adjusted for age

273 (years), sex, study site, self-reported race/ethnic background, and 11 principal components, and
274 analyses were restricted to genetic variants with a minor allele frequency of at least 0.05 (due to
275 low sample size). For each GWAS, we used the fully-adjusted two-stage procedure for rank-
276 normalizing residuals in association analyses (40) to identify genetic variants associated with
277 transcript expression. For PBMCs, we used transcript measures from the two MESA visits with
278 RNA-seq data to increase power. To do this, we removed related individuals, and used a
279 random effect model that accounted for individuals. Summary statistics from the GWAS for each
280 transcript were used to develop PRS weights for the corresponding transcript. Next, we
281 constructed tPRS in MESA. We applied clump and threshold implemented in PRSice2 v2.3.1.e
282 (41) using clumping parameters $R^2=0.1$, distance of 250Kb, and three p-value thresholds (5×10^{-8} , 10^{-7} , 10^{-6}). For each transcript, we constructed the three tPRS in WHI. A tPRS with the
283 smallest p-value in association with the transcript in WHI, and also having p-
284 value $< 0.05/3 = 0.017$, was selected and considered validated. We also computed FDR-adjusted
285 p-values based on all constructed tPRS (3 candidate tPRS per gene across all genes). To test
286 the association of the tPRS with transcript in WHI, we used logistic mixed models, executed with
287 the GENESIS R package (42) version 2.16.1. Each tPRS served as the exposure, and
288 transcripts served as the outcome, here too using the two-stage procedure for rank-
289 normalization (40). Relatedness was modeled via a sparse kinship matrix among TOPMed WHI
290 individuals. We selected transcripts with p-value < 0.017 for follow-up analysis.

291
292
293 We validated that our approach to construct tPRS is robust. We compared a few polygenic
294 prediction models developed using bulk RNAseq in monocytes. First, the prediction model
295 developed using prediXcan based on the MESA dataset (43,44), with weights provided in the
296 predictDB database (<http://www.predictdb.org/>). Second, our approach above using genome-
297 wide SNPs (including trans-eQTLs), and third, a similar clump and threshold approach as above
298 limited to cis-eQTLs defined as SNPs within 1Mbp of the start and end position of the transcript

299 (the definition used by prediXcan). We focused on monocytes for this comparison because
300 prediXcan models were only published based on monocytes.

301

302

303 **Using tPRS to identify reverse association between gene expression and SDB traits**

304 We constructed generalized tPRS in HCHS/SOL. We used HCHS/SOL genotypes imputed to
305 the TOPMed freeze5b reference panel. Prior to tPRS construction, we filtered SNPs with impu-
306 tation quality <0.8, minor allele frequency <5%, missingness rate >0.01. As illustrated in column
307 B of **Figure 2**, We identified potential reverse causation, where gene expression alters SDB, by
308 using the tPRS constructed in HCHS/SOL as instrumental variables (IVs) and testing their asso-
309 ciation with their respective SDB phenotypes in HCHS/SOL. We used logistic mixed models, ex-
310 ecuted with the GENESIS R package (42) version 2.16.1. Each tPRS served as the exposure,
311 and the relevant SDB phenotype served as the outcome. To account for skewness of the SDB
312 phenotypes, we used the two-stage procedure for rank-normalization (40). Relatedness was
313 modeled via a sparse kinship matrix, household sharing, and block unit sharing among
314 HCHS/SOL individuals. Association analyses were adjusted for age, sex, study site, His-
315 panic/Latino background, the first 5 PCs of the genetic data, and log of the sampling weights
316 used to sample HCHS/SOL individuals into the study. Because the tPRS represent a genetic
317 proxy for gene expression, if a tPRS was found to be associated with a SDB phenotype (p-
318 value<0.05), it provided evidence that the transcript contributed to the SDB phenotype, rather
319 than vice versa. However, as illustrated in diagrams A.2 and A.3 in **Figure 2** for sleep-transcript
320 association, bidirectional associations are also plausible.

321

322 **Associations between tPRS and metabolites**

323 Treating tPRS as genetic IVs for gene expression, we estimated associations between tPRS
324 and all identified (named) metabolites with < 25% missing values in HCHS/SOL. We used

325 robust survey models implemented in the R survey package version 4.0 (45), accounting for
326 HCHS/SOL study design (probability sampling and clustering) and providing associations
327 generalizable to the HCHS/SOL target population. For each metabolite, we first imputed
328 observations with missing values of that metabolite with its minimum value observed in the
329 sample, under the assumption that missing values are due to concentrations being below the
330 detection limit, and then rank-normalized it across the sample. We used the same covariates as
331 before: age, sex, study site, Hispanic background, and the first 5 PCs of the genetic data.
332 Furthermore, we adjusted for BMI depending on the original association of the SDB phenotype
333 and the transcript (BMI unadjusted or BMI adjusted). For each transcript, we corrected
334 metabolite associations to account for FDR using the Benjamini-Hochberg (BH) procedure (39).
335 Associations were considered significant if the FDR p-value was <0.05 .

336

337 **Association analyses of SDB traits with selected metabolites to verify a complete associ-** 338 **ation chain**

339 To further validate a complete association “chain” as detailed in **Figure 2**, we performed
340 association analyses between the SDB phenotypes and metabolites identified in the tPRS
341 analysis. Associations between SDB phenotypes and metabolites used a survey sampling
342 approach to account for HCHS/SOL sampling design and obtained estimates generalizable to
343 the HCHS/SOL target population. Thus, we used the survey R package (46) with each individual
344 weighted by their sampling weights, and clustering accounted for when computing robust
345 standard errors. Analyses were adjusted for age, sex, study site, Hispanic/Latino background,
346 and BMI depending on the original detected SDB-transcript association (BMI unadjusted or BMI
347 adjusted). If an SDB phenotype was associated with the metabolite ($p\text{-value}<0.05$), we
348 interpreted this as validation of a SDB association with this metabolite via the transcript-level
349 chain.

350

351 RESULTS

352 Sample characteristics

353 Characteristics of the MESA population that participated in the TOPMed omics study, the sleep
354 study, and the smaller T-cells and monocytes analyses are provided in **Table S1**; characteristics
355 of the HCHS/SOL participants with genetic and metabolite data are provided in **Table S2**.

356 MESA individuals are a multi-ethnic sample, 69 years old on average during MESA exam 5, and
357 52% female. HCHS/SOL individuals are from diverse Hispanic/Latino backgrounds with a mean
358 age of 46 years during the baseline exam, and 59% female. SDB phenotypes were more severe
359 in MESA, with average AHI=18.6, MinO2=83, and AvgO2=94.1, in contrast to HCHS/SOL with
360 average AHI=6.4, MinO2=87.1, and AvgO2=96.4, consistent with the older age of the MESA
361 sample. Characteristics of the WHI participants with RNA-seq data used to validate the tran-
362 script PRS are provided in **Table S3**. WHI individuals are from a multi-race and ethnic sample
363 and are 80 years old on average at the Long-Life Study exam when RNA was extracted, and
364 are all females.

365

366 SDB phenotypes for oxyhemoglobin saturation and AHI are linked to tissue-specific 367 changes in the transcriptome

368 In MESA, we identified 96 and 24 differentially expressed transcripts (**Table S4, S5 and Table**
369 **S6, S7**) with FDR p-value < 0.1 in unadjusted and adjusted BMI analyses, respectively, in the
370 different cell types. **Table 1** reports the top differentially expressed transcripts (FDR p-value
371 <0.05). Three transcripts, *AJUBA* (Ajuba LIM Protein), *ZNF665* (Zinc Finger Protein 665), and
372 *TMC3-AS1* (TMC3 Antisense RNA 1, a long non-coding RNA), are significantly associated with
373 AvgO2 and AHI in both analyses, in the direction of reduced expression with worse SDB
374 measures (higher AHI, lower AvgO2).

375

376

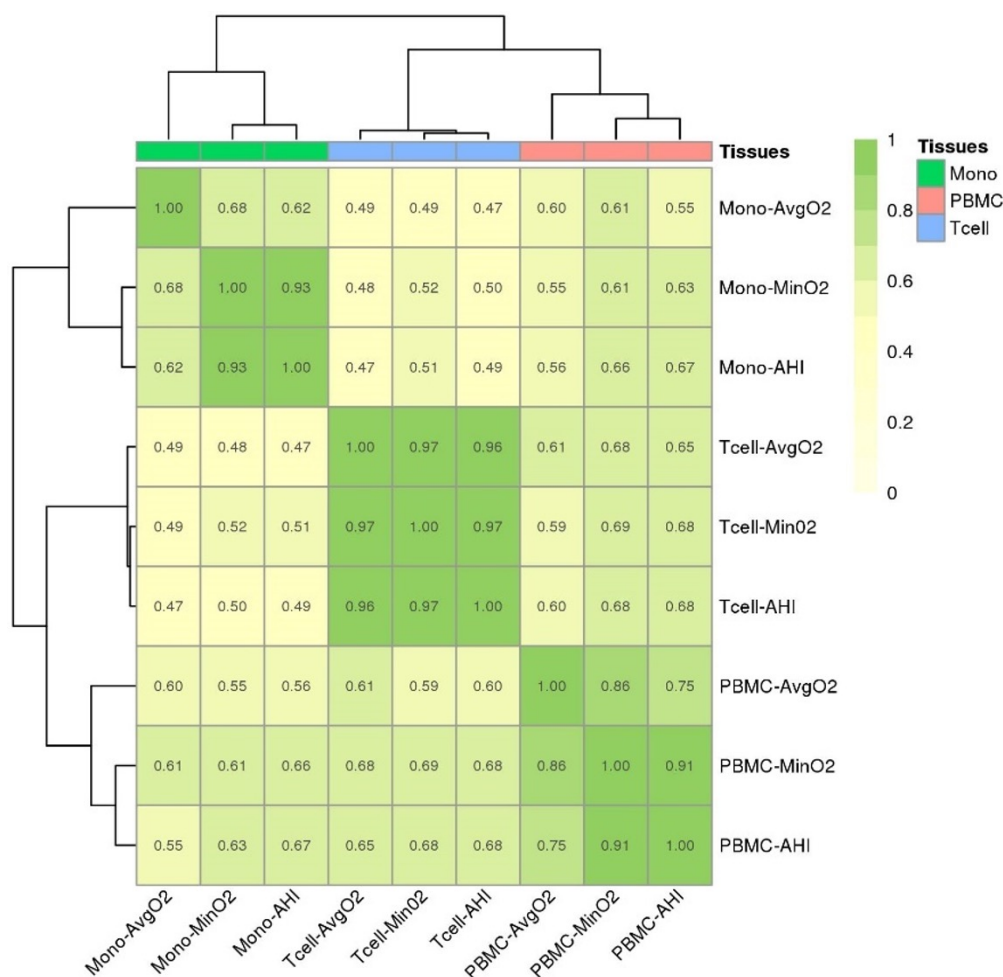
Table 1. Top results from the tissue-specific transcriptome-wide gene expression analysis of SDB phenotypes (FDR p<0.05) in MESA.					
<i>Unadjusted for BMI</i>					
Gene	Adj LFC	P-value	FDR p-value	SDB Trait	Cell Type
<i>AJUBA</i>	0.104	2.59E-06	0.050	AvgO2	PBMCs
<i>PDGFC</i>	-0.016	2.59E-06	0.050	MinO2	PBMCs
<i>SIAE</i>	-0.011	1.26E-06	0.020	MinO2	Monocytes
<i>EMP1</i>	-0.030	6.91E-06	0.050		
<i>LHFPL2</i>	-0.014	9.42E-06	0.050		
<i>ZNF665</i>	-0.011	2.74E-07	0.003		
<i>FAM106A</i>	-0.032	7.69E-06	0.047	AHI	T-cells
<i>TMC3-AS1</i>	-0.022	2.74E-07	0.003		
<i>Adjusted for BMI</i>					
Gene	Adj LFC	P-value	FDR- p-value	SDB Trait	Cell Type
<i>AJUBA</i>	0.116	2.59E-07	0.005	AvgO2	PBMCs
<i>ZNF665</i>	-0.012	5.49E-07	0.010	AHI	T-cells
<i>DUX4L27</i>	-0.029	6.59E-06	0.043		
<i>TMC3-AS1</i>	-0.021	7.14E-06	0.043		
<p>The table provides results from analyses of unadjusted (n=8 transcripts) and adjusted (n=4 transcripts) for BMI. “Adj LFC” is the covariate-adjusted log2-fold change in gene expression per 1 unit increase in SDB exposure. P-value is the raw p-value, and FDR p-value is the p-value following FDR adjustment using the Benjamini-Hochberg procedure. For AvgO2 and MinO2, negative Adj LFC indicates increased expression with worse SDB symptoms. For AHI, positive Adj LFC indicates increased expression with worse SDB symptoms.</p>					

377

378 To visualize gene expression and compare across SDB traits and cell types, Spearman
379 correlation of log-fold change in expression of all SDB-associated transcripts (n=96 transcripts
380 FDR p<0.1) was illustrated with a heatmap for BMI unadjusted (**Figure S3**) and BMI adjusted
381 (**Figure S4**) analyses, clustered using hierarchical clustering based on the correlation between
382 the log-fold estimates. These results illustrate concordant and discordant patterns of differential
383 gene expression by cell type (PBMCs, monocytes, and T-cells) and SDB trait (AvgO2, MinO2,
384 and AHI). There are a few striking differences in gene expression, particularly the increased
385 expression of *FAM106A*, *DNAJA3*, *BCDIN3D-AS1*, *TGFBRAP1*, *BEND5*, *TTC24*, *TMC3-AS1*,
386 *LINC00235*, *SEC14L2*, *ARHGEF9*, *TSHZ1*, and *LA16c-312E8.4* in T-cells compared to
387 monocytes and PBMCs. To further investigate the overall patterns in gene expression in relation
388 to tissue type and SDB traits, a heatmap of the Spearman correlation of the log-fold expression
389 estimates of SDB phenotypes was plotted **Figure 3**. Within cell types, the SDB traits AHI and
390 MinO2 had the highest correlation for gene expression (Spearman R^2 between 0.91 to 0.97),
391 whereas AvgO2 had lower correlations with AHI and MinO2, especially in monocytes. In
392 addition, a heatmap of estimated log-fold gene expression change (FDR p-value < 0.1) with
393 SDB phenotypes across tissues adjusted for BMI is shown in **Figure S5**. The correlations
394 between the SDB effect estimates for gene expression across cell types are different, and
395 generally higher, from the phenotypic correlations between the SDB phenotypes, which are at
396 the range of 0.53 to 0.73 Spearman R^2 (**Figure S6**). When computing correlations over all
397 genes, estimated associations between AvgO2 with gene expression had almost no correlation
398 with the other phenotypes (**Figure S7**). The same patterns (although slightly attenuated) are
399 observed in BMI adjusted analyses (**Figure S8**).

400

Figure 3: Spearman correlations between estimated log-fold changes in gene expression across SDB phenotypes and blood cell types without BMI adjustment in MESA



Heatmap illustrating the Spearman correlations of log-fold change of transcript expression by tissue type (monocytes, T-cells, PBMCs) and SDB phenotype (AvgO2, MinO2, AHI) in MESA. Correlations were computed over genes with FDR $p < 0.1$. Color legend portrays Spearman R^2 (no/weak correlation = light yellow; complete/strong correlation = green). Estimated AHI effect sizes were flipped prior to computation of correlations so that they match the direction of MinO2 and AvgO2.

401

402

403 Construction and validation of transcript PRS

404 We constructed tPRS for gene expression in monocytes using a few methods, focusing on

405 transcripts that were associated with SDB exposures in our analysis. The performance of

406 constructed tPRS was evaluated against whole-blood gene expression levels in n=1,269 WHI
407 participants. **Figure S9** visualizes the results, demonstrating that tPRS constructed using the
408 clump and threshold for genome-wide SNPs, including trans-eQTLs and tPRS focusing on cis-
409 eQTLs have similar results, and the same generalization rate as that of the prediXcan-based
410 tPRS. However, prediXcan tPRS had opposite direction of association with one of the
411 transcripts in WHI, and, both cis-eQTLs based tPRS (prediXcan and clump and threshold) were
412 not available for some transcripts due to lack of transcript-associated SNPs near the coding
413 region. Thus, we moved forward with the genome-wide approach. Of the 96 tPRS (BMI
414 unadjusted analysis) and 24 tPRS (BMI-adjusted analysis) tested, 26 and 9 tPRS were
415 associated ($p < 0.017$) with gene expression (**Tables S8-S9**) in whole blood and considered
416 validated as IVs.

417

418 **Evidence of causal association between transcripts and SDB phenotypes**

419 We tested the association of the validated tPRS, constructed in a cell-specific manner, with SDB
420 phenotypes in HCHS/SOL (**Table S10**). Of the 26 tested in BMI unadjusted analysis, 3 tPRS
421 showed evidence of reverse association with SDB phenotypes ($p\text{-value} < 0.05$), supporting a
422 causal relationship between expression of these transcripts and SDB traits. Among them, the
423 strongest association was of the tPRS for *P2RX4* (Purinergic Receptor P2X 4) in PBMC, in its
424 association with AvgO2, one standard deviation (SD) increase in the PRS was associated with
425 increased 1.9% AvgO2. Additionally, tPRS for *SEC14L2* (SEC14 Like Lipid Binding 2) was
426 negatively associated with AHI in T-cells and tPRS for *TUBB6* (Tubulin Beta 6 Class V) was
427 positively associated with MinO2 in monocytes. After BMI adjustment, only *P2RX4* remained
428 positively associated with AvgO2 in PBMCs ($p\text{-value} < 0.05$), as shown in **Table S11**.

429

430 **Evidence of causal association between transcripts and metabolites**

431 We tested the relation between each validated tPRS and metabolites in HCHS/SOL. The tPRS
432 for P2RX4 and CTD-2366F13.1 (also known as MOCS2-DT, MOCS2 Divergent Transcript)
433 were associated with a total of 6 and 7 metabolites in unadjusted BMI and adjusted BMI
434 analyses (FDR p-value <0.05, Table S12 and Tables S13), respectively; the association
435 “chains” are visualized in **Figure 4**. Of 7 metabolites, 3 of them (butyrylcarnitine, linoleoyl-
436 arachidonoyl-glycerol (18:2/20:4) , and palmitoleoyl-linoleoyl-glycerol (16:1/18:2)) were also
437 associated with AvgO2 (Table S15). However, the AvgO2-metabolite associations did not
438 remain after BMI adjustment, suggesting that BMI, rather than SDB, may be driving these
439 associations (**Table S16**). Of the transcripts, P2RX4 had evidence of a complete chain of
440 association with SDB and metabolites (p-value <0.05) in the BMI unadjusted analysis.
441

442

Figure 4. Identified association chains between AvgO2, transcripts, and metabolites

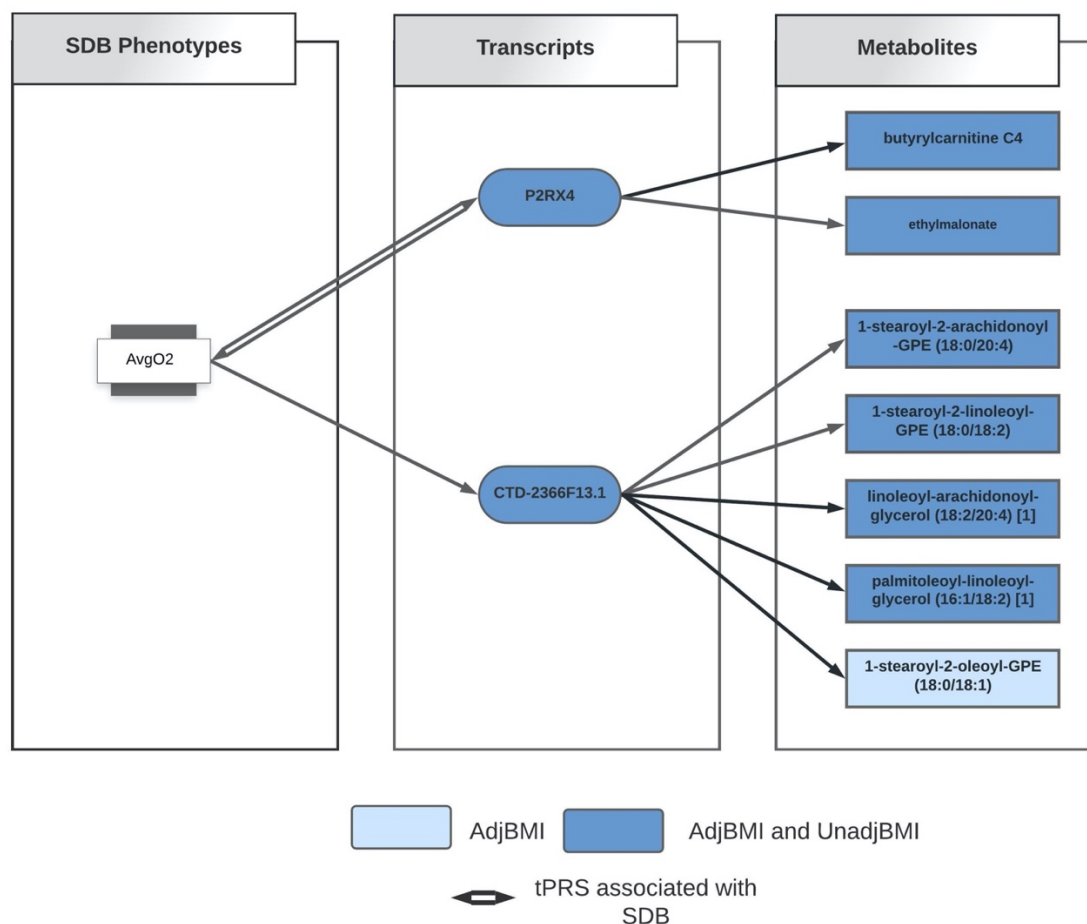


Diagram illustrating the “association chain” relationships between AvgO2, tPRS, and metabolites in BMI unadjusted and BMI adjusted analyses. Dark blue color indicates association between tPRS and metabolites in both BMI unadjusted and adjusted analyses; The light blue color indicates the association between tPRS and metabolite only in BMI adjusted analysis.

443

444 **DISCUSSION**

445 Here, we conducted a robust analysis of SDB phenotypes and their multi-omics correlates. We
446 first identified transcriptome-wide tissue-specific changes in gene expression associated with
447 sleep-related oxyhemoglobin saturation traits and AHI in MESA and then used those transcripts
448 to develop genetic proxies for gene expression (tPRS). Next, we generalized and validated
449 some of the tPRS in WHI. Finally, we utilized the validated tPRS to further study SDB

450 phenotypes and metabolite associations in HCHS/SOL. Our results support SDB-related
451 leukocyte alterations in gene expression and highlight signaling pathways related to
452 inflammation, thrombosis, and neurotransmission.
453
454 SDB traits were associated with differential expression of many transcripts across three blood
455 cell types (**Table S4 and S5**, 96 genes with FDR p-value<0.1). Of the top transcripts (8 genes
456 with FDR p-value <0.05), *AJUBA* expression was positively associated with higher AvgO2 and
457 *PDGFC* expression was negatively associated with higher MinO2 in PBMCs. *AJUBA* is a
458 scaffold protein in the family of LIM domain-containing protein, considered key regulators of the
459 hypoxic response (47). Recent research supports a role for AJUBA in interacting with retinoic
460 acid receptor signaling in an in vitro model (48) and a role for indirectly limiting inflammation by
461 maintaining mitochondrial quality control in a mouse model (49); therefore, greater *AJUBA*
462 expression may be associated with increased AvgO2 through pathways related to inflammation
463 and retinoic acid. *PDGFC* encodes platelet derived growth factor C (50) and is upregulated
464 during hypoxia in tumor cells (51). The association between higher *PDGFC* expression and
465 lower minimum oxygen saturation (MinO2) supports a role for *PDGFC* signaling in SDB-related
466 hypoxia. In monocytes, *EMP1* was linked to MinO2, and prior studies have shown that *EMP1*
467 expression increases during sleep loss and during hypoxia in cancer tissues (52,53).
468 Correlations between leukocyte subsets and trait-specific gene expression (**Figure 3**) supported
469 an overall pattern of similarity between cell populations, but also highlighted some striking
470 differences. For example, the gene *HPCAL4* (Hippocalcin-Like Protein 4) has high expression
471 changes with “worse” SDB phenotypes (higher AHI, lower AvgO2 and MinO2) in monocytes, but
472 weak associations in T-cells and PBMCs. However, *FAM106A* and *SERPINE2* have lower
473 expression with worse SDB phenotypes in T-cells, but weak associations in PBMCs and
474 monocytes. *SERPINE2* (54) encodes glia-derived nexin (GDN, also referred to as protease
475 nexin-1) with mixed evidence as a genetic factor for COPD (55,56). *SERPINE2/GDN* inhibits the

476 hypoxia-triggered serine protease thrombin (57,58), and a mouse model of *SERPINE2*
477 deficiency causes excess thrombin activity and overproduction of cytokines in the lungs (59),
478 suggesting a role for *SERPINE2* in airway inflammation. Since OSA may be associated with a
479 procoagulant state featuring high thrombin levels (60,61), the hypoxemia associated with OSA
480 may lead to increased thrombin levels, affecting *SERPINE2* expression in T-cells.

481
482 Of the top differentially expressed genes in MESA whose tPRS was validated in an independent
483 cohort (WHI), only *P2RX4* was found to have a complete “chain” of association with SDB and
484 metabolites when tested in another independent study, HCHS/SOL. The PRS for *P2RX4* was
485 positively associated with AvgO₂, both with and without adjustment for BMI, and as such it is a
486 candidate contributor to oxyhemoglobin saturation in SDB. *P2RX4* encodes a purinergic
487 receptor for ATP, P₂X₄, which may play a role in the neuroprotective effects of hypoxic
488 preconditioning (62,63). *P2RX4* was further negatively associated with metabolite
489 butyrylcarnitine, an indicator of fatty acid metabolism previously linked to BMI (64). Higher
490 oxygen tension promotes increased ATP production (65), which may in turn promote increased
491 *P2RX4* expression (P₂X₄ as a purinergic ATP receptor) and decreased butyrylcarnitine (66).
492 *P2RX4* can have beneficial or detrimental effects depending on context. A mouse model of
493 genetically increased *P2XR4* expression led to enhanced cardiovascular function (67) and prior
494 research supports a role for *P2XR4* in heart contractility (68), suggesting that *P2XR4* may
495 positively drive AvgO₂ via bolstering cardiac force. Likewise, butyrylcarnitine was found to be
496 associated with time of cardiac isovolumetric relaxation and may be a marker of heart failure
497 (69), further linking *P2XR4* to cardiac function. While *CTD-2366F13.1* (*MOCS2-DT*) was not
498 associated with SDB traits, two of the four metabolites negatively associated with *CTD-*
499 *2366F13.1* were also associated with AvgO₂: linoleoyl-arachidonoyl- glycerol (18:2/20:4) and
500 palmitoleoyl-linoleoyl- glycerol (16:1/18:2). Levels of 2-Arachidonoylglycerol, an agonist of the
501 CB1 and CB2 cannabinoid receptors, are increased in the brain during ischemia (70) and in

502 macrophages in response to oxidative stress (71). Therefore, greater AvgO2 levels may result
503 in decreased linoleoyl-arachidonoyl- glycerol (18:2/20:4) levels.

504
505 There are several strengths and some limitations of our analysis. Our unique study design
506 exploited a stepwise discovery/validation approach across multiple studies and optimized the
507 availability of SDB-related datasets to study omics markers and SDB. First, we identified SDB-
508 related transcripts. Next, we utilized genetic associations with gene expression to construct
509 tPRS, serving as “genetic IVs”: exposure variables that are likely associated with the gene
510 transcripts and are specific to them, thus allowing for downstream association analysis and
511 causal inference using these IVs instead of the transcript themselves (72). The idea of using
512 genetic variants as IVs is often used in Mendelian Randomization (MR) analysis. Our analysis is
513 different than standard one-sample MR in that we did not estimate the effect of the transcript on
514 the outcome, because we did not have access to RNA-seq in HCHS/SOL. However, for
515 causality inference, it is sufficient to test the IV association with the outcome of interest (73). We
516 then studied the evidence for the effect of gene expression on SDB using the tPRS. Still, the
517 exact form of association between the gene expression and SDB traits cannot be determined
518 (**Figure 2**). For example, if no tPRS-SDB association was detected, it is possible that this was
519 due to lack of power. Even in the absence of tPRS-SDB association, the association between
520 the SDB and tPRS can be due to either causal effect of SDB on tPRS, or confounding by a
521 common cause of both. Notably, during the WHI validation step, many transcripts did not have
522 significant tPRS associations and therefore were not carried forward for the genetic association
523 analysis in HCHS/SOL; lack of validation may be due to cell type differences, as we validated
524 tPRS in WHI, where gene expression was measured in whole blood, unlike measurement in
525 specific cell types in MESA. Finally, we leveraged the validated tPRS to test for associations of
526 gene transcript expression with metabolites and connect possible “chains” of associations. All
527 included cohorts are large and represent diverse populations in the U.S. Our sleep cohorts,

528 HCHS/SOL and MESA, have objective sleep phenotype measurement without prior selection of
529 participants based on specific phenotypes. Other limitations of our study include high multiple
530 testing burden, performing procedures with multiple steps, utilizing multiple data in constructing
531 tPRS, and differences in sample timing between blood sample collection and overnight PSG in
532 MESA. However, genetic data should not be affected by differences in timing, and chronic
533 conditions like SDB may be stable over time, making this limitation less of a concern. It is
534 notable that the three blood cell types used in MESA are not distinct: PBMCs include monocytes
535 and T-cells. Further, both monocytes and T-cells are also composed of more granular cell types.
536 Statistical analyses within one cell type are generally powered to detect associations that hold
537 across the component, more granular, cell types, and some cell type-specific associations may
538 be masked. Overall, we utilize robust statistical methods and objective measures, integrating
539 across multiple layers of biological measures, to interrogate the mechanisms driving SDB-
540 related morbidity.

541

542 **CONCLUSION**

543 In summary, we examined multiple levels of biological information to investigate signaling
544 mechanisms underlying SDB traits to better understand drivers of morbidity in SDB. Our results
545 highlight differential gene expression by circulating leukocyte populations in relation to multiple
546 SDB traits related to hypoxia, neurotransmission, and thrombolytic activity. Analyses with
547 validated tPRS in independent cohorts support a mechanistic role for *P2XR4* purinergic
548 signaling in SDB, a gene known to influence cardiac function, which is relevant to SDB as both
549 a risk factor and outcome. Overall, we applied novel, robust methods to integrate multi-omic
550 data and SDB data to discover mechanisms underlying multiple SDB traits. Our multi-
551 dimensional approach using large population cohorts is a promising approach to unravel
552 biological underpinnings of complex human disorders.

553

554 **ACKNOWLEDGEMENTS**

555 The authors thank the staff and participants of HCHS/SOL, MESA, and WHI, for their important
556 contributions. We gratefully acknowledge the investigators and participants who provided
557 biological samples and data for TOPMed.

558

559 **FUNDING**

560 This work was supported by the National Heart Lung and Blood Institute grants R35HL135818
561 to S.R., T32-HL007901 to D.C-T., and R21HL145425 to T.S. Molecular data for the Trans-
562 Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung
563 and Blood Institute (NHLBI). Genome Sequencing for “NHLBI TOPMed: Multi-Ethnic Study of
564 Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Broad Institute Genomics
565 Platform (HHSN268201500014C). RNA-Seq for “NHLBI TOPMed: Multi-Ethnic Study of
566 Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Northwest Genomics Center
567 (HHSN268201600032I). Genome Sequencing for “NHLBI TOPMed: Women’s Health Initiative
568 (WHI)” (phs001237.v3.p1) was performed at the Broad Institute Genomics Platform
569 (HHSN268201500014C). RNA-Seq for “NHLBI TOPMed: Women’s Health Initiative (WHI)”
570 (phs001237.v3.p1) was performed at the Broad Institute Genomics Platform
571 (HHSN268201600034I). Core support including centralized genomic read mapping and
572 genotype calling, along with variant quality metrics and filtering were provided by the TOPMed
573 Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core
574 support including phenotype harmonization, data management, sample-identity QC, and
575 general program coordination were provided by the TOPMed Data Coordinating Center
576 (R01HL-120393; U01HL-120393; contract HHSN268201800001I). The MESA projects are
577 conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in
578 collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis
579 (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute

580 (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts
581 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160,
582 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006,
583 N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-
584 HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079,
585 UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other
586 investigators, the staff, and the participants of the MESA study for their valuable
587 contributions. A full list of participating MESA investigators and institutes can be found at
588 <http://www.mesa-nhlbi.org>. The WHI program is funded by the National Heart, Lung, and Blood
589 Institute, National Institutes of Health, U.S. Department of Health and Human Services through
590 contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004,
591 75N92021D00005. The Hispanic Community Health Study/Study of Latinos is a collaborative
592 study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the
593 University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami
594 (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine
595 (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago
596 (HHSN268201300003I / N01- HC-65236 Northwestern Univ), and San Diego State University
597 (HHSN268201300005I / N01-HC-65237). The following Institutes/Centers/Offices have
598 contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on
599 Minority Health and Health Disparities, National Institute on Deafness and Other
600 Communication Disorders, National Institute of Dental and Craniofacial Research, National
601 Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological
602 Disorders and Stroke, NIH Institution-Office of Dietary Supplements. The Genetic Analysis
603 Center at the University of Washington was supported by NHLBI and NIDCR contracts
604 (HHSN268201300005C AM03 and MOD03). Support for metabolomics data was graciously
605 provided by the JLH Foundation (Houston, Texas).

606

607

608 **AUTHOR CONTRIBUTIONS**

609 T.S. supervised the research. N.K. and T.S. conceptualized and conducted the biostatistical and
610 bioinformatics analyses. N.K., T.S., and D.A.W. interpreted the results and wrote and edited the
611 manuscript. Y.Z., B.Y, B.C., H.W., H.M.O-B, A.P.R., A.R.R., J.D.S., J.C., M.D., P.C.Z., R.K.,
612 C.K., S.S.R., J.I.R., S.A.G, and S.R. read and approved the final manuscript.

613

614 **COI**

615 The authors declare that they have no competing interests.

616

617 **DATA AVAILABILITY STATEMENT**

618 MESA, HCHS/SOL and WHI data are available through application to dbGaP according to the
619 study specific accessions. MESA phenotypes are available in: phs000209; WHI phenotypes:
620 phs000200; and HCHS/SOL phenotypes: phs000810. HCHS/SOL genotyping data: phs000880.
621 MESA and WHI RNA-seq data has been deposited and will become available through the TOP-
622 Med according to the study specific accessions; MESA: phs001416 ; WHI: phs000972.
623 HCHS/SOL metabolomics data are available via data use agreement with the HCHS/SOL Data
624 Coordinating Center at the University of North Carolina at Chapel Hill, see collaborators web-
625 site: <https://sites.csc.unc.edu/hchs/>. Data needed to construct the tPRS are publicly available
626 on the repository https://github.com/nkurniansyah/SDB_Multi_Omics.

627

628 **Bibliography**

629

- 630 1. Rundo JV. Obstructive sleep apnea basics. *Cleve Clin J Med*. 2019 Sep;86(9 Suppl 1):2–9.
- 631 2. Mehra R, Stone KL, Blackwell T, Ancoli Israel S, Dam T-TL, Stefanick ML, et al.
632 Prevalence and correlates of sleep-disordered breathing in older men: osteoporotic
633 fractures in men sleep study. *J Am Geriatr Soc*. 2007 Sep;55(9):1356–64.
- 634 3. Redline S, Kump K, Tishler PV, Browner I, Ferrette V. Gender differences in sleep
635 disordered breathing in a community-based sample. *Am J Respir Crit Care Med*. 1994
636 Mar;149(3 Pt 1):722–6.
- 637 4. Osorio RS, Gumb T, Pirraglia E, Varga AW, Lu S-E, Lim J, et al. Sleep-disordered
638 breathing advances cognitive decline in the elderly. *Neurology*. 2015 May 12;84(19):1964–
639 71.
- 640 5. Botros N, Concato J, Mohsenin V, Selim B, Doctor K, Yaggi HK. Obstructive sleep apnea
641 as a risk factor for type 2 diabetes. *Am J Med*. 2009 Dec;122(12):1122–7.
- 642 6. Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between
643 sleep-disordered breathing and hypertension. *N Engl J Med*. 2000 May 11;342(19):1378–
644 84.
- 645 7. Li X, Sotres-Alvarez D, Gallo LC, Ramos AR, Aviles-Santa L, Perreira KM, et al.
646 Associations of Sleep-disordered Breathing and Insomnia with Incident Hypertension and
647 Diabetes. The Hispanic Community Health Study/Study of Latinos. *Am J Respir Crit Care Med*.
648 2021 Feb 1;203(3):356–65.
- 649 8. Chami HA, Fontes JD, Vasan RS, Keaney JF, O’Connor GT, Larson MG, et al. Vascular
650 inflammation and sleep disordered breathing in a community-based cohort. *Sleep*. 2013
651 May 1;36(5):763–768C.
- 652 9. Unnikrishnan D, Jun J, Polotsky V. Inflammation in sleep apnea: an update. *Rev Endocr
653 Metab Disord*. 2015 Mar;16(1):25–34.
- 654 10. de Paula LKG, Alvim RO, Pedrosa RP, Horimoto ARVR, Krieger JE, Oliveira CM, et al.
655 Heritability of OSA in a rural population. *Chest*. 2016 Jan 6;149(1):92–7.
- 656 11. Carmelli D, Colrain IM, Swan GE, Bliwise DL. Genetic and environmental influences in
657 sleep-disordered breathing in older male twins. *Sleep*. 2004 Aug 1;27(5):917–22.
- 658 12. Ryan S, Cummins EP, Farre R, Gileles-Hillel A, Jun JC, Oster H, et al. Understanding the
659 pathophysiological mechanisms of cardiometabolic complications in obstructive sleep
660 apnoea: towards personalised treatment approaches. *Eur Respir J*. 2020 Aug 6;56(2).
- 661 13. Baguet J-P, Hammer L, Lévy P, Pierre H, Launois S, Mallion J-M, et al. The severity of
662 oxygen desaturation is predictive of carotid wall thickening and plaque occurrence. *Chest*.
663 2005 Nov;128(5):3407–12.
- 664 14. Náchér M, Serrano-Mollar A, Farré R, Panés J, Seguí J, Montserrat JM. Recurrent
665 obstructive apneas trigger early systemic inflammation in a rat model of sleep apnea.
666 *Respir Physiol Neurobiol*. 2007 Jan 15;155(1):93–6.
- 667 15. Dematteis M, Godin-Ribuot D, Arnaud C, Ribouot C, Stanke-Labesque F, Pépin J-L, et al.
668 Cardiovascular consequences of sleep-disordered breathing: contribution of animal models
669 to understanding the human disease. *ILAR J*. 2009;50(3):262–81.
- 670 16. Dyugovskaya L, Lavie P, Lavie L. Phenotypic and functional characterization of blood
671 gammadelta T cells in sleep apnea. *Am J Respir Crit Care Med*. 2003 Jul 15;168(2):242–9.
- 672 17. Dyugovskaya L, Lavie P, Lavie L. Increased adhesion molecules expression and
673 production of reactive oxygen species in leukocytes of sleep apnea patients. *Am J Respir*

- 674 Crit Care Med. 2002 Apr 1;165(7):934–9.
- 675 18. Minoguchi K, Tazaki T, Yokoe T, Minoguchi H, Watanabe Y, Yamamoto M, et al.
676 Elevated production of tumor necrosis factor-alpha by monocytes in patients with
677 obstructive sleep apnea syndrome. *Chest*. 2004 Nov;126(5):1473–9.
- 678 19. Bergeron C, Kimoff J, Hamid Q. Obstructive sleep apnea syndrome and inflammation. *J*
679 *Allergy Clin Immunol*. 2005 Dec;116(6):1393–6.
- 680 20. Gharib SA, Seiger AN, Hayes AL, Mehra R, Patel SR. Treatment of obstructive sleep
681 apnea alters cancer-associated transcriptional signatures in circulating leukocytes. *Sleep*.
682 2014 Apr 1;37(4):709–14, 714A.
- 683 21. Perry JC, Guindalini C, Bittencourt L, Garbuio S, Mazzotti DR, Tufik S. Whole blood
684 hypoxia-related gene expression reveals novel pathways to obstructive sleep apnea in
685 humans. *Respir Physiol Neurobiol*. 2013 Dec 1;189(3):649–54.
- 686 22. Turnbull CD, Lee LYW, Starkey T, Sen D, Stradling J, Petousi N. Transcriptomics
687 Identify a Unique Intermittent Hypoxia-mediated Profile in Obstructive Sleep Apnea. *Am J*
688 *Respir Crit Care Med*. 2020 Jan 15;201(2):247–50.
- 689 23. Sofer T, Li R, Joehanes R, Lin H, Gower AC, Wang H, et al. Transcriptional survey of
690 peripheral blood links lower oxygen saturation during sleep with reduced expressions of
691 CD1D and RAB20 that is reversed by CPAP therapy. *medRxiv*. 2019 Aug 12;
- 692 24. Polotsky VY, Bevans-Fonti S, Grigoryev DN, Punjabi NM. Intermittent hypoxia alters
693 gene expression in peripheral blood mononuclear cells of healthy volunteers. *PLoS One*.
694 2015 Dec 14;10(12):e0144725.
- 695 25. Alterki A, Joseph S, Thanaraj TA, Al-Khairi I, Cherian P, Channanath A, et al. Targeted
696 Metabolomics Analysis on Obstructive Sleep Apnea Patients after Multilevel Sleep
697 Surgery. *Metabolites*. 2020 Sep 1;10(9).
- 698 26. Xu H, Zheng X, Jia W, Yin S. Chromatography/Mass Spectrometry-Based Biomarkers in
699 the Field of Obstructive Sleep Apnea. *Medicine*. 2015 Oct;94(40):e1541.
- 700 27. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-
701 Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*. 2002 Nov
702 1;156(9):871–81.
- 703 28. Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/Ethnic
704 Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA).
705 *Sleep*. 2015 Jun 1;38(6):877–88.
- 706 29. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The Women’s
707 Health Initiative recruitment methods and results. *Ann Epidemiol*. 2003 Oct;13(9
708 Suppl):S18-77.
- 709 30. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglius ML, Giachello
710 AL, et al. Design and implementation of the Hispanic Community Health Study/Study of
711 Latinos. *Ann Epidemiol*. 2010 Aug;20(8):629–41.
- 712 31. Lavange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, et al.
713 Sample design and cohort selection in the Hispanic Community Health Study/Study of
714 Latinos. *Ann Epidemiol*. 2010 Aug;20(8):642–9.
- 715 32. Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al. Genetic
716 diversity and association studies in US hispanic/latino populations: applications in the
717 hispanic community health study/study of latinos. *Am J Hum Genet*. 2016 Jan
718 7;98(1):165–84.
- 719 33. Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. Use of >100,000

- 720 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome
721 sequences improves imputation quality and detection of rare variant associations in
722 admixed African and Hispanic/Latino populations. *PLoS Genet.* 2019 Dec
723 23;15(12):e1008500.
- 724 34. Feofanova EV, Chen H, Dai Y, Jia P, Grove ML, Morrison AC, et al. A Genome-wide
725 Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic
726 Community Health Study/Study of Latinos. *Am J Hum Genet.* 2020 Nov 5;107(5):849–63.
- 727 35. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or
728 without a reference genome. *BMC Bioinformatics.* 2011 Aug 4;12:323.
- 729 36. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted
730 ultrahigh performance liquid chromatography/electrospray ionization tandem mass
731 spectrometry platform for the identification and relative quantification of the small-
732 molecule complement of biological systems. *Anal Chem.* 2009 Aug 15;81(16):6656–67.
- 733 37. Ohta T, Masutomi N, Tsutsui N, Sakairi T, Mitchell M, Milburn MV, et al. Untargeted
734 metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer
735 344 male rats. *Toxicol Pathol.* 2009 Jun;37(4):521–35.
- 736 38. Sofer T, Kurniansyah N, Aguet F, Ardlie K, Durda P, Nickerson DA, et al. Benchmarking
737 association analyses of continuous exposures with RNA-seq in observational studies. *Brief*
738 *Bioinformatics.* 2021 May 20;
- 739 39. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful
740 approach to multiple testing. *Journal of the Royal Statistical Society: Series B*
741 *(Methodological).* 1995 Jan;57(1):289–300.
- 742 40. Sofer T, Zheng X, Gogarten SM, Laurie CA, Grinde K, Shaffer JR, et al. A fully adjusted
743 two-stage procedure for rank-normalization in genetic association studies. *Genet*
744 *Epidemiol.* 2019 Jan 17;43(3):263–75.
- 745 41. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data.
746 *Gigascience.* 2019 Jul 1;8(7).
- 747 42. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association
748 testing using the GENESIS R/Bioconductor package. *Bioinformatics.* 2019 Dec
749 15;35(24):5346–8.
- 750 43. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al.
751 A gene-based association method for mapping traits using reference transcriptome data.
752 *Nat Genet.* 2015 Sep;47(9):1091–8.
- 753 44. Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, et al. Genetic
754 architecture of gene expression traits across diverse populations. *PLoS Genet.* 2018 Aug
755 10;14(8):e1007586.
- 756 45. Lumley T. Package “survey.” R package version. 2015;
- 757 46. Lumley T. *Complex surveys: a guide to analysis using R.* 2011;
- 758 47. Bridge KS, Sharp TV. Regulators of the hypoxic response: a growing family. *Future*
759 *Oncol.* 2012 May;8(5):491–3.
- 760 48. Hou Z, Peng H, White DE, Negorev DG, Maul GG, Feng Y, et al. LIM protein Ajuba
761 functions as a nuclear receptor corepressor and negatively regulates retinoic acid signaling.
762 *Proc Natl Acad Sci USA.* 2010 Feb 16;107(7):2938–43.
- 763 49. Ponia SS, Robertson SJ, McNally KL, Subramanian G, Sturdevant GL, Lewis M, et al.
764 Mitophagy antagonism by ZIKV reveals Ajuba as a regulator of PINK1 signaling, PKR-
765 dependent inflammation, and viral invasion of tissues. *Cell Rep.* 2021 Oct

- 766 26;37(4):109888.
- 767 50. Reigstad LJ, Varhaug JE, Lillehaug JR. Structural and functional specificities of PDGF-C
768 and PDGF-D, the novel members of the platelet-derived growth factors family. *FEBS J.*
769 2005 Nov;272(22):5723–41.
- 770 51. Clara CA, Marie SKN, de Almeida JRW, Wakamatsu A, Oba-Shinjo SM, Uno M, et al.
771 Angiogenesis and expression of PDGF-C, VEGF, CD105 and HIF-1 α in human
772 glioblastoma. *Neuropathology.* 2014 Aug;34(4):343–52.
- 773 52. Arnardottir ES, Nikonova EV, Shockley KR, Podtelezchnikov AA, Anafi RC, Tanis KQ, et
774 al. Blood-gene expression reveals reduced circadian rhythmicity in individuals resistant to
775 sleep deprivation. *Sleep.* 2014 Oct 1;37(10):1589–600.
- 776 53. Chen A, Sceneay J, Gødde N, Kinwel T, Ham S, Thompson EW, et al. Intermittent
777 hypoxia induces a metastatic phenotype in breast cancer. *Oncogene.* 2018
778 Aug;37(31):4214–25.
- 779 54. Monard D. SERPINE2/Protease Nexin-1 in vivo multiple functions: Does the puzzle make
780 sense? *Semin Cell Dev Biol.* 2017 Feb;62:160–9.
- 781 55. Demeo DL, Mariani TJ, Lange C, Srisuma S, Litonjua AA, Celedon JC, et al. The
782 SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum*
783 *Genet.* 2006 Feb;78(2):253–64.
- 784 56. Chappell S, Daly L, Morgan K, Baranes TG. The SERPINE2 gene and chronic obstructive
785 pulmonary disease. *American journal of.* 2006;
- 786 57. Baker JB, Low DA, Simmer RL, Cunningham DD. Protease-nexin: a cellular component
787 that links thrombin and plasminogen activator and mediates their binding to cells. *Cell.*
788 1980 Aug;21(1):37–45.
- 789 58. Görlach A, Diebold I, Schini-Kerth VB, Berchner-Pfannschmidt U, Roth U, Brandes RP, et
790 al. Thrombin Activates the Hypoxia-Inducible Factor-1 Signaling Pathway in Vascular
791 Smooth Muscle Cells. *Circ Res.* 2001 Jul 6;89(1):47–54.
- 792 59. Solleti SK, Srisuma S, Bhattacharya S, Rangel-Moreno J, Bijli KM, Randall TD, et al.
793 Serpine2 deficiency results in lung lymphocyte accumulation and bronchus-associated
794 lymphoid tissue formation. *FASEB J.* 2016 Jul;30(7):2615–26.
- 795 60. Zolotoff C, Bertoletti L, Gozal D, Mismetti V, Flandrin P, Roche F, et al. Obstructive
796 Sleep Apnea, Hypercoagulability, and the Blood-Brain Barrier. *J Clin Med.* 2021 Jul
797 14;10(14).
- 798 61. Liak C, Fitzpatrick M. Coagulability in obstructive sleep apnea. *Can Respir J.* 2011
799 Dec;18(6):338–48.
- 800 62. Garcia-Guzman M, Soto F, Gomez-Hernandez JM, Lund PE, Stühmer W. Characterization
801 of recombinant human P2X4 receptor reveals pharmacological differences to the rat
802 homologue. *Mol Pharmacol.* 1997 Jan;51(1):109–18.
- 803 63. Ozaki T, Muramatsu R, Sasai M, Yamamoto M, Kubota Y, Fujinaka T, et al. The P2X4
804 receptor is required for neuroprotection via ischemic preconditioning. *Sci Rep.* 2016 May
805 13;6:25893.
- 806 64. Moore SC, Matthews CE, Sampson JN, Stolzenberg-Solomon RZ, Zheng W, Cai Q, et al.
807 Human metabolic correlates of body mass index. *Metabolomics.* 2014 Apr 1;10(2):259–69.
- 808 65. Bardsley EN, Pen DK, McBryde FD, Ford AP, Paton JFR. The inevitability of ATP as a
809 transmitter in the carotid body. *Auton Neurosci.* 2021 Sep;234:102815.
- 810 66. Zolkipli-Cunningham Z, Naviaux JC, Nakayama T, Hirsch CM, Monk JM, Li K, et al.
811 Metabolic and behavioral features of acute hyperpurinergia and the maternal immune

- 812 activation mouse model of autism spectrum disorder. *PLoS One*. 2021 Mar
813 18;16(3):e0248771.
- 814 67. Yang A, Sonin D, Jones L, Barry WH, Liang BT. A beneficial role of cardiac P2X4
815 receptors in heart failure: rescue of the calsthestrin overexpression model of
816 cardiomyopathy. *Am J Physiol Heart Circ Physiol*. 2004 Sep;287(3):H1096-103.
- 817 68. Hu B, Mei QB, Yao XJ, Smith E, Barry WH, Liang BT. A novel contractile phenotype
818 with cardiac transgenic expression of the human P2X4 receptor. *FASEB J*. 2001
819 Dec;15(14):2739–41.
- 820 69. Razavi AC, Bazzano LA, He J, Fernandez C, Whelton SP, Krousel-Wood M, et al. Novel
821 findings from a metabolomics study of left ventricular diastolic function: the bogalusa
822 heart study. *J Am Heart Assoc*. 2020 Feb 4;9(3):e015118.
- 823 70. Melis M, Pillolla G, Bisogno T, Minassi A, Petrosino S, Perra S, et al. Protective activation
824 of the endocannabinoid system during ischemia in dopamine neurons. *Neurobiol Dis*. 2006
825 Oct;24(1):15–27.
- 826 71. Matthews AT, Lee JH, Borazjani A, Mangum LC, Hou X, Ross MK. Oxyradical stress
827 increases the biosynthesis of 2-arachidonoylglycerol: involvement of NADPH oxidase. *Am*
828 *J Physiol Cell Physiol*. 2016 Dec 1;311(6):C960–74.
- 829 72. von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as
830 instrumental variables. *J Health Econ*. 2016 Jan;45:131–48.
- 831 73. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges
832 in mendelian randomization. *Epidemiology*. 2014 May;25(3):427–35.

833
834

835

836

837

838

839

840

841

842

843

844

845

846

847

848 **An integrated multi-omics analysis of sleep-disordered breathing traits across mul-**
849 **ti-ple blood cell types**

850
851 Kurniansyah et al.

852		
853	Supplemental Methods	35
854	The Multi-Ethnic Study of Atherosclerosis (MESA)	35
855	The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)	36
856	Genotyping and imputation in HCHS/SOL	37
857	The Women’s Health Initiative (WHI).....	37
858	RNA sequencing in WHI	38
859	Supplemental figures.....	39
860	Figure S1: MESA data flow across various measures in the 1 st and 5 th exam, and the sleep	
861	ancillary study.....	39
862	Figure S2: HCHS/SOL data flow across genotyping, metabolomics, and sleep data.	40
863	Figure S3: Heatmap of estimated log-fold gene expression change with SDB phenotypes	
864	across tissues without BMI adjustment in MESA	41
865	Figure S4: Heatmap of estimated log-fold gene expression change with SDB phenotypes	
866	across tissues from BMI-adjusted analysis in MESA.....	42
867	Figure S5: Spearman correlations between estimated log-fold changes in gene expression	
868	across SDB phenotypes and tissues in BMI adjusted analysis in MESA (top genes).....	43
869	Table S6. Correlation between the SDB phenotypes.....	44
870	Figure S7: Spearman correlations between estimated log-fold changes in gene expression	
871	across SDB phenotypes and tissues in analysis without BMI adjustment (all genes).....	45
872	Figure S8: Spearman correlations between estimated log-fold changes in gene expression	
873	across SDB phenotypes and tissues in analysis BMI adjustment (all genes).	46
874	Figure S9. Comparison of the associations between monocyte-based tPRSs and whole-blood	
875	gene expression in WHI	47
876	References.....	48

877
878

879 **Supplemental Methods**

880
881 **The Multi-Ethnic Study of Atherosclerosis (MESA)**
882

883 MESA is a longitudinal cohort study (27), established in 2000, that prospectively collected risk
884 factors for development of subclinical and clinical cardiovascular disease among participants in
885 six field centers across the United States (Baltimore City and Baltimore County, MD; Chicago,
886 IL; Forsyth County, NC; Los Angeles County, CA; Northern Manhattan and the Bronx, NY; and
887 St. Paul, MN). The 1st and 5th MESA exams took place between 2000-2002, and 2010-2012,
888 respectively, and whole blood was drawn from participants in both exams. For about 1,400
889 participants, blood was used later for RNA extraction and/or proteomics in at least one of the
890 exams. In addition, a sleep study ancillary to MESA occurred shortly after MESA exam 5 during
891 2010-2013. Sleep study participants underwent single night in-home polysomnography
892 (Compumedics Somte Systems, Abbotsville, Australia, AU), as previously described (28). The
893 number of individuals with each type of data and at each time point (exam 1 and exam 5) varies.
894 **Figure S1** in the Supplementary Information visualizes the data flow and overlaps across the
895 various measures used in this study: whole-genome genotyping, RNA-seq, and sleep. The
896 study was approved by Institutional Review Boards in all study centers and participants
897 provided written informed consent.

898

899 [The Hispanic Community Health Study/Study of Latinos \(HCHS/SOL\)](#)

900

901 The HCHS/SOL is a longitudinal cohort study of U.S. Hispanics/Latinos (30,31) recruited from
902 four geographic regions: Bronx NY, Chicago IL, Miami FL, and San Diego CA. The HCHS/SOL
903 baseline exam occurred on 2008-2011, where 16,415 participants were enrolled via multi-stage
904 probability sampling. HCHS/SOL individuals who consented further participated in an in-home
905 sleep study, using a validated type 3 home sleep apnea test recording airflow (via nasal
906 pressure), oximetry, position, and snoring (ARES Unicorder 5.2; B-Alert). Genetic data were
907 measured and imputed to the TOPMed freeze 5b reference panel as previously described, for
908 individuals who consented at baseline (32,33). More information about genotyping and

909 imputation is provided in the Supplementary Information. Metabolomic data were also measured
910 for n~4,000 individuals selected at random out of those with genetic data (34). **Figure S2** in
911 the Supplementary Information provides the data flow in HCHS/SOL, focusing on individuals
912 with genetic data and wide consent for genetic data sharing. All participants provided written
913 informed consent at their recruitment site and the study was approved by the institutional review
914 boards at all participating institutions.

915 [Genotyping and imputation in HCHS/SOL](#)

916
917 Blood was drawn from HCHS/SOL participants during the baseline exam. Individuals who con-
918 sented to genetic studies were genotyped using an Illumina Omni2.5M array, which included
919 150,000 custom-selected Single Nucleotide Polymorphisms (SNPs) including ancestry-informa-
920 tive and Amerindian-specific variants. Global ancestry proportions measuring the proportion of
921 the genome inherited from European, African, and Amerindian ancestors and genetic principal
922 components were computed as previously reported [1]. The genotypes were imputed to the
923 Trans-Omics in Precision Medicine (TOPMed) freeze 5b reference panel as described in [2].

924 925 [The Women's Health Initiative \(WHI\)](#)

926
927 The WHI is a prospective national health study focused on identifying optimal strategies for pre-
928 venting chronic diseases that are the major causes of death and disability in postmenopausal
929 women [3]. The WHI initially recruited 161,808 women between 1993 and 1997 with the goal of
930 including a socio-demographically diverse population with racial/ethnic minority groups propor-
931 tionate to the total minority population of US women aged 50-79 years. The WHI consists of two
932 major parts: a set of randomized Clinical Trials and an Observational Study. The WHI Clinical
933 Trials (CT; N=68,132) includes three overlapping components, each a randomized controlled
934 comparison: the Hormone Therapy Trials (HT), Dietary Modification Trial, and Calcium and Vita-
935 min D Trial. A parallel prospective observational study (OS; N = 93,676) examined biomarkers

936 and risk factors associated with various chronic diseases. While the HT trials ended in the mid-
937 2000s, active follow up of the WHI-CT and WHI-OS cohorts has continued for over 25 years
938 with the accumulation of large numbers of diverse clinical outcomes, risk factor measurements,
939 medication use, and many other types of data.

940 A total of 11,071 WHI participants have whole-genome sequencing data via TOPMed, and
941 1,274 of these participants have RNA-seq measured in venous blood via TOPMed.

942
943 [RNA sequencing in WHI](#)

944
945 RNA-seq was performed via the Trans-Omics in Precision Medicine (TOPMed) program. The
946 WHI RNA samples (N=1,335) were collected from Long Life Study (LLS) participants as part of
947 the LLS Blood Protocol using the PreAnalytiX PAXgene blood tubes, a collection system de-
948 signed to preserve RNA from whole blood. After collection in participant's homes throughout the
949 US, PAXgene tubes were the last of five tubes drawn from each participant, mixed carefully (in-
950 verted 8-10 times), kept at room temperature for a minimum of 2 hours post draw, and shipped
951 overnight with cool packs to the Fred Hutch Specimen Processing Lab (SPL). Upon receipt at
952 the SPL, PAXgene tubes were stored at -80 degrees C until they could be transferred to the
953 Fred Hutch Public Health Sciences Biomarker Lab, where the vials were kept frozen at -80 de-
954 grees C. Within about a month of collection, the lab extracted total RNA, including miRNA, using
955 the PreAnalytiX method (*PAXgene Blood miRNA Kit Handbook, Qiagen, 05/2009*) designed for
956 use with the PAXgene blood collection tubes. A qualitative assessment by agarose gel electro-
957 phoresis of RNA integrity was done at the time of extraction. The RNA was quantified by
958 NanoDrop. The elution volume of 76 μ L of extracted RNA was divided between two RNA 'Par-
959 ent' vials without further dilution, frozen at -80 degrees C, and shipped overnight on dry ice to
960 the WHI biorepository for long-term storage at -80 degrees C. RNA sequencing for WHI was
961 performed at the Broad Institute using the unified TOPMed protocols. More information about

962 RNA sequencing protocols in TOPMed is available here [https://github.com/broadinstitute/gtex-](https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md)
963 [pipeline/blob/master/TOPMed_RNAseq_pipeline.md](https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md).

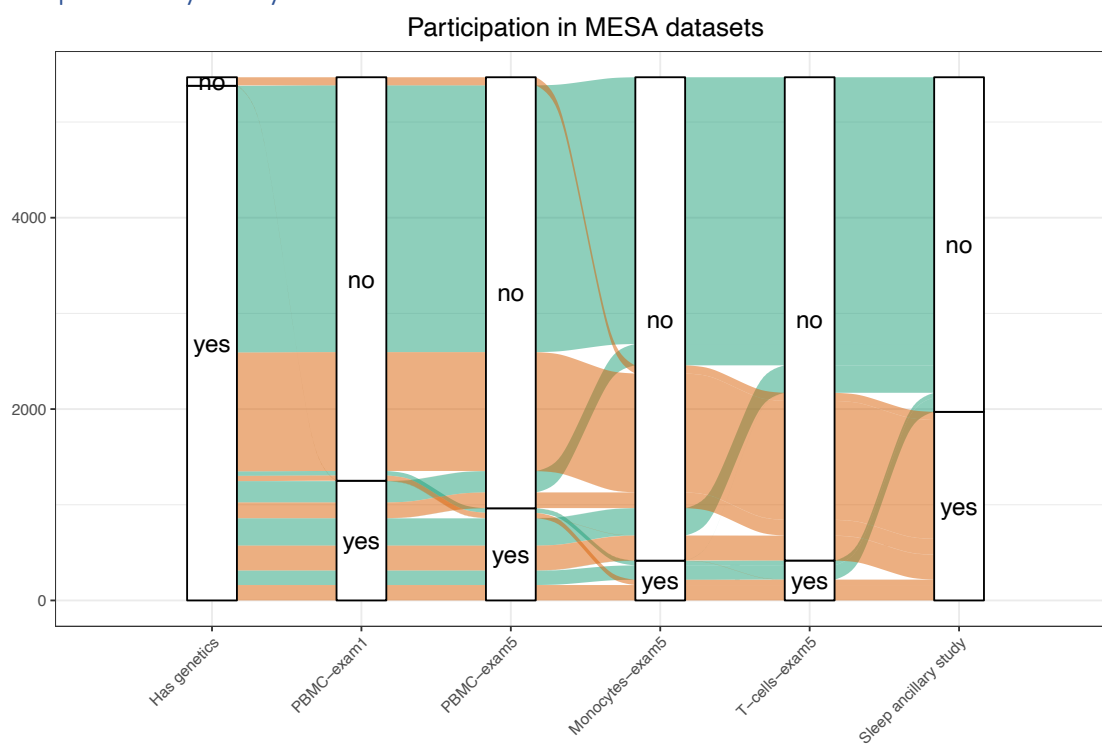
964

965

966 Supplemental figures

967

Figure S1: MESA data flow across various measures in the 1st and 5th exam, and the sleep ancillary study.

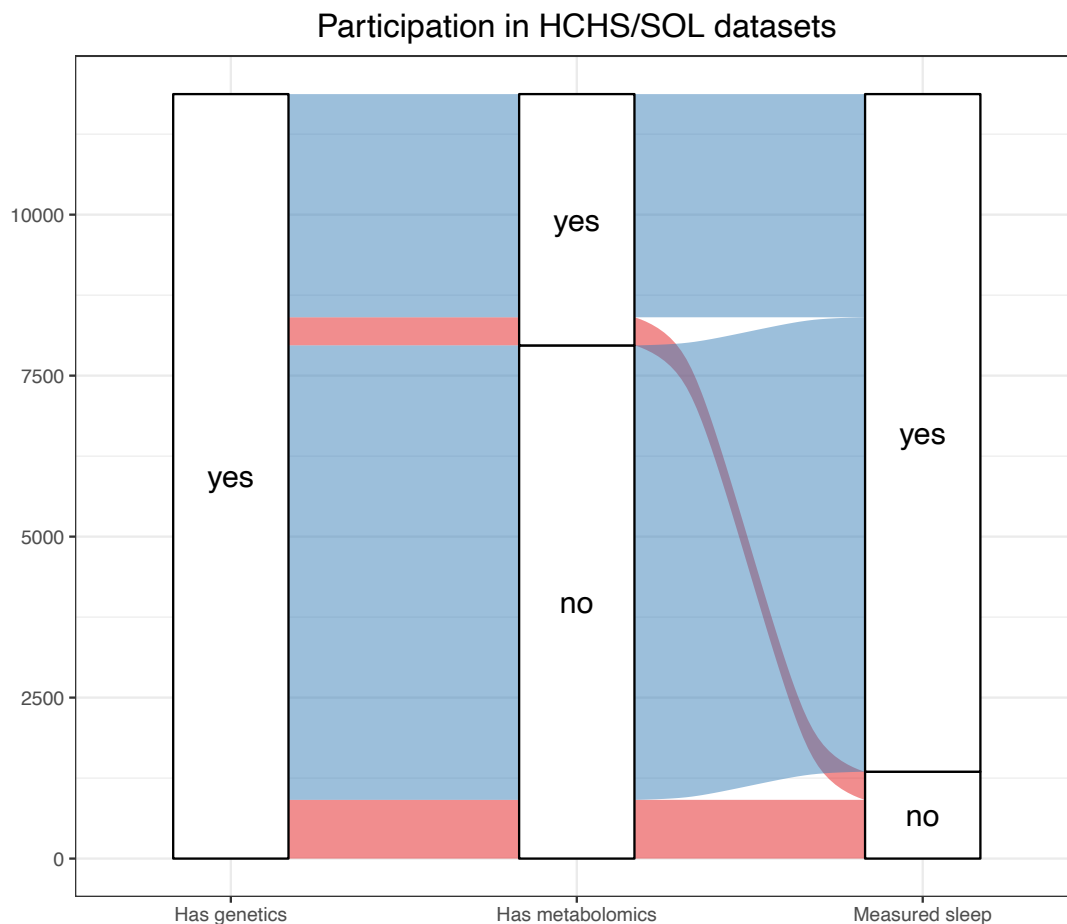


The figure focuses on n=5,468 individuals having at least one of the displayed measures.

968

969

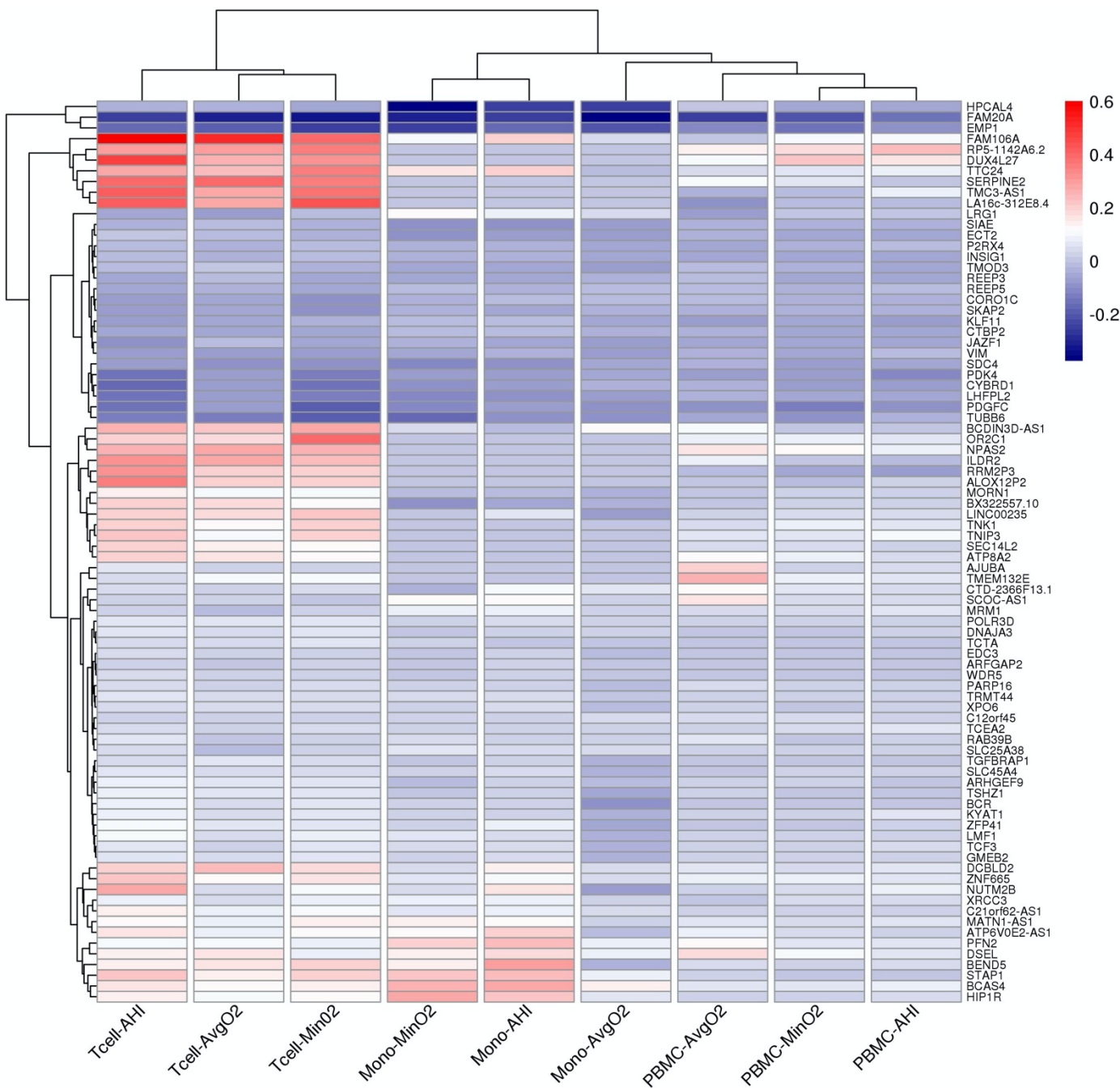
Figure S2: HCHS/SOL data flow across genotyping, metabolomics, and sleep data.



The figure focuses on n=11,872 with genetic data.

970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987

Figure S3: Heatmap of estimated log-fold gene expression change with SDB phenotypes across tissues without BMI adjustment in MESA



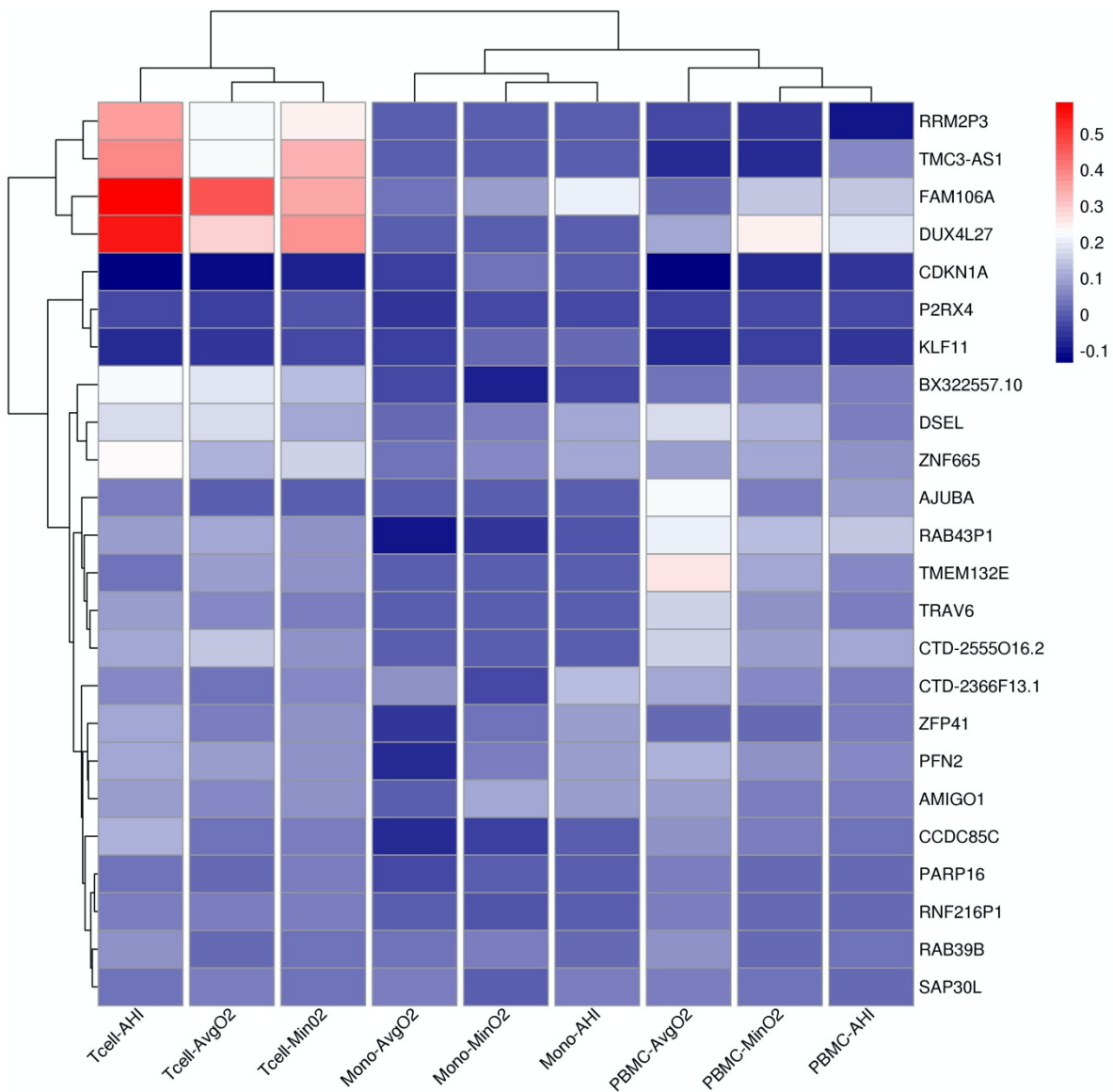
Genes displayed in this figure are those that had FDR p-value<0.1 in association analysis without BMI adjustment. FDR adjustment was computed separately in each set of associations defined by cell type and SDB phenotype. Red color indicated

988

989

990

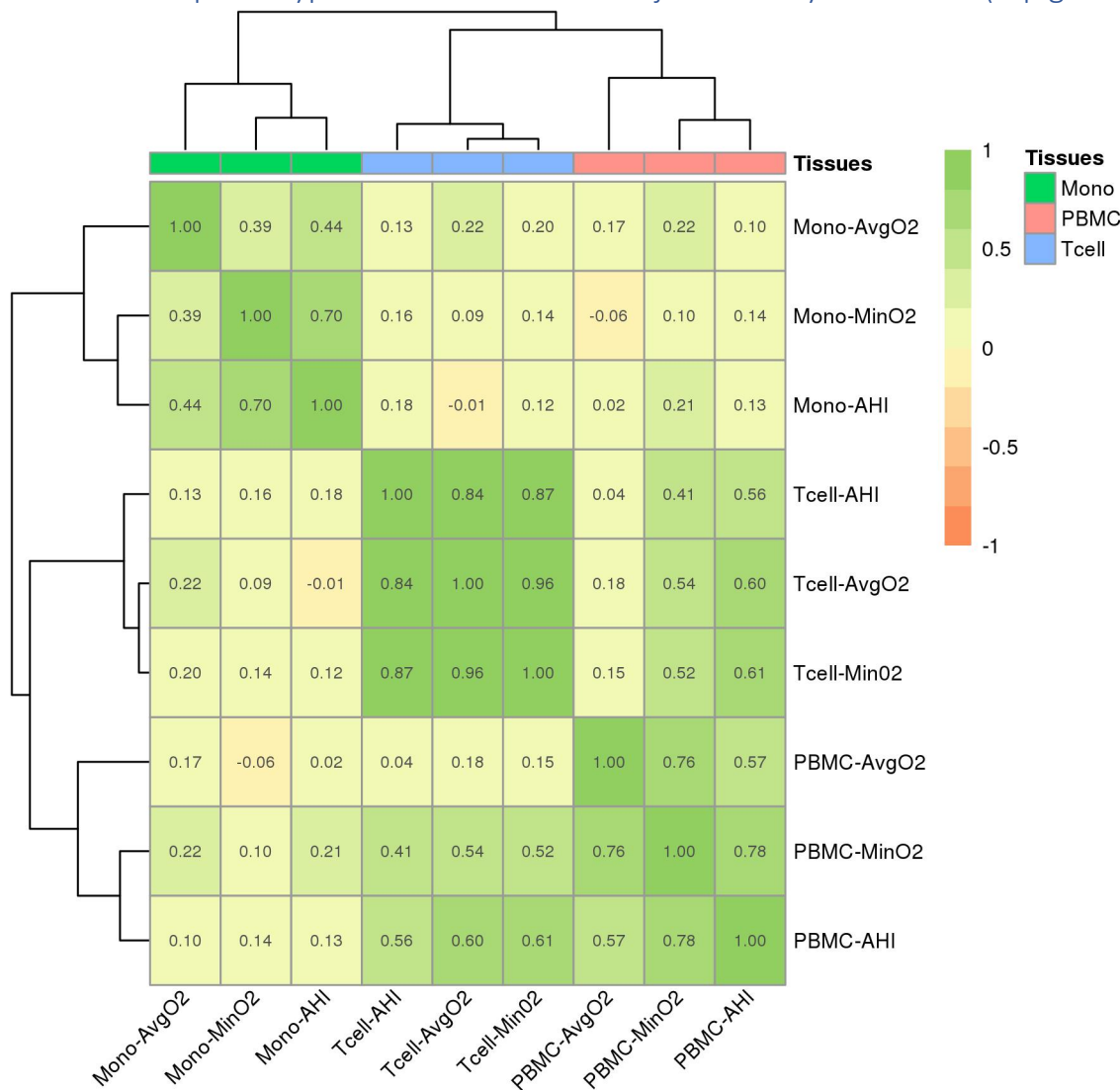
Figure S4: Heatmap of estimated log-fold gene expression change with SDB phenotypes across tissues from BMI-adjusted analysis in MESA



Genes displayed in this figure are those that had FDR p-value<0.1 in association analysis with BMI adjustment. FDR adjustment was computed separately in each set of associations defined by cell type and SDB phenotype.

991
992
993
994
995
996
997

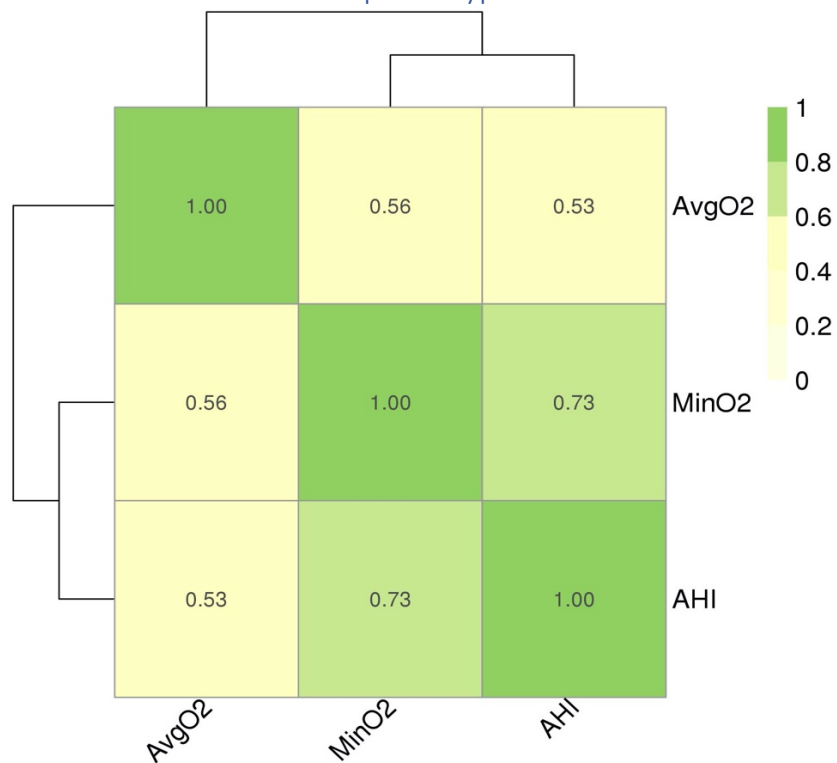
Figure S5: Spearman correlations between estimated log-fold changes in gene expression across SDB phenotypes and tissues in BMI adjusted analysis in MESA (top genes)



Heatmap illustrating the Spearman correlations of log-fold change of transcript expression by tissue type (monocytes, T-cells, PBMCs) and SDB phenotype (AvgO2, MinO2, AHI) in MESA. Correlations were computed over genes with FDR $p < 0.1$. Color legend portrays Spearman R^2 (no/weak correlation = light yellow; complete/strong correlation = green). Estimated AHI effect sizes are flipped prior to computation of correlations so that they match the direction of MinO2 and AvgO2.

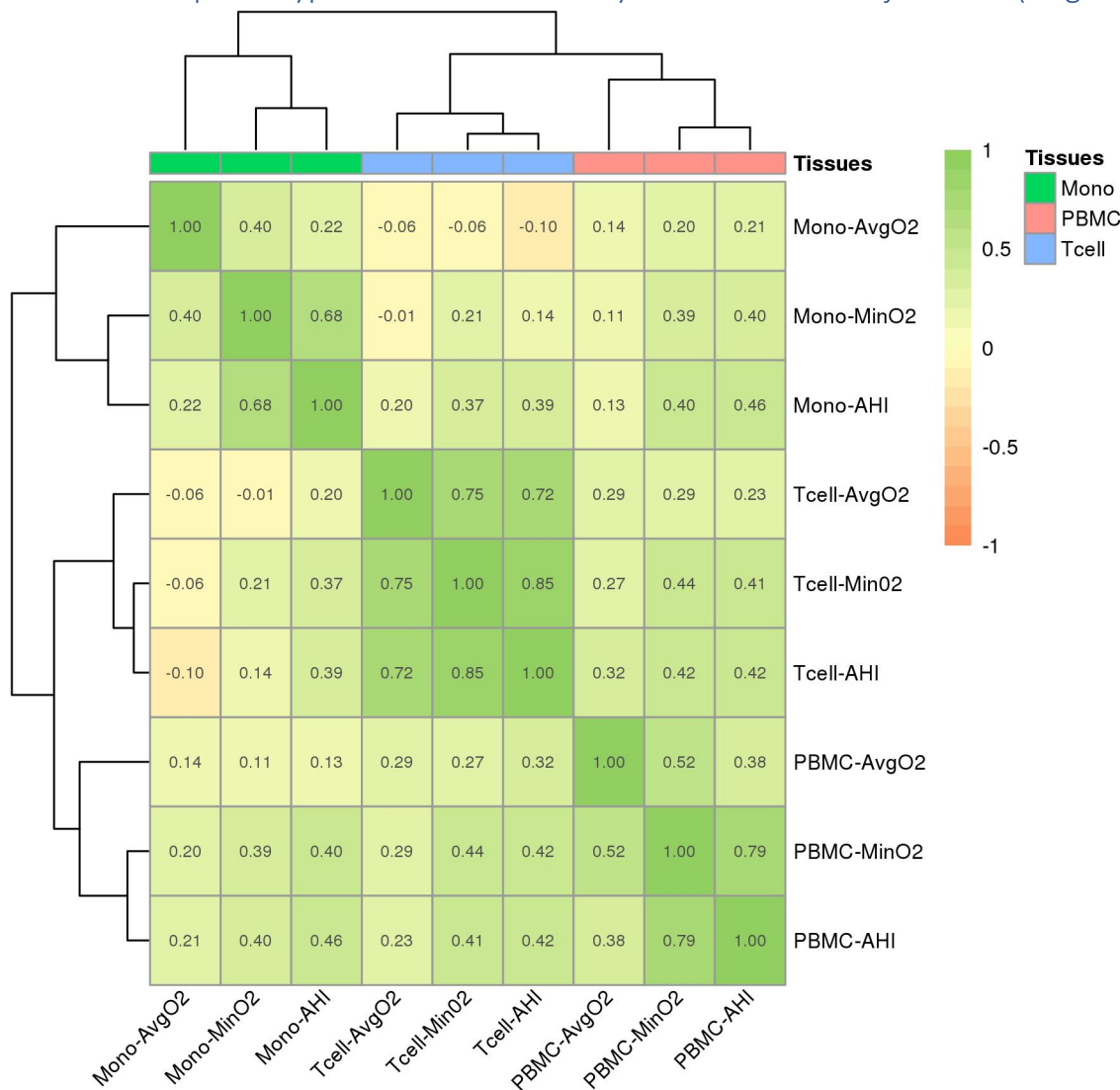
998
999
1000
1001

Table S6. Correlation between the SDB phenotypes



Heatmap illustrating the Spearman correlations SDB phenotype (AvgO2, MinO2, AHI). Color legend portrays Spearman R^2 (no/weak correlation = light yellow; complete/strong correlation = green).

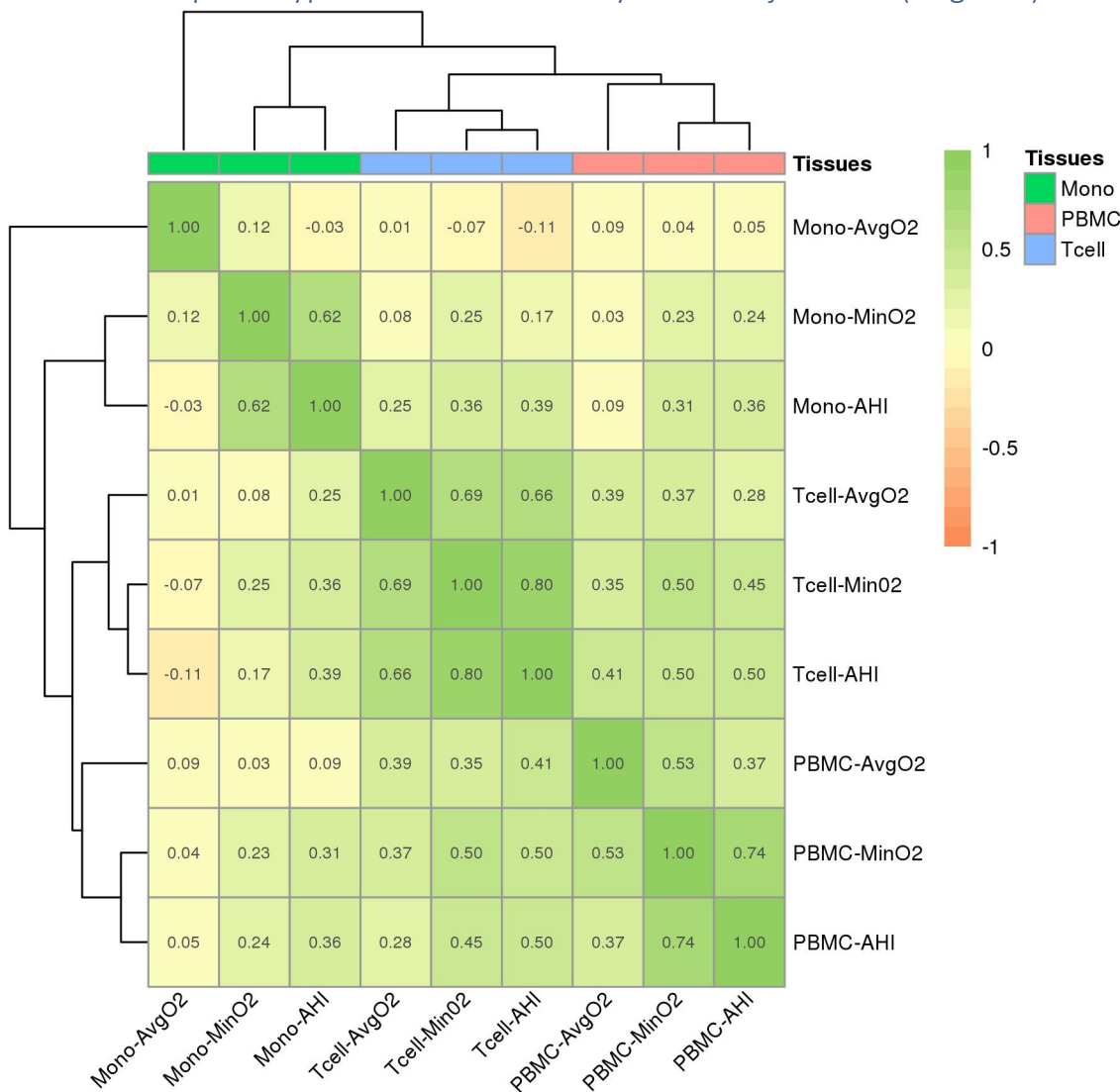
Figure S7: Spearman correlations between estimated log-fold changes in gene expression across SDB phenotypes and tissues in analysis without BMI adjustment (all genes).



Heatmap illustrating the Spearman correlations of log-fold change of transcript expression by tissue type (monocytes, T-cells, PBMCs) and SDB phenotype (AvgO2, MinO2, AHI). Color legend portrays Spearman R^2 (no/weak correlation = light yellow; complete/strong correlation = green).

1003
1004
1005
1006
1007
1008

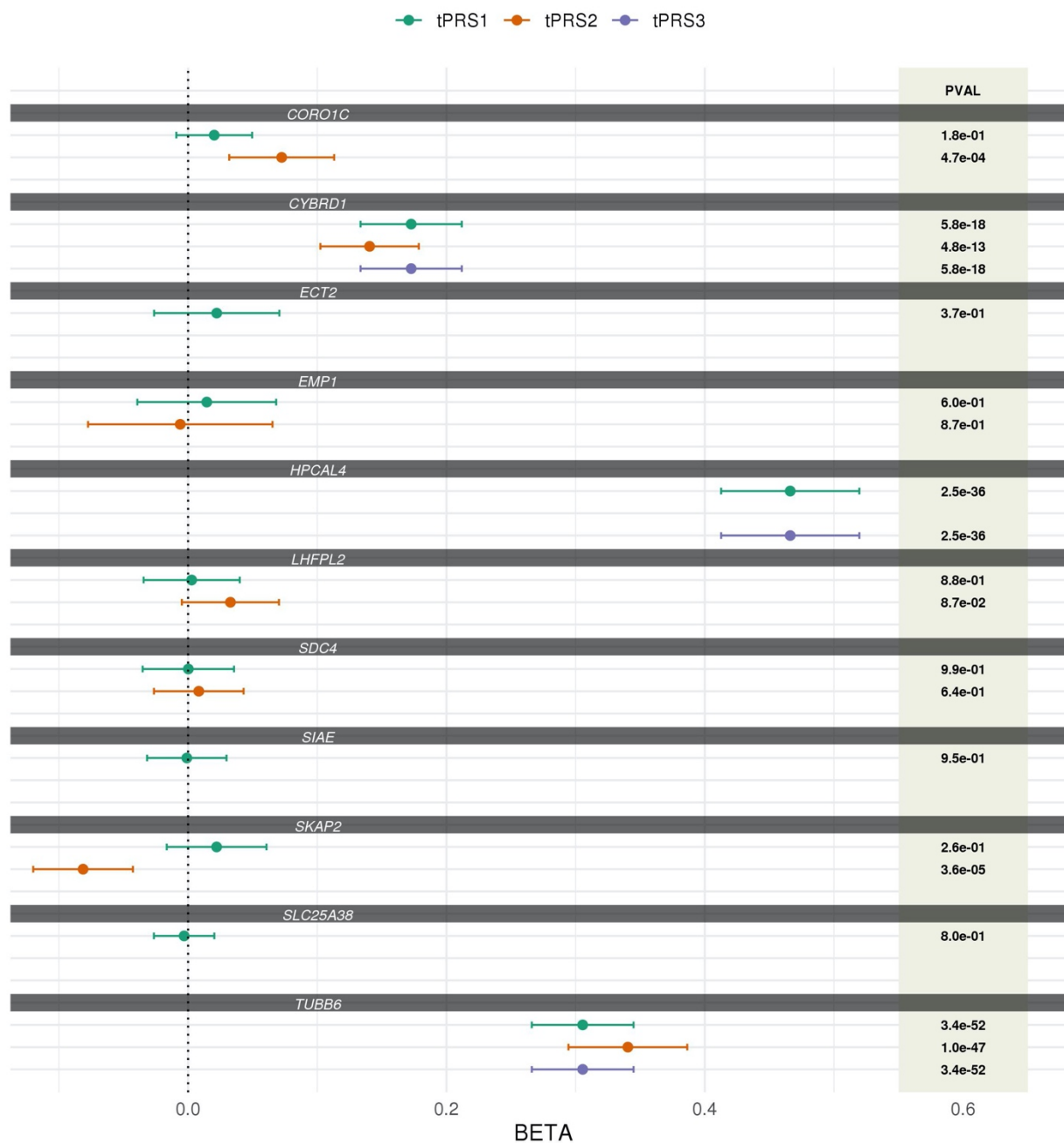
Figure S8: Spearman correlations between estimated log-fold changes in gene expression across SDB phenotypes and tissues in analysis BMI adjustment (all genes).



Heatmap illustrating the Spearman correlations of log-fold change of transcript expression by tissue type (monocytes, T-cells, PBMCs) and SDB phenotype (AvgO2, MinO2, AHI). Color legend portrays Spearman R^2 (no/weak correlation = light yellow; complete/strong correlation = green).

1009
1010
1011
1012
1013
1014
1015
1016
1017

Figure S9. Comparison of the associations between monocyte-based tPRSs and whole-blood gene expression in WHI



For each transcript associated with an SDB phenotype, the figure provides the estimated association effect, 95% confidence interval, and p-value, of tPRSs constructed in different approaches with whole-blood transcript expression in WHI. tPRS1 and tPRS3 were constructed using the clump and threshold approach implemented in PRSice2 using summary statistics from GWAS of transcript expression in monocytes in MESA, and with clumping guided by LD in MESA (the same individuals used for GWAS). tPRS1 allows for genome-wide SNPs, and tPRS3 focused on cis-eQTLs. tPRS2 is the prediXcan model (also using cis-eQTLs only). For tPRS1 and 3 we considered three p-value threshold (5×10^{-8} , 10^{-7} , and 10^{-6}), and the tPRS with smallest p-value is displayed. PRS associations were estimated in models adjusted for sex, age, study site, race/ethnic background, batch effects, and 11 ancestral principal components.

1018
1019

1020 **References**

- 1021
- 1022 1 Conomos, M.P. *et al.* (2016) Genetic diversity and association studies in US hispanic/latino
1023 populations: applications in the hispanic community health study/study of latinos. *Am. J.*
1024 *Hum. Genet.* 98, 165–184
 - 1025 2 Kowalski, M.H. *et al.* (2019) Use of >100,000 NHLBI Trans-Omics for Precision Medicine
1026 (TOPMed) Consortium whole genome sequences improves imputation quality and detection
1027 of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.*
1028 15, e1008500
 - 1029 3 Hays, J. *et al.* (2003) The Women’s Health Initiative recruitment methods and results. *Ann*
1030 *Epidemiol* 13, S18-77
- 1031