Genes associated with depression and coronary artery disease are enriched for inflammation and cardiomyopathy-associated pathways.

Kritika Singh[1,2], Hyunjoon Lee[3,4,5]*, Julia M Sealock[1,2]*, Tyne Miller-Flemming[1,2], Peter Straub[1,2], Nancy J. Cox[1,2], Quinn S. Wells[6], Jordan W. Smoller[3,4,5], Emily C. Hodges[2,7], Lea K. Davis[1,2,8,9,10]

*These authors contributed equally

#Corresponding author

[1]Division of Genetic Medicine, Department of Medicine
Vanderbilt University Medical Center
Nashville, TN, 37232, USA

[2]Vanderbilt Genetics Institute
Vanderbilt University Medical Center
Nashville, TN, 37232, USA

[3]Psychiatric and Neurodevelopmental Genetics Unit
Center for Genomic Medicine
Massachusetts General Hospital
Boston, MA, 02114, USA

[4]Center for Precision Psychiatry
Department of Psychiatry
Massachusetts General Hospital
Boston, MA, 02114, USA

[5]Stanley Center for Psychiatric Research
Broad Institute of Harvard and MIT
Cambridge, MA, 02114,USA

[6]Division of Cardiovascular Medicine, Department of Medicine,
Vanderbilt University Medical Center,
Nashville, TN, 37232, USA

[7]Department of Biochemistry,
Vanderbilt University School of Medicine,

36 Nashville, TN, 37232, USA
37
38 [8]Department of Molecular Physiology and Biophysics
39 Vanderbilt University Medical Center
40 Nashville, TN, 37232, USA
41
42 [9]Department of Psychiatry and Behavioral Sciences
43 Vanderbilt University Medical Center
44 Nashville, TN, 37232, USA
45
46 [10]Departments of Medicine and Biomedical Informatics
47 Vanderbilt University Medical Center
48 Nashville, TN, 37232, USA
49
50 **Total word count of manuscript: 4,656 Words**
51
52 Short Title: Genes shared between depression and CAD
53

Abstract (140 words)

Background: Depression and Coronary Artery Disease (CAD) are highly comorbid conditions. Approximately 40% of individuals who have one diagnosis will also develop the other within their lifetime. Prior research indicates that polygenic risk for depression increases the odds of developing CAD even in the absence of clinical depression. However, the specific genes and pathways involved in comorbid depression-CAD remain unknown.

Results: We identified genes that are significantly associated with both depression and CAD, and are enriched for pathways involved in inflammation and for previous association with cardiomyopathy. We observed increased rate of prevalent, but not incident, cardiomyopathy cases in individuals with comorbid depression-CAD compared to those with CAD alone in three electronic large health record (EHR) datasets.

Conclusions: The results of our study implicate genetically regulated inflammatory mechanisms in depression-CAD. Our results also raise the hypothesis that depression-associated CAD may be enriched for cardiomyopathy.

69 Clinical Perspective

70 *A. What's New?*

71 *1.* Gene associations shared between depression and CAD are enriched for prior

72 association with cardiomyopathy phenotypes.

73 *2.* Cardiomyopathy is significantly more prevalent in individuals with comorbid

74 depression-CAD than in CAD or depression alone.

75 *B. What are the Clinical Implications?*

76 1. Our work suggests that individuals with comorbid depression-CAD may benefit from

77 screening for cardiomyopathy.

78

79 Non-Standard Abbreviations and Acronyms

Abbreviation	Definition
CAD	Coronary artery disease
CI	Confidence interval
CAD	Coronary Artery Disease
CVD	Cardiovascular disease
EHR	Electronic health record
GWAS	Genome-wide association study
ICD	International Classification of Diseases
LD	Linkage disequilibrium
MD	Major depression
MDD	Major depressive disorder
MHC	Major histocompatibility complex
MR	Mendelian randomisation
OR	Odds ratio
PGS	Polygenic score
SNP	Single nucleotide polymorphism
VUMC	Vanderbilt University Medical Center

80

81 **Introduction**

82 Chronic complex diseases such as cardiovascular disease (CVD) are the primary drivers
83 of premature death among individuals with psychiatric disorders. Approximately 17 to 44
84 percent of patients with coronary artery disease (CAD), the most common type of CVD, also
85 have a diagnosis of major depression (MDD), a common mental health diagnosis[1]. Several
86 studies suggest that these two conditions are biologically related. A diagnosis of major
87 depression is associated with increased 18-month cardiac mortality among CAD patients; and
88 CAD with comorbid MDD reduces lifespan by 15-20 years[2–5]. However, despite the frequency
89 and high mortality rate of comorbid CAD with depression and/or MDD, hereafter referred to as
90 (major) depressive CAD or (m)dCAD, the biological relationship between these conditions
91 remains poorly understood.

92 Previously, we demonstrated that common genetic risk for depression identified by
93 genome-wide association studies (GWAS) and quantified in the form of polygenic risk scores
94 (PRS), is associated with a diagnosis of CAD and myocardial infarction in a healthcare-based
95 clinical population[6]. High genetic liability to MDD was associated with increased risk of
96 cardiovascular disease even among patients with no history of psychiatric illness and after
97 accounting for cardiovascular disease risk factors[6]. Shared inflammatory processes provide one
98 possible explanation for these findings[7,8]. Studies show that as CAD progresses, changes in the
99 levels of inflammatory biomarkers such as C-Reactive protein (CRP), leukocytes, monocytes and
100 inflammation associated prothrombotic markers including platelets, can be observed[9–11].
101 Similarly, onset and severity of depression is associated with changes in the levels of immune
102 and inflammatory factors including leukocytes, CRP, and platelets[12–22]. Neuroinflammation and

103 peripheral inflammation are hypothesized to play an important role in both MDD and CAD
104 respectively, providing a potential common biological pathway that may link
105 neuroinflammation in depression together with atherosclerotic inflammation in CAD[7,8].
106 Consortium efforts over the past decade have yielded results from large genome-wide
107 association studies of depression, and independently, CAD[23,24]. These studies provide a map of
108 statistical associations between common single nucleotide polymorphisms (SNPs) and each
109 diagnosis, respectively. However, the genes that are implicated by these SNP associations must
110 also be identified to define the biological pathways that might be shared between these
111 conditions. Transcriptome wide association scans (TWAS) employ functional data to map SNPs
112 to genes based on their ability to regulate gene expression[25]. In this study, we first used the
113 TWAS method S-MultiXcan to map depression and CAD SNP-associations to the genes they
114 regulate using cross-tissue expression quantitative trait loci (eQTL) annotations[25]. We then
115 performed a series of enrichment and pathway analyses to characterize the gene set that was
116 significantly associated with both depression and CAD. Our findings suggest that genes
117 associated with both CAD and depression are enriched for both inflammatory pathways and
118 genes previously associated with cardiomyopathy.
119 Next, we developed a clinical hypothesis based on these findings and tested it in three
120 separate large electronic health records (EHR) data sets –Vanderbilt University Medical Center
121 EHR, All of Us Research Program, and Mass-General Brigham EHR. At each site, we find that the
122 rate of prevalent cardiomyopathy is significantly higher in subjects with (m)dCAD compared to
123 those with CAD alone. Taken together, results of our study link depression and CAD through

124 genetically controlled inflammatory mechanisms, and further suggest a relationship between
125 (m)dCAD and cardiomyopathy.

126 **Materials and Methods**

127 *Genome-wide Association Study Data – Coronary Artery Disease*

128 We utilized publicly available summary statistics from a recently published GWAS of
129 CAD as the foundation for our MetaXcan and S-MultiXcan analyses[23]. This study included a
130 genome-wide meta-analysis of CAD cases and controls from the UK Biobank (UKBB) and
131 CARDIoGRAMplusC4D[23]. CAD was defined in the UKBB by ICD10 codes (I21 – I25), procedural
132 codes (K40-K46, K49, K50 and K75), and self-report of heart attack/myocardial infarction,
133 coronary angioplasty +/- stent, coronary artery bypass graft surgery, or triple heart bypass.
134 Controls were defined by the absence of features used to define cases and the absence of any
135 family history of "heart disease". Application of these criteria resulted in 34,541 CAD cases and
136 261,984 controls. The second sample used in this meta-analysis was the CARDIoGRAMplusC4D
137 GWAS which itself was a meta-analysis of twenty-eight studies of European or South Asian
138 descent populations. CAD was operationally defined differently across the cohorts, but all
139 definitions relied on clinical or procedural markers of CAD, evidence of myocardial infarction, or
140 imaging data consistent with a CAD diagnosis (PMC3679547). The CARDIoGRAMplusC4D GWAS
141 included 63,746 CAD cases and 130,681 controls. The meta-analysis of UKBB and
142 CARDIoGRAMplusC4D totaled 122,733 CAD cases and 424,528 controls.

143

144 *Genome-wide Association Study Data – Depression*

145 We used summary statistics from the PGC and UKBB as the base for the MetaXcan
146 analysis of depression[24]. The genome-wide association meta-analysis for depression included
147 170,756 cases and 329,443 controls. These results included summary statistics from a meta-
148 analysis of the 33 cohorts of the Psychiatric Genomics Consortium as described in Wray et al.
149 (2018), (https://doi.org/10.1038/s41588-018-0090-3) and the broad depression phenotype in
150 the full release of the UKBB as described in (Howard et al. (2018),
151 https://doi.org/10.1038/s41467-018-03819-3). The broad depression phenotype in UKBB
152 included self-reported help-seeking behavior for "nerves, anxiety, tension or depression (once
153 at any visit)" from a general practice physician or a psychiatrist. Secondly, a subject was defined
154 as a case if there was a primary or secondary diagnosis of a depressive mood disorder from
155 linked hospital admission records (UK Biobank fields: 41202 and 41204; ICD codes: F32—Single
156 episode depression, F33—Recurrent depression, F34—Persistent mood disorders, F38—Other
157 mood disorders and F39—Unspecified mood disorders). The remaining respondents were
158 classed as controls if they answered "No" to both questions on all assessments. For the PGC
159 sample, cases were those who met international consensus criteria (DSM-IV, ICD-9, or ICD-10)
160 for a lifetime diagnosis of MDD established using structured diagnostic instruments from
161 assessments by trained interviewers, clinician-administered checklists, or medical record
162 review. Controls in most samples were screened for the absence of lifetime MDD (22/29
163 samples) or randomly selected from the population.
164
165 Statistical Analysis of Genetic Data
166 *S-MultiXcan Analysis of Depression and CAD GWAS Summary Statistics*

167 Predicted expression models of 22,207 genes across 49 tissues were developed
168 (supplementary materials) as previously described using data from the Genotype Tissue
169 Expression Project[25,26]. These models were then used to train gene-based predicted
170 associations for each set of summary statistics described above[25,26]. The subsequent
171 associations with CAD and depression were tested using a multivariate regression model
172 (Summary-MultiXcan or S-MultiXcan) which uses cross-tissue eQTL information[25]. S-MultiXcan
173 infers the gene-level MultiXcan association results, using univariate S-PrediXcan results and LD
174 information from a reference panel (i.e., the GTEx data set). Correlations between tissues are
175 accounted for using a pseudo-inverse approach which applies a singular value decomposition
176 (SVD) of the covariance matrix to keep only the components of large variation.
177 We used a Bonferroni correction to adjust the statistical significance threshold for
178 20,971 gene-based tests in the CAD MetaXcan analysis ($p < 2.38e-06$) and 20,945 gene-based
179 tests in the depression MetaXcan analysis ($p < 2.38e-06$).
180
181 *Enrichment Analysis of Gene-based Association Results (Depression and CAD)*
182 After identifying genes associated with either depression or CAD, we determined
183 whether genes associated with both CAD and depression (n=185) were more abundant than
184 expected by chance, based on the number of genes significantly associated with each
185 phenotype independently. We performed a hypergeometric test using the hyper R package and
186 the publicly available website http://nemates.org/MA/progs/overlap_stats.html.
187
188 *Gene Set Enrichment Analyses (GSEA)*

189 We used the Gene Set Enrichment Analysis Web-based Tool (http://www.gsea-
190 msigdb.org/gsea/msigdb/annotate.jsp) to functionally annotate the pathways for the 185 genes
191 associated with *both* depression and CAD. We restricted our analysis to the 'canonical
192 pathways' class with FDR q-values less than 0.05 and the GTEx compendium expression profiles.
193 There were 2,982 pathways, composed of 40,786 genes, tested in the null 'background' set,
194 provided by GSEA.
195
196 Description of Clinical Populations
197 *Vanderbilt University Medical Center (VUMC) Electronic Health Record (EHR)*
198 Vanderbilt University Medical Center (VUMC) is a tertiary care center that provides
199 inpatient and outpatient care in middle Tennessee and surrounding communities. The VUMC
200 electronic health record (EHR) system was established in 1990 and includes data on billing
201 codes from the International Classification of Diseases, 9th and 10th editions (ICD-9 and ICD-
202 10), Current Procedural Terminology (CPT) codes, laboratory values, reports, and clinical
203 documentation. A fully de-identified mirror image of the EHR, called the Synthetic Derivative, is
204 available to Vanderbilt faculty for research purposes[27].
205
206 *Massachusetts General Brigham (MGB) EHR*
207 Mass General Brigham (MGB) is a hospital network that includes Massachusetts General
208 Hospital, Brigham and Women's hospital, and other community and specialty hospitals in
209 Boston area. The data source of MGB EHR is MGB Research Patient Data Registry (RPDR;
210 https://rpdrssl.partners.org), an EHR database which spans more than 20 years of data from

211 over 6.5 million patients and includes data on diagnoses (billing codes; ICD-9 and ICD-10 codes),
212 procedures (CPT codes), laboratory values, and clinical notes. Data floors were applied to the
213 MGB RPDR to include only patients with at least one clinical note and three visits since 2005
214 with visit more than 30 days apart.
215
216 *All of Us Research Program*
217 We used data from the *All of Us* Research Program, a multi-ancestry population-based cohort
218 that contains various forms of information on individuals[28]. This data repository enrolls
219 participants 18 years of age or older and contains their demographic and medical record-based
220 information. The data for disease diagnosis was collected using billing codes and was the
221 converted to phecodes.
222
223 *Phenotypic analysis in the Synthetic Derivative*
224 Informed by results of the genomic analyses, we first investigated the prevalence of
225 cardiomyopathy in three mutually exclusive groups of patients in the VUMC-EHR including
226 those who had a diagnosis of (1) depression or major depression *without* comorbid CAD (dep),
227 (2) CAD *without* comorbid depression (CAD), and (3) CAD *with* comorbid depression or major
228 depression ((m)dCAD)(Table S1). To reduce missingness of clinical data and enrich the sample
229 for patients who receive their primary care at VUMC, we applied a simple "medical home"
230 heuristic requiring the presence of any five codes on different days over a period of at least
231 three years. We selected the cohort of individuals meeting these criteria who also had not been
232 included in the VUMC BioBank which resulted in a total of 988,002 individuals. The genotyped

233 sample was held out of this analysis to maintain its independence for future related research
234 studies including genomic data.
235 We defined dep-only, CAD-only, and (m)dCAD using phecodes which are higher order
236 combinations of at least two related ICD codes, occurring on two different days, using the R
237 PheWAS package. Specifically, “CAD-only” cases required the presence of phecode 411.4 and
238 absence of 296.2 phecode (N = 47,351), “dep-only” cases included phecode 296.2 and absence
239 of 411.4 phecode (N = 57,069) and (m)dCAD required the presence of a phecode for both CAD
240 and depression (N = 6,725). We next calculated the prevalence of cardiomyopathy, indicated by
241 the presence of one or more of the following phecodes "425.1", "425.11", "425.12", "425.8",
242 "425". We then compared the prevalence of cardiomyopathy diagnoses between each
243 diagnostic group using Pearson’s Chi-squared test and Fisher’s exact T-test.
244 We then used a logistic regression model to determine whether cardiomyopathy was
245 indeed more prevalent among patients with (m)dCAD, compared to CAD alone, after controlling
246 for potential confounding. Cases were defined as those with cardiomyopathy, controls were
247 those without cardiomyopathy. The exposure was defined as depression, and features known
248 to associate with both depression and cardiomyopathy, age, race, sex, record median BMI, type
249 2 diabetes (T2D; Phecode = 250.2), tobacco use disorder (TUD; Phecode = 318) and
250 hypertension (Phecodes = 401, 415.21, 453), were included as covariates.
251 Next, we used the same regression-based approach to test whether *incident*
252 cardiomyopathy was increased among individuals with (m)dCAD compared to CAD alone. All
253 individuals with cardiomyopathy coded prior to depression or CAD (N = 14,583) were excluded.
254 Hence the cardiomyopathy exposure was required to be coded after both depression and CAD.

255 This resulted in a total of 2,345 cardiomyopathy cases and 45,330 cardiomyopathy controls. We
256 then fitted two multivariable logistic regression models to test whether depression increased
257 the odds of incident cardiomyopathy among patients with (m)dCAD compared to patients with
258 CAD alone, after adjusting for the same previously described confounders.
259
260 Sensitivity analyses using Severe Depression.
261 Previous literature points towards progression and worsening of CAD with increasing
262 depression severity. To test whether cardiomyopathy more commonly co-occurred with severe
263 depression, we redefined our exposure using a stricter MDD definition requiring the presence
264 of 296.22 phecode and ICD9/10 codes (Table S8) which dropped the sample size from 57,069
265 depression-exposed patients to 41,423 MDD-exposed patients. The same regression models
266 described above were fitted to the data.
267
268 Phenotypic Replication in MGB and *All of Us* Data
269 We replicated the EHR analyses in two external cohorts, *All of Us* and MGB. We used the
270 same phecode based definitions to define our inclusion and exclusion criteria, case and control
271 labels, and exposures. Sensitivity analyses were similarly performed in these data by restricting
272 the exposure definition to include individuals with only severe depression.
273
274
275
276 Results

277 Genes Associated with Both Depression and CAD

278 We intersected the 22,207 predicted expression gene models with the summary
279 statistics of CAD and depression phenotypes as described above. While the tissues that best
280 predicted gene expression differed between CAD and depression (Figure S4); the mean
281 (Spearman's rank = 0.433), standard deviation (Spearman's rank = 0.565), minimum
282 (Spearman's rank = 0.389), and maximum (Spearman's rank = 0.373) association Z-score
283 statistics demonstrated strong correlation between depression and CAD across all tissues
284 tested (Figure S3).

285 Results from the 20,971 gene-based tests of association with CAD, and 20,945 gene-
286 based tests of association with depression are available in Table S2 and Table S3,
287 respectively[23,24]. We identified a total of 1,455 genes significantly associated with CAD ($p <$
288 $2.38e-06$) and 928 genes significantly associated with depression ($p < 2.38e-06$) illustrated in
289 Figures S1 and S2, respectively. We next identified the genes that were significantly associated
290 with *both* depression and CAD. A total of 20,944 genes were tested in common between
291 depression and CAD, and of these, 185 genes identified by their gene names were significantly
292 associated with *both* CAD and depression (Figure 2; Table S4). The genes associated with both
293 depression and CAD were distributed across the genome (Figure 2).

294

295 Enrichment and Pathway Analysis of 185 Genes Associated with Both Depression and CAD

296 We found that there were three times more genes associated with both CAD and
297 depression than expected by chance given the number of genes significantly associated with
298 each phenotype independently (hypergeometric test, $p<1.718e-43$) (Figure 3). Inflammatory

299 (FDR q-value < 0.05) and cardiomyopathy-associated pathways (FDR q value < 0.05) were
300 overrepresented among genes associated with both CAD and depression (Table 1). Five genes
301 from the shared set of 185 (2.7%) were annotated to cardiomyopathy-associated pathways.
302 Four of the 5 genes (*ATP2A2, ITGB4, SGCD*, and *CACNB4)* were annotated to all three
303 cardiomyopathy pathways (Table S4). Thirteen additional genes (7%) were annotated to
304 adaptive immune pathways (Table S4).
305
306 Clinical Cardiomyopathy in (m)dCAD
307 Analyses performed in the EHR returned significant differences in the prevalence of
308 cardiomyopathy between depression-only, CAD-only and (m)dCAD cohorts (Chisq P < 2.2e-16)
309 (Table S5). Almost 15% of individuals with (m)dCAD had a cardiomyopathy diagnosis, compared
310 to 11.2% in the CAD cohort and only 1.59 % in the depression cohort. The results remained
311 significant when we compared only the (m)dCAD group to the CAD group (Fisher's exact test p <
312 2.2e-16) (Table S5). We then replicated these findings in the MGH EHR and the All of Us Data.
313 We next restricted the depression phenotype to major depression (phecode 296.22) and
314 observed the same pattern in the prevalence of cardiomyopathy between the groups (Table
315 S6).
316
317 Prevalence of Clinical Cardiomyopathy after Adjusting for known Confounders
318 Results of the multivariable logistic model indicate that there is a 27% increase in the
319 odds of prevalent cardiomyopathy with a diagnosis of (m)dCAD compared to CAD after
320 adjusting for age, race, sex, hypertension diagnosis, smoking status, type-2-diabetes and BMI

321 (see Methods for details) (p-value = 3.6e-09, OR = 1.27, 95% CI = 1.17-1.3). This finding was
322 replicated in the *All of Us* (p-value = 0.012, OR = 1.16, 95%CI = 1.03-1.30) and MGH (p-value <
323 2e-16, OR = 1.32, 95%CI = 1.28-1.37) data sets.

324 Sensitivity analyses restricting the criteria for depression to "major depression"
325 preserved the effect estimate and significance of the original finding (OR = 1.14, 95%CI = 1.06-
326 1.26, p = 6.5e-03)(Figure 4). This finding was again replicated in the MGH cohort (OR = 1.23,
327 95%CI = 1.13-1.18, p < 2e-16). While the effect estimate was similar in the *All of Us* data set, the
328 finding did not reach statistical significance likely due to a reduced sample size (OR = 1.13,
329 95%CI = 0.97-1.32, p-value = 0.12)

330

331 <u>Rate of Incident Cardiomyopathy in (m)dCAD compared to CAD</u>

332 Lastly, we tested the hypothesis that (m)dCAD is a risk factor for the development of
333 *subsequent* cardiomyopathy. Using a multivariable logistic regression model, we found no
334 increase in the odds of incident cardiomyopathy among (m)dCAD patients compared to CAD
335 patients, after adjusting for known confounders in the VUMC EHR (OR = 0.96, 95% CI = 0.84-
336 1.09, p-value = 0.53) (Figure 5) and in the *All of Us* data (OR = 0.864, 95%C.I. = 0.730-1.021, p-
337 value = 0.086). In the MGH data, not only did we observe a lack of evidence for increased
338 incident cardiomyopathy subsequent to (m)dCAD, in fact we observed that (m)dCAD cases were
339 significantly *less* likely to be diagnosed with new onset cardiomyopathy compared to CAD alone
340 (OR = 0.783, 95%C.I. = 0.743-0.825, p-value < 2e-16).

341 After restricting the definition of depression to "major depression" in the VUMC data,
342 the odds of incident cardiomyopathy were elevated but not significantly so (OR = 1.16, 95% C.I.

343 = 1.00-1.34, p-value = 0.04). However, again, in the *All of Us* (OR = 0.66, 95% C.I. = 0.51-0.84, p-
344 value = 9e-4) and MGH (OR = 0.54, 95% C.I. = 0.50-0.58, p-value < 2 e-16) data, the rate of
345 incident cardiomyopathy was lower in (m)dCAD than in CAD alone.
346
347
348 Discussion
349 Depression and CAD are highly comorbid conditions and the molecular mechanisms
350 underlying this common comorbidity remain understudied. Here we present an integrative
351 genotypic approach to identify genes that map to SNP associations for each condition, followed
352 by a phenotypic analysis to test a hypothesis motivated by the genetic findings. By performing a
353 cross-tissue transcriptome-wide association scan using large scale GWAS data for depression
354 and CAD, we identified 185 genes (three time more than expected by chance) that were
355 significantly associated with both depression and CAD. Interestingly, the genes associated with
356 both phenotypes were not necessarily those most strongly associated with depression or CAD
357 independently. On balance, these results suggest that comorbid (m)dCAD may have a partially
358 unique genetic architecture which overlaps a subset of genes associated with depression and
359 CAD.
360 Pathway analyses indicate that the genes involved in (m)dCAD are enriched for
361 inflammatory and cardiomyopathy-associated pathways. Specifically, we observed an
362 enrichment for adaptive immune system genes which can be understood in the context of
363 previous studies that have highlighted immune dysfunction in both depression and CAD[15,29–36].
364 Moreover, research has shown that depression is associated with low-grade chronic

365 inflammation whereas CAD has phases of both, acute and chronic inflammation[36–38]. Hence the
366 enrichment of adaptive immune system genes suggests that depression and CAD could be
367 linked through exposure to low-grade chronic inflammation.

368 There were six genes *ATP2A2, ITGB4, SGCD, CACNB4, ACTN2*, and *TGFB1* associated with
369 both CAD and depression, which are known to be involved in cardiomyopathy (Table S4).
370 Differential expression of *ATP2A*, which encodes a Ca^{2+} pump on the endoplasmic reticulum, is
371 associated with cardiac phenotypes including dilated cardiomyopathy[39,,40]; and heterozygous
372 conditional knockout mice demonstrate an essential role for *ATP2A* in neuronal calcium
373 homeostasis resulting in behavioral phenotypes[41]. Similarly, *CACNB4* is the one of the most
374 abundant voltage-gated calcium channel subunits expressed in the brain where it is critical for
375 presynaptic signaling, and co-located in the heart where early data suggests it may be involved
376 in cardiac contraction[42–44]. The gene locus was also recently implicated in idiopathic
377 cardiomyopathy in individuals of African-American ancestry[45]. *ACTN2*, a cardiac-specific
378 structural protein, is classically involved in familial hypertrophic, dilated, and arrhythmogenic
379 cardiomyopathy[46–48] while *TGFB1* is a cytokine with a variety of cellular functions implicated in
380 cardiac hypertrophy and hypertrophic cardiomyopathy along with congenital heart disease[49,50].
381 Mutations in delta-sarcoglycan (*SGCD)* result in severe neuromuscular disease in humans and
382 knockout mice develop early progressive cardiomyopathy and dilated cardiomyopathy[51,52].
383 However, less is known about the impact of differences in expression of these genes (*ACTN2*,
384 *TGFB1*, and *SGCD*) on depression.

385 Interpreted in the context of previous studies, our genetic findings suggest that
386 depression and CAD could share low grade chronic inflammation[36,53–55]. Moreover, genetic

387 results of our study raised the hypothesis that a predisposition to both depression and CAD
388 (clinically observed as (m)dCAD) may further predispose individuals to cardiomyopathy.
389 Canonically, viral infections are considered the most common trigger and cause of
390 inflammatory cardiomyopathy thus leading to immune mechanisms which potentially damage
391 the myocardial function[56–63]. However, an alternative model consistent with our findings could
392 position chronic low-grade inflammation as a shared risk factor for depression, CAD, and
393 cardiomyopathy. We investigated the hypothesis that cardiomyopathy co-occurs more
394 commonly with (m)dCAD than with CAD, by harnessing the power of large-scale electronic
395 health record data at VUMC, MGB, and *All of Us*. These three systems represent both hospital
396 ascertainment (VUMC and MGB) and volunteer ascertainment (*All of Us*). Results of these
397 analyses suggest that while prevalent cardiomyopathy does in fact co-occur more frequently
398 with (m)dCAD than with CAD alone, the order of events remains unclear. For example, it
399 remains possible that (a) cardiomyopathy increases the risk of co-morbid depression in CAD, (b)
400 treatment for depression provides some protection against subsequent cardiomyopathy, or (c)
401 both cardiomyopathy and (m)dCAD share additional risk factors. While our work demonstrating
402 the clinical co-occurrence suggests that cardiomyopathy patients may benefit from depression
403 screening, future work is needed to disentangle their precise relationship and inform clinically
404 translatable mitigation strategies.
405 Despite its strengths, our study is limited by the well-established caveats of EHR based
406 phenotyping which itself suffers from individual bias of the clinician. TWAS analyses rely on
407 SNP-based predictive models of mRNA trained using mostly European-ancestry individuals in
408 GTEx v8 and assumes additivity of SNP effects on gene expression, which ignores the possibility

409 of epistatic and gene-environment interactions. Online pathway tools are prone to knowledge
410 bias in selection of the "validated" genes annotated to pathways.
411 Nevertheless, these analyses highlight the unique genetic and phenotypic architecture
412 of (m)dCAD compared to CAD and depression alone. We add to the literature the observation
413 that prevalent, but perhaps not incident cardiomyopathy, is also more common in patients with
414 (m)dCAD compared to CAD alone. Importantly, this hypothesis was motivated by genetic
415 findings suggesting that the genes associated with both depression and CAD are enriched for
416 pathways with a known role in immune and cardiomyopathy-associated biological processes.
417 This data-driven approach highlights the power of *in silico* transcriptome-wide studies to
418 motivate hypotheses that can be tested in large EHR databases.
419
420 Acknowledgements
421
422 **Funding**
423 KS is funded by the American Heart Association Fellowship AHA827137. LKD is supported by
424 R56MH120736. JWS and LKD are supported in part by NIMH R01 H118233. JMS is funded by
425 1F31MH124306-01A1. QSW was supported by NIH 1R01HL140074.
426
427 <u>CTSA (SD, Vanderbilt Resources)</u>
428 The deidentified EHR used at VUMC was supported by the National Center for Research
429 Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing

Translational Sciences, Grant 2 UL1 TR000445-06. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

BioVU

The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at https://victr.vumc.org/biovu-funding/.

The *All of Us* Research Program is supported by grants through 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 OD026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA#: AOD 16037; Federally Qualified 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. In addition to the funded partners, the *All of Us Research Program* would not be possible without the contributions made by its participants.

Declarations

453 Contributions

454 K.S. and L.K.D. conceptualized and designed the work. K.S, J.M.S, and H.L. implemented the
455 computational procedures and performed data analysis. N.J.C, Q.S.W., and E.C.H provided
456 important clinical and intellectual insights. All authors read, edited, and approved the final
457 manuscript.

458

459 Corresponding author

460 Correspondence to Lea K. Davis at lea.k.davis@vumc.org

461

462 **Ethics declarations**

463 Ethics approval and consent to participate

464 This study was reviewed by the VUMC IRB and designated as non-human subjects research
465 because of the use of fully de-identified data (IRB# 190418 and IRB# 201609).

466

467 Competing interests

468 JWS is a member of the Scientific Advisory Board of Sensorium Therapeutics (with equity), and
469 has received an honorarium for an internal seminar Tempus Labs. He is PI of a collaborative
470 study of the genetics of depression and bipolar disorder sponsored by 23andMe for which
471 23andMe provides analysis time as in-kind support but no payments.

472 **Supplementary Materials**

473 Supplementary Methods

Supplementary Figure 1

Supplementary Figure 2

Supplementary Figure 3

Supplementary Figure 4

Supplementary Table 1

Supplementary Table 2

Supplementary Table 3

Supplementary Table 4

Supplementary Table 5

Supplementary Table 6

485 **References**

486 1. Kendler, K. S., Gardner, C. O., Fiske, A. & Gatz, M. Major depression and coronary artery
487 disease in the Swedish twin registry: phenotypic, genetic, and environmental sources of
488 comorbidity. *Arch. Gen. Psychiatry* **66**, 857–863 (2009).
489 2. Glassman, A. H. Depression and cardiovascular comorbidity. *Dialogues Clin. Neurosci.* **9**, 9–17
490 (2007).
491 3. Sullivan, M. *et al.* Depression in coronary heart disease. What is the appropriate diagnostic
492 threshold? *Psychosomatics* **40**, 286–292 (1999).
493 4. O'Neil, A. *et al.* Depression is a risk factor for incident coronary heart disease in women: An
494 18-year longitudinal study. *J. Affect. Disord.* **196**, 117–124 (2016).
495 5. Weeke, A., Juel, K. & Vaeth, M. Cardiovascular death and manic-depressive psychosis. *J.*
496 *Affect. Disord.* **13**, 287–292 (1987).
497 6. Dennis, J. *et al.* Genetic risk for major depressive disorder and loneliness in sex-specific
498 associations with coronary artery disease. *Mol. Psychiatry* **26**, 4254–4264 (2021).
499 7. Mazereeuw, G. *et al.* Platelet activating factors in depression and coronary artery disease: a
500 potential biomarker related to inflammatory mechanisms and neurodegeneration. *Neurosci.*
501 *Biobehav. Rev.* **37**, 1611–1621 (2013).
502 8. McCaffery, J. M. *et al.* Common genetic vulnerability to depressive symptoms and coronary
503 artery disease: a review and development of candidate genes related to inflammation and
504 serotonin. *Psychosom. Med.* **68**, 187–200 (2006).
505 9. Upadhyay, R. K. Emerging risk biomarkers in cardiovascular diseases and disorders. *J. Lipids*
506 **2015**, 971453 (2015).

507 10. Shimokawa, H., Aarhus, L. L. & Vanhoutte, P. M. Porcine coronary arteries with
508 regenerated endothelium have a reduced endothelium-dependent responsiveness to
509 aggregating platelets and serotonin. *Circ. Res.* **61**, 256–270 (1987).
510 11. Brydon, L., Magid, K. & Steptoe, A. Platelets, coronary heart disease, and stress. *Brain.*
511 *Behav. Immun.* **20**, 113–119 (2006).
512 12. Huang, M. *et al.* Longitudinal association of inflammation with depressive symptoms: A
513 7-year cross-lagged twin difference study. *Brain. Behav. Immun.* **75**, 200–207 (2019).
514 13. Beydoun, M. A. *et al.* White blood cell inflammatory markers are associated with
515 depressive symptoms in a longitudinal study of urban adults. *Transl. Psychiatry* **6**, e895
516 (2016).
517 14. Kim, J.-M. *et al.* Changes in pro-inflammatory cytokine levels and late-life depression: A
518 two year population based longitudinal study. *Psychoneuroendocrinology* **90**, 85–91 (2018).
519 15. Bränn, E. *et al.* Inflammatory markers in late pregnancy in association with postpartum
520 depression-A nested case-control study. *Psychoneuroendocrinology* **79**, 146–159 (2017).
521 16. Bremmer, M. A. *et al.* Inflammatory markers in late-life depression: results from a
522 population-based study. *J. Affect. Disord.* **106**, 249–255 (2008).
523 17. Martínez-Cengotitabengoa, M. *et al.* Peripheral Inflammatory Parameters in Late-Life
524 Depression: A Systematic Review. *Int. J. Mol. Sci.* **17**, E2022 (2016).
525 18. Gheysarzadeh, A. *et al.* Serum-based microRNA biomarkers for major depression: MiR-
526 16, miR-135a, and miR-1202. *J. Res. Med. Sci. Off. J. Isfahan Univ. Med. Sci.* **23**, 69 (2018).

527 19. Schmidt, H. D., Shelton, R. C. & Duman, R. S. Functional biomarkers of depression:
528 diagnosis, treatment, and pathophysiology. *Neuropsychopharmacol. Off. Publ. Am. Coll.*
529 *Neuropsychopharmacol.* **36**, 2375–2394 (2011).
530 20. Nurden, A. T. The biology of the platelet with special reference to inflammation, wound
531 healing and immunity. *Front. Biosci. Landmark Ed.* **23**, 726–751 (2018).
532 21. Tsao, C.-W., Lin, Y.-S., Chen, C.-C., Bai, C.-H. & Wu, S.-R. Cytokines and serotonin
533 transporter in patients with major depression. *Prog. Neuropsychopharmacol. Biol. Psychiatry*
534 **30**, 899–905 (2006).
535 22. Pandey, G. N. *et al.* Hyperactive phosphoinositide signaling pathway in platelets of
536 depressed patients: effect of desipramine treatment. *Psychiatry Res.* **105**, 23–32 (2001).
537 23. van der Harst, P. & Verweij, N. Identification of 64 Novel Genetic Loci Provides an
538 Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **122**, 433–
539 443 (2018).
540 24. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102
541 independent variants and highlights the importance of the prefrontal brain regions. *Nat.*
542 *Neurosci.* **22**, 343–352 (2019).
543 25. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves
544 association detection. *PLOS Genet.* **15**, e1007889 (2019).
545 26. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–
546 585 (2013).
547 27. Bowton, E. *et al.* Biobanks and Electronic Medical Records: Enabling Cost-Effective
548 Research. *Sci. Transl. Med.* **6**, (2014).

549 28. The All of Us Research Program Investigators. The “All of Us” Research Program. *N. Engl.*
550 *J. Med.* **381**, 668–676 (2019).
551 29. McCaffery, J. M. *et al.* Common genetic vulnerability to depressive symptoms and
552 coronary artery disease: a review and development of candidate genes related to
553 inflammation and serotonin. *Psychosom. Med.* **68**, 187–200 (2006).
554 30. Upadhyay, R. K. Emerging risk biomarkers in cardiovascular diseases and disorders. *J.*
555 *Lipids* **2015**, 971453 (2015).
556 31. Shimokawa, H., Aarhus, L. L. & Vanhoutte, P. M. Porcine coronary arteries with
557 regenerated endothelium have a reduced endothelium-dependent responsiveness to
558 aggregating platelets and serotonin. *Circ. Res.* **61**, 256–270 (1987).
559 32. Brydon, L., Magid, K. & Steptoe, A. Platelets, coronary heart disease, and stress. *Brain.*
560 *Behav. Immun.* **20**, 113–119 (2006).
561 33. Huang, M. *et al.* Longitudinal association of inflammation with depressive symptoms: A
562 7-year cross-lagged twin difference study. *Brain. Behav. Immun.* **75**, 200–207 (2019).
563 34. Beydoun, M. A. *et al.* White blood cell inflammatory markers are associated with
564 depressive symptoms in a longitudinal study of urban adults. *Transl. Psychiatry* **6**, e895
565 (2016).
566 35. Kim, J.-M. *et al.* Changes in pro-inflammatory cytokine levels and late-life depression: A
567 two year population based longitudinal study. *Psychoneuroendocrinology* **90**, 85–91 (2018).
568 36. Sealock, J. M. *et al.* Use of the PsycheMERGE Network to Investigate the Association
569 Between Depression Polygenic Scores and White Blood Cell Count. *JAMA Psychiatry* **78**, 1365
570 (2021).

571 37. Twig, G. *et al.* White Blood Cell Count and the Risk for Coronary Artery Disease in Young
572 Adults. *PLoS ONE* **7**, e47183 (2012).
573 38. Zebrack, J. S., Muhlestein, J. B., Horne, B. D. & Anderson, J. L. C-reactive protein and
574 angiographic coronary artery disease: independent and additive predictors of risk in subjects
575 with angina. *J. Am. Coll. Cardiol.* **39**, 632–637 (2002).
576 39. Lu, Q. *et al.* Intranuclear cardiac troponin I plays a functional role in regulating Atp2a2
577 expression in cardiomyocytes. *Genes Dis.* S2352304221000611 (2021)
578 doi:10.1016/j.gendis.2021.04.007.
579 40. Alimadadi, A., Munroe, P. B., Joe, B. & Cheng, X. Meta-Analysis of Dilated
580 Cardiomyopathy Using Cardiac RNA-Seq Transcriptomic Datasets. *Genes* **11**, 60 (2020).
581 41. Nakajima, K. *et al.* Brain-specific heterozygous loss-of-function of ATP2A2, endoplasmic
582 reticulum Ca2+ pump responsible for Darier's disease, causes behavioral abnormalities and a
583 hyper-dopaminergic state. *Hum. Mol. Genet.* **30**, 1762–1772 (2021).
584 42. Coste de Bagneaux, P. *et al.* A homozygous missense variant in CACNB4 encoding the
585 auxiliary calcium channel beta4 subunit causes a severe neurodevelopmental disorder and
586 impairs channel and non-channel functions. *PLOS Genet.* **16**, e1008625 (2020).
587 43. Andrade, A. *et al.* Genetic Associations between Voltage-Gated Calcium Channels and
588 Psychiatric Disorders. *Int. J. Mol. Sci.* **20**, 3537 (2019).
589 44. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids*
590 *Res.* **36**, D480–D484 (2007).
591 45. Xu, H. *et al.* A Genome-Wide Association Study of Idiopathic Dilated Cardiomyopathy in
592 African Americans. *J. Pers. Med.* **8**, 11 (2018).

593 46. Good, J.-M. *et al.* ACTN2 variant associated with a cardiac phenotype suggestive of left-
594 dominant arrhythmogenic cardiomyopathy. *Hear. Case Rep.* **6**, 15–19 (2020).
595 47. Haywood, N. J. *et al.* Hypertrophic cardiomyopathy mutations in the calponin-homology
596 domain of ACTN2 affect actin binding and cardiomyocyte Z-disc incorporation. *Biochem. J.*
597 **473**, 2485–2493 (2016).
598 48. Tiso, N. Identification of mutations in the cardiac ryanodine receptor gene in families
599 affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum. Mol.*
600 *Genet.* **10**, 189–194 (2001).
601 49. Patel, R. *et al.* Variants of Trophic Factors and Expression of Cardiac Hypertrophy in
602 Patients with Hypertrophic Cardiomyopathy. *J. Mol. Cell. Cardiol.* **32**, 2369–2377 (2000).
603 50. Yadav, M. L., Bhasker, A. N., Kumar, A. & Mohapatra, B. Identification and
604 characterization of genetic variants of TGFB1 in patients with congenital heart disease. *Meta*
605 *Gene* **31**, 100987 (2022).
606 51. Bauer, R., MacGowan, G. A., Blain, A., Bushby, K. & Straub, V. Steroid treatment causes
607 deterioration of myocardial function in the -sarcoglycan-deficient mouse model for dilated
608 cardiomyopathy. *Cardiovasc. Res.* **79**, 652–661 (2008).
609 52. Rutschow, D. *et al.* S151A δ-sarcoglycan mutation causes a mild phenotype of
610 cardiomyopathy in mice. *Eur. J. Hum. Genet.* **22**, 119–125 (2014).
611 53. Lynall, M.-E. *et al.* Peripheral Blood Cell–Stratified Subgroups of Inflamed Depression.
612 *Biol. Psychiatry* **88**, 185–196 (2020).
613 54. McNally, L., Bhagwagar, Z. & Hannestad, J. Inflammation, Glutamate, and Glia in
614 Depression: *A Literature Review*. *CNS Spectr.* **13**, 501–510 (2008).

615 55. Troubat, R. *et al.* Neuroinflammation and depression: A review. *Eur. J. Neurosci.* **53**,
616 151–171 (2021).
617 56. Eitel, I. *et al.* Inflammation in takotsubo cardiomyopathy: insights from cardiovascular
618 magnetic resonance imaging. *Eur. Radiol.* **20**, 422–431 (2010).
619 57. Adamo, L., Rocha-Resende, C., Prabhu, S. D. & Mann, D. L. Reappraising the role of
620 inflammation in heart failure. *Nat. Rev. Cardiol.* **17**, 269–285 (2020).
621 58. Maisch, B., Ristic, A. D. & Pankuweit, S. [Inflammatory cardiomyopathy and
622 myocarditis]. *Herz* **42**, 425–438 (2017).
623 59. Krejci, J., Mlejnek, D., Sochorova, D. & Nemec, P. Inflammatory Cardiomyopathy: A
624 Current View on the Pathophysiology, Diagnosis, and Treatment. *BioMed Res. Int.* **2016**,
625 4087632 (2016).
626 60. Tschöpe, C. *et al.* Myocarditis and inflammatory cardiomyopathy: current evidence and
627 future directions. *Nat. Rev. Cardiol.* **18**, 169–193 (2021).
628 61. Eitel, I. *et al.* Clinical characteristics and cardiovascular magnetic resonance findings in
629 stress (takotsubo) cardiomyopathy. *JAMA* **306**, 277–286 (2011).
630 62. Scally, C. *et al.* Myocardial and Systemic Inflammation in Acute Stress-Induced
631 (Takotsubo) Cardiomyopathy. *Circulation* **139**, 1581–1592 (2019).
632 63. Rroku, A., Grahl, S., Landmesser, U. & Heidecker, B. A case report of myocardial
633 inflammation in takotsubo syndrome. A chicken-or-the-egg phenomenon. *Int. J. Cardiol.*
634 *Heart Vasc.* **39**, 100958 (2022).
635

636 **Figures and Tables**

637

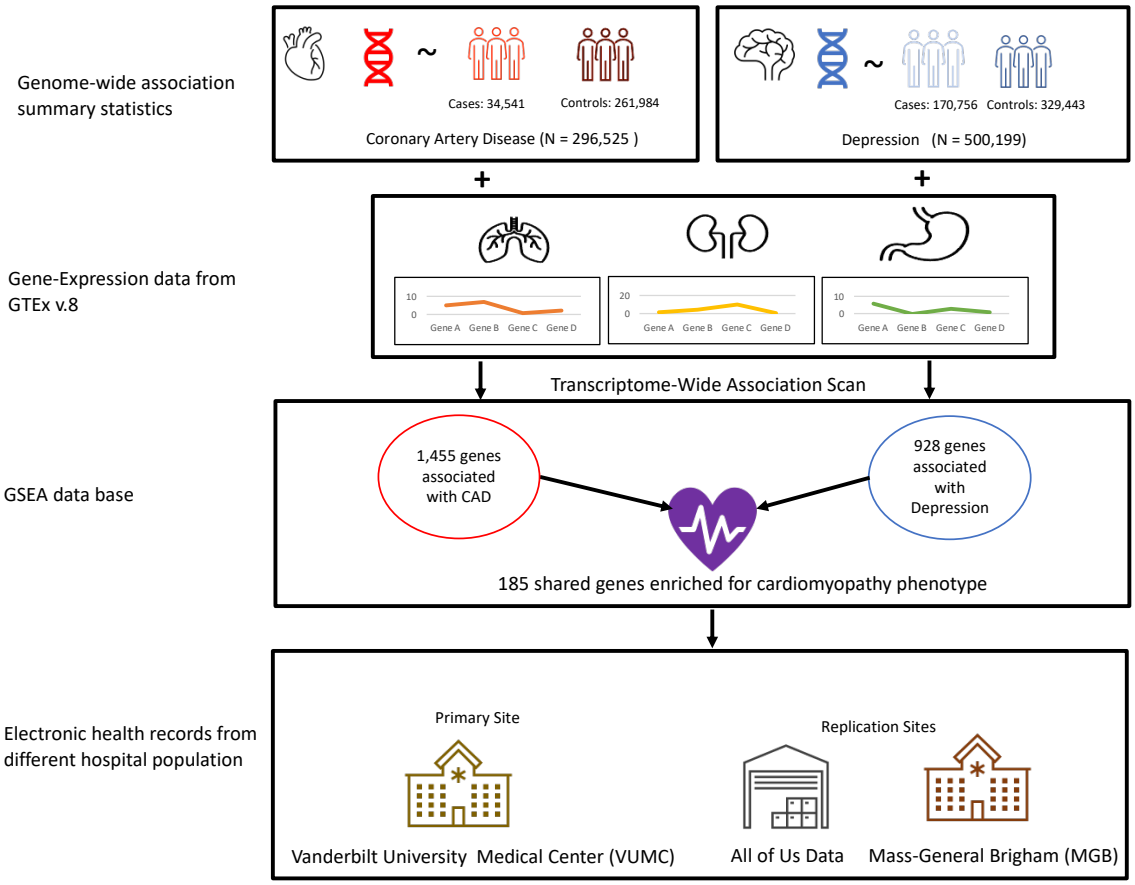


Figure 1

Title: Overall Schematic of the Study

Legend : We performed a transcriptome-wide association study (TWAS) of genetically regulated expression (GReX) using the GTEx v.8 database and genome-wide summary statistics of Depression and Coronary Artery Disease separately. We derived genes that were shared between Depression and CAD, we performed pathway and enrichment analyses. Our results show that genes associated with both depression and CAD are associated with cardiomyopathy associated phenotypes.

638

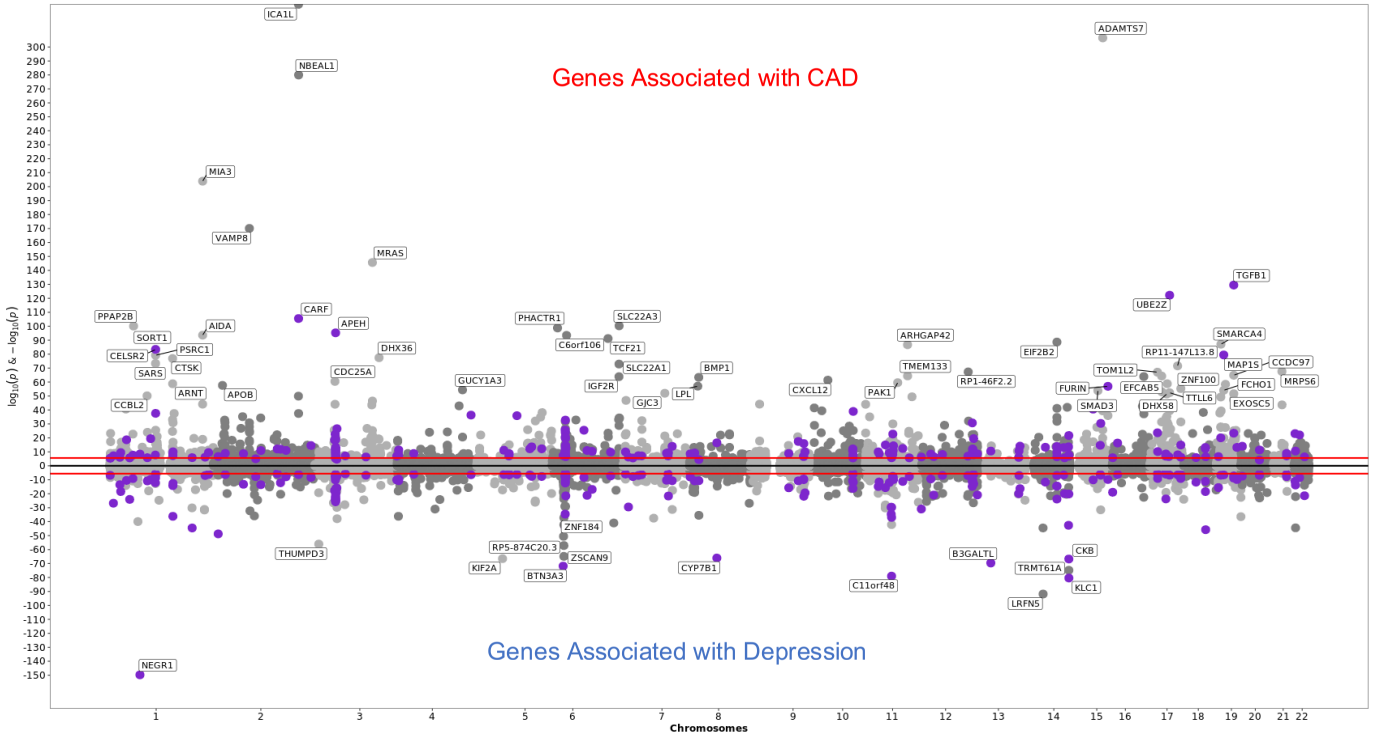


Figure 2

Title: Genes associated with Depression and CAD using S-MultiXcan

Legend: A Miami plot of genes associated with both Depression and CAD via S-MultiXcan Each data point represents a gene grouped by chromosome (x-axis) and p value (y-axis) of the association with the phenotype. The top half of the graph are genes associated with CAD and the lower half of the graph represents genes associated with Depression. The purple dots represent the 185 genes shared between Depression and CAD. Genes with –log(p values) greater that 50 are labelled. The light and dark grey represents the non-shared genes on alternating chromosomes with light grey with being the non-shared genes on odd chromosomes and dark grey representing the non-shared genes on the even numbered chromosomes

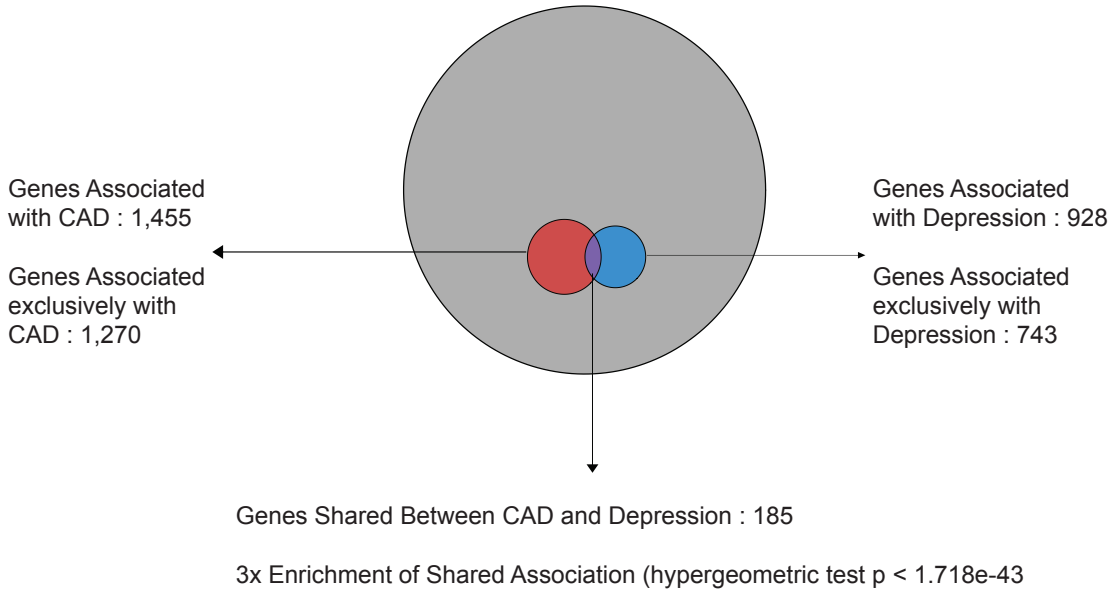


Figure 3

Title : Enrichment test of the genes shared between Depression and CAD

Legend: Eulerr Plot to highlight the results of enrichment test of the genes associated with both depression and CAD. The grey circle represents the set of 20,944 genes that were tested for association with both phenotypes. The red portion represents the 1,270 genes significantly associated with CAD only. The blue portion represents the 743 genes significantly associated with depression only. The purple portion represents the 185 genes associated with both depression and CAD. Red + purple represents all CAD associated genes (1455) and blue + purple represents all genes associated with depression (928).

640

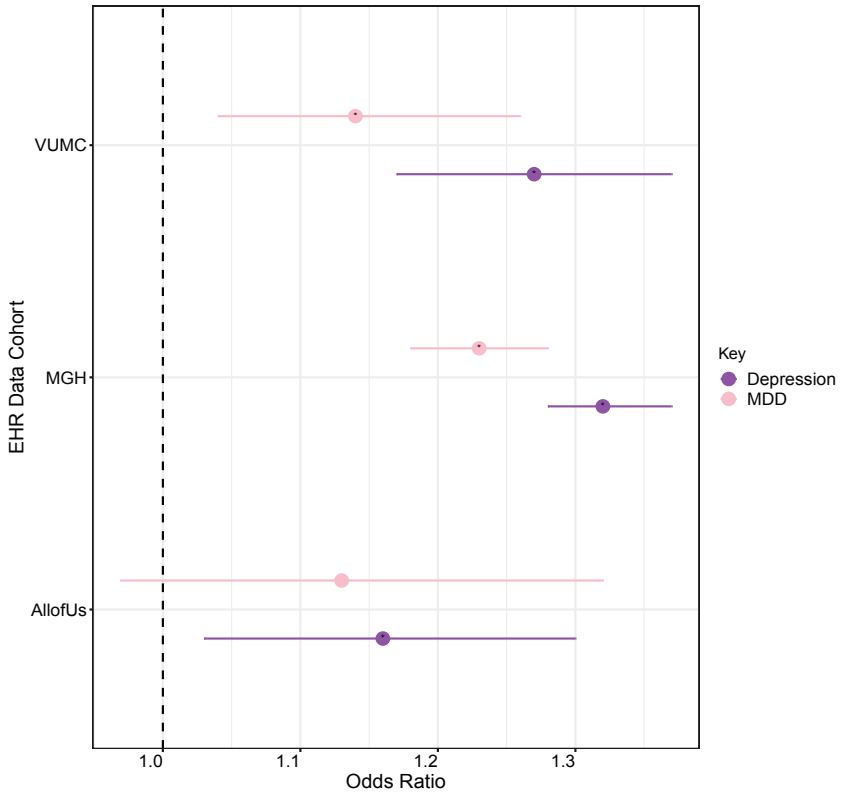


Figure 4

Title: Forest plot illustrating odds of prevalent cardiomyopathy in (m)dCAD compared to CAD

Legend: The association between (m)dCAD vs CAD and prevalent cardiomyopathy after controlling for race, median age across the medical record, sex, T2D, hypertension, and smoking (ever/never) at each of the three sites (VUMC, MGB, and All of Us). Purple represents results from an analysis in which depression was broadly defined while light pink represents results from an analysis restricted to major depression. Whiskers indicate 95% CIs.

641

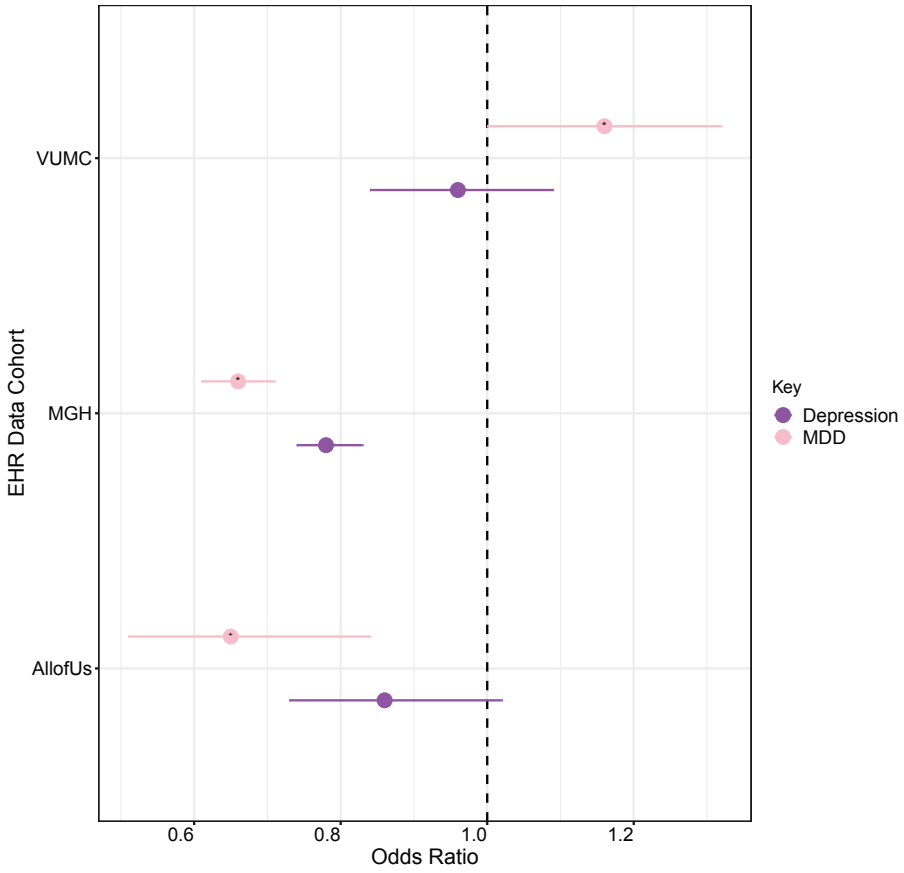


Figure 5

Title: Forest plot illustrating odds of incident cardiomyopathy in (m)dCAD compared to CAD

Legend: The association between (m)dCAD vs CAD and incident cardiomyopathy after controlling for race, median age across the medical record, sex, T2D, hypertension, and smoking (ever/never) at each of the three sites (VUMC, MGB, and All of Us). Purple represents results from an analysis in which depression was broadly defined while light pink

642

Gene Set Name	No. of Genes in the set (K)	Number of Genes in Overlap (k)	k/K	FDR Value
Post-translational protein modification	1435	22	0.0152	1.39e-03
Arrhythmogenic right ventricular cardiomyopathy (ARCV)	76	5	0.0657	1.87e-02
Transport of small molecules	728	13	0.0179	1.87e-02
Hypertrophic cardiomyopathy (HCM)	83	5	0.0602	2.13e-02
Dilated cardiomyopathy	90	5	0.0556	2.62e-02
Adaptive Immune System	825	13	0.0158	3.79e-02

Table 1: Gene Set Enrichment Analysis of the Shared Genes

Legend: We used the 185 genes shared between Depression and CAD as the input for GESA (Canonical Pathways) browser and identified that the shared genes are associated with immune and cardiomyopathy associated phenotypes. K refers to the number of genes in the GSEA database for that particular pathway. k refers to the number of genes from our shared 185 gene-set in the particular pathway

	Depression (broad)	Major Depression	CAD	Depression and CAD	Major Depression and CAD
Sample sizes					
Total (N)	57,069	41,423	47,349	6,725	4,175
Cardiomyopathy	809 (1.41%)	537 (1.29%)	5,314 (11.22%)	1,001 (14.88%)	589 (14.10%)
Demographics					
Female (%)	68.64	69.1	32.86%	48.91%	47.04%
Median Age, years (IQR)	41(24.5 – 65.5)	39 (22.5 – 55.5)	68 (60 – 76)	65 (56.5-72.5)	64 (56 – 72)
Median BMI (IQR)	27.34	27.28	28.81	29.29	29.68
Underweight	2,071 (3.63%)	1,655 (4.0%)	377 (0.79%)	64 (0.95%)	39 (0.93%)
Normal	17,437 (30.56%)	12,964 (31.3%)	8,673 (18.31%)	1,320 (19.62%)	794 (19.02%)
Overweight	14,154 (24.80%)	10,204 (24.63%)	15,520 (32.77%)	1,979 (29.43%)	1,291 (30.93%)
Obese	19,247 (33.73%)	14,203 (34.28%)	17,423 (36.8%)	2,812 (41.81%)	1,951 (46.73%)
Unknown	4,160 (7.29%)	2,397 (5.78%)	5,356 (11.31%)	550 (8.18%)	100 (2.4%)
Cardiovascular Comorbidities					
Hypertension (N, %)	19,784 (34.66%)	13,513 (32.62%)	38,905 (82.17%)	6,200 (92.2%)	3,837 (91.9%)
Type 2 Diabetes (N, %)	7,452 (13.06%)	5,116 (12.35%)	15,207 (32.11%)	3,227 (47.99%)	2,062 (49.39%)
Tobacco Use Disorder (N, %)	8,363 (14.65%)	6,555 (15.82%)	8,482 (17.91%)	2,399 (35.67%)	1,816 (43.5%)

Table 2: Demographic table for Phenotypes in VUMC

This table describes the demographics for Depression, Major Depression, CAD, Comorbid Depression-CAD and Comorbid Major Depression-CAD in VUMC EHR data.

656 Supplementary Materials

657 of

658 Genes associated with depression and coronary artery disease are enriched for

659 inflammation and cardiomyopathy-associated pathways.

660 Kritika Singh[1,2], Hyunjoon Lee[3,4,5]*, Julia M Sealock[1,2]*, Tyne Miller-Flemming[1,2], Peter

661 Straub[1,2], Nancy J. Cox[1,2], Quinn S. Wells[6], Jordan W. Smoller[3,4,5], Emily C. Hodges[2,7], Lea

662 K. Davis[1,2,8,9,10]

663 *These authors contributed equally

664 #Corresponding author

665 [1]Division of Genetic Medicine, Department of Medicine
666 Vanderbilt University Medical Center
667 Nashville, TN, 37232, USA
668
669 [2]Vanderbilt Genetics Institute
670 Vanderbilt University Medical Center
671 Nashville, TN, 37232, USA
672
673 [3]Psychiatric and Neurodevelopmental Genetics Unit
674 Center for Genomic Medicine
675 Massachusetts General Hospital
676 Boston, MA, 02114, USA
677
678 [4]Center for Precision Psychiatry
679 Department of Psychiatry
680 Massachusetts General Hospital
681 Boston, MA, 02114, USA
682
683 [5]Stanley Center for Psychiatric Research
684 Broad Institute of Harvard and MIT
685 Cambridge, MA, 02114,USA
686
687 [6]Division of Cardiovascular Medicine, Department of Medicine,
688 Vanderbilt University Medical Center,

Nashville, TN, 37232, USA

[7]Department of Biochemistry,
Vanderbilt University School of Medicine,
Nashville, TN, 37232, USA

[8]Department of Molecular Physiology and Biophysics
Vanderbilt University Medical Center
Nashville, TN, 37232, USA

[9]Department of Psychiatry and Behavioral Sciences
Vanderbilt University Medical Center
Nashville, TN, 37232, USA

[10]Departments of Medicine and Biomedical Informatics
Vanderbilt University Medical Center
Nashville, TN, 37232, USA

707 Description of Clinical Populations

708 *Vanderbilt University Medical Center (VUMC) Electronic Health Record (EHR)*

709 Vanderbilt University Medical Center (VUMC) is a tertiary care center that provides
710 inpatient and outpatient care in middle Tennessee and surrounding communities. The VUMC
711 electronic health record (EHR) system was established in 1990 and includes data on billing
712 codes from the International Classification of Diseases, 9th and 10th editions (ICD-9 and ICD-
713 10), Current Procedural Terminology (CPT) codes, laboratory values, reports, and clinical
714 documentation. A fully de-identified mirror image of the EHR, called the Synthetic Derivative, is
715 available to Vanderbilt faculty for research purposes. A data floor (i.e., "medical home")
716 heuristic of any five codes on different days over a period of at least three years was imposed
717 to enrich the sample for primary care.

718

719 *Massachusetts General Hospital (MGH)*

720 Mass General Brigham (MGB) is a hospital network that includes Massachusetts General
721 Hospital, Brigham and Women's hospital, and other community and specialty hospitals in
722 Boston area. The data source of MGB EHR is MGB Research Patient Data Registry (RPDR;
723 https://rpdrssl.partners.org), an EHR database which spans more than 20 years of data from
724 over 6.5 million patients and includes data on diagnoses (billing codes; ICD-9 and ICD-10 codes),
725 procedures (CPT codes), laboratory values, and clinical notes. Data floors were applied to the
726 MGB RPDR to include only patients with at least one clinical note and three visits since 2005
727 with visit more than 30 days apart.

728

729 *All of Us Research Program*

730 All of Us Research Program, a population-based cohort that contains demographic, EHR,
731 and survey information on participants. Participants enroll digitally through the All of Us
732 website. After completion of the consent modules and enrollment, participants are given
733 several health-measuring surveys to be completed. As of July 2019, the All of Us program had
734 enrolled more than 175,000 core participants and more than 230,000 total participants.
735

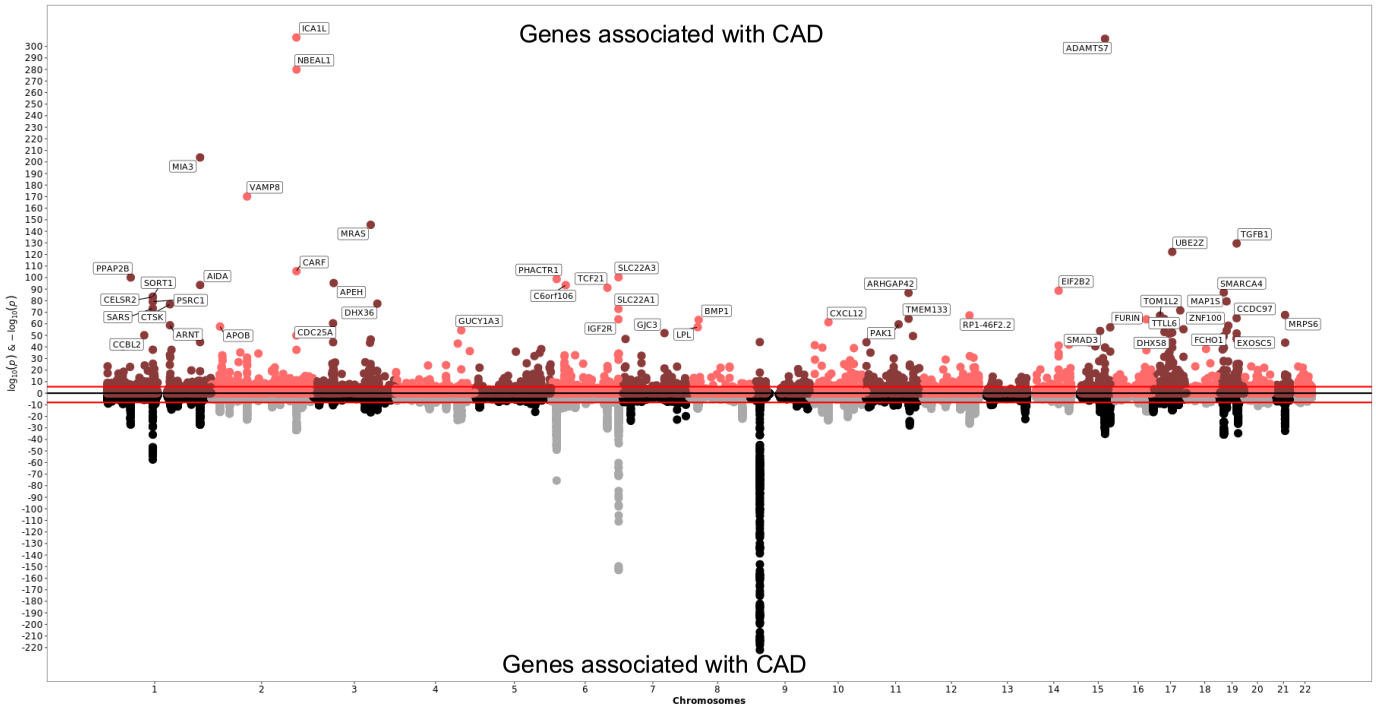


Figure S1

Title: Genes associated with CAD using S-MultiXcan

Legend: A Miami plot of genes and SNPs associated with CAD via S-MultiXcan. Above the x-axis, each data point represents a gene grouped by chromosome (x-axis) and p value (y-axis) of the association with the phenotype. The light red and darl red are alternating colors for genes on odd and even chromosomes. Below the x-axis, each data point represents the SNPs associated with CAD from the input GWAS. Genes with –log(p values) greater that 50 are labelled. The black and grey are alternating colors for SNPs on odd and even chromosomes

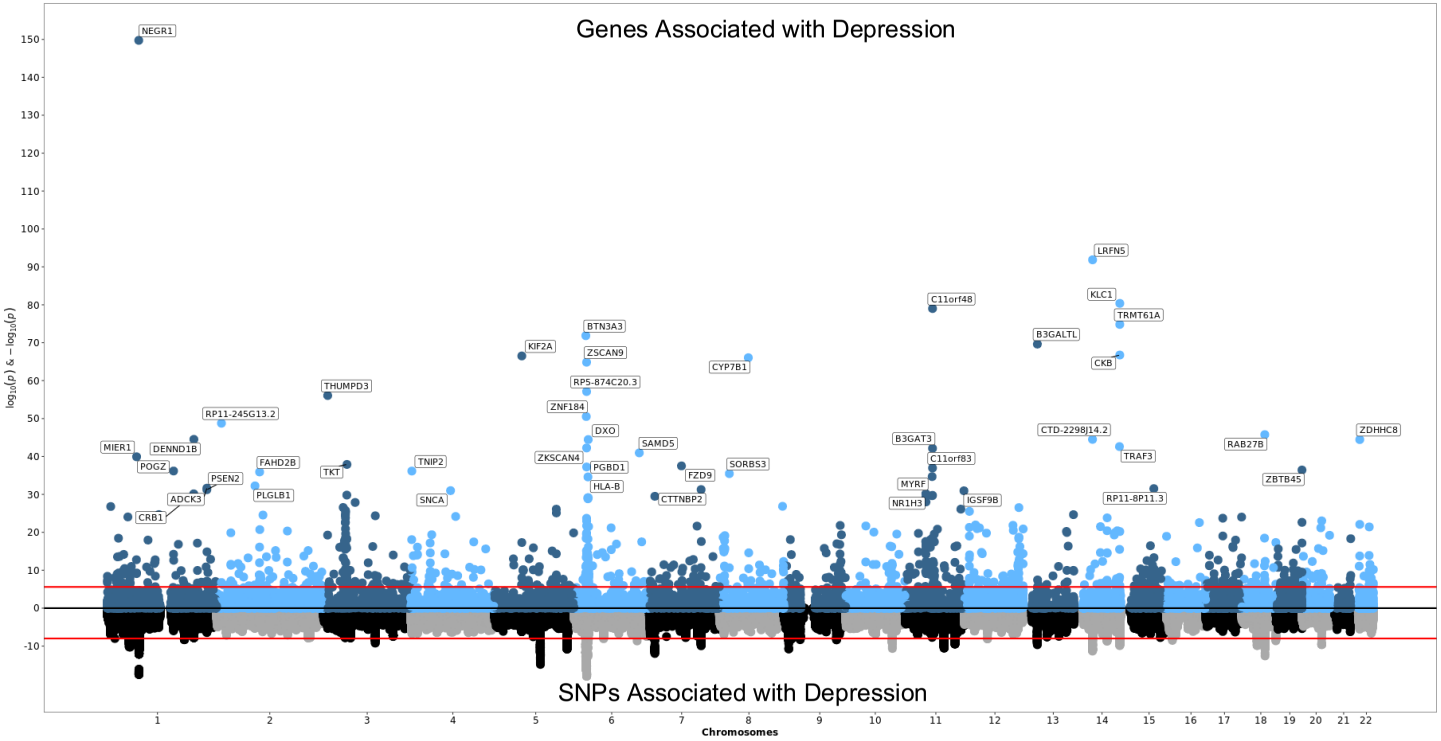


Figure S2

Title: Genes associated with Depression using S-MultiXcan

Legend: A Miami plot of genes and SNPs associated with Depression via S-MultiXcan. Above the x-axis, each data point represents a gene grouped by chromosome (x-axis) and p value (y-axis) of the association with the phenotype. The light red and dark blue are alternating colors for genes on odd and even chromosomes Below the x-axis, each data point represents the SNPs associated with Depression from the input GWAS. Genes with –log(p values) greater that 50 are labelled. The black and grey are alternating colors for SNPs on odd and even chromosomes

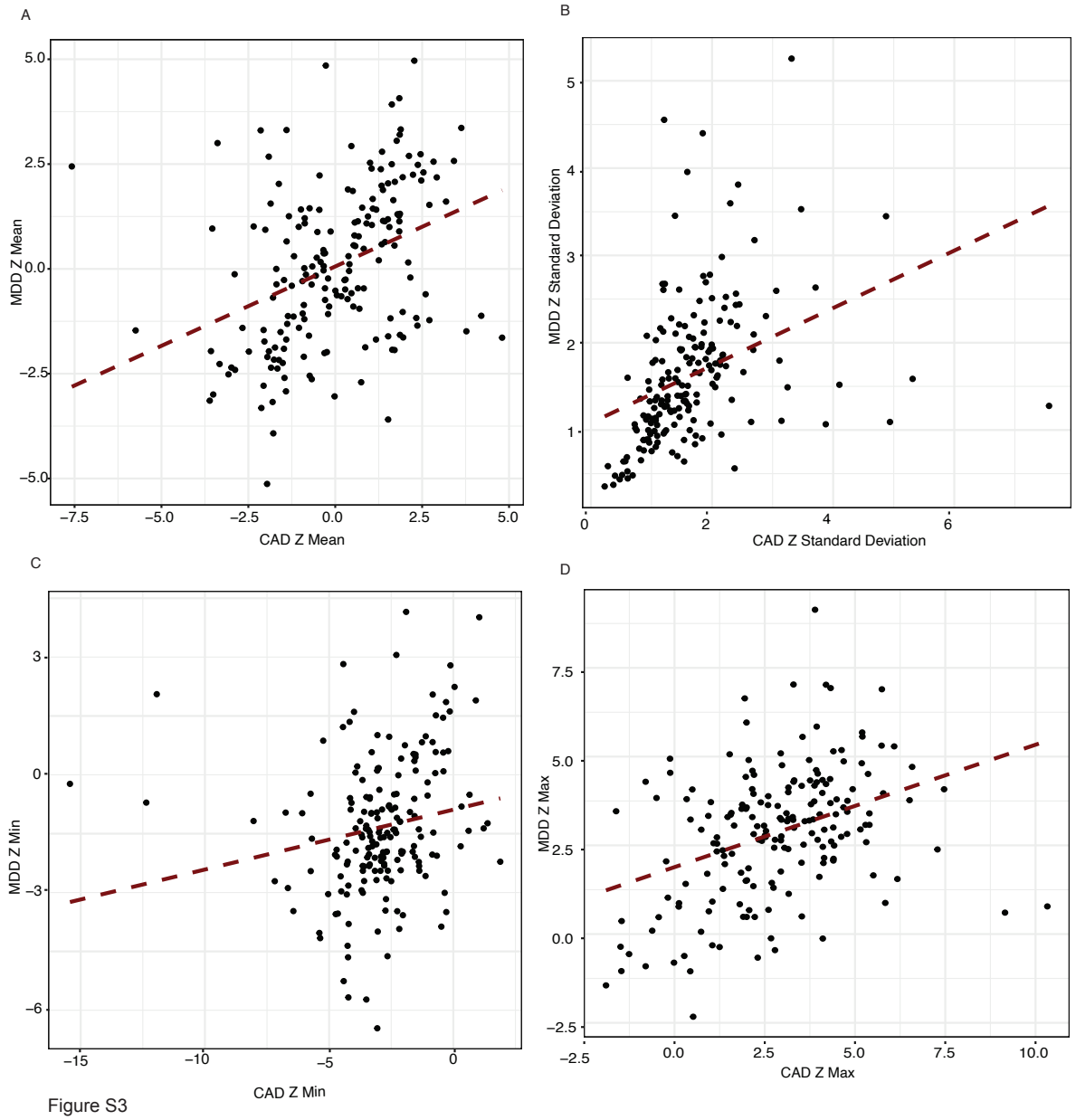
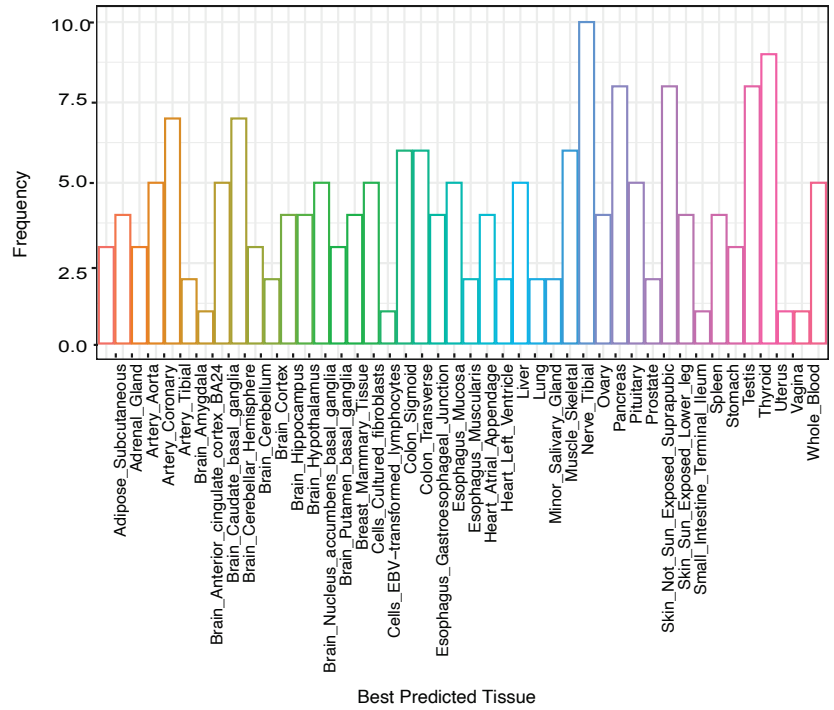


Figure S3

Title: Correlation plots between the test statistics of Depression and CAD for the 185 genes associated with both the phenotypes

Legend: (a) Correlation Plot of Z Mean of Depression and CAD for the 185 genes shared between both the phenotypes (Spearman's coefficient = 0.433)
(b) Correlation Plot of Z Standard Deviation of Depression and CAD for the 185 genes shared between both the phenotypes (Spearman's coefficient = 0.565)
(c) Correlation Plot of Z Min of Depression and CAD for the 185 genes shared between both the phenotypes (Spearman's coefficient = 0.389)
(d) Correlation Plot of Z Max of Depression and CAD for the 185 genes shared between both the phenotypes (Spearman's coefficient = 0.373)



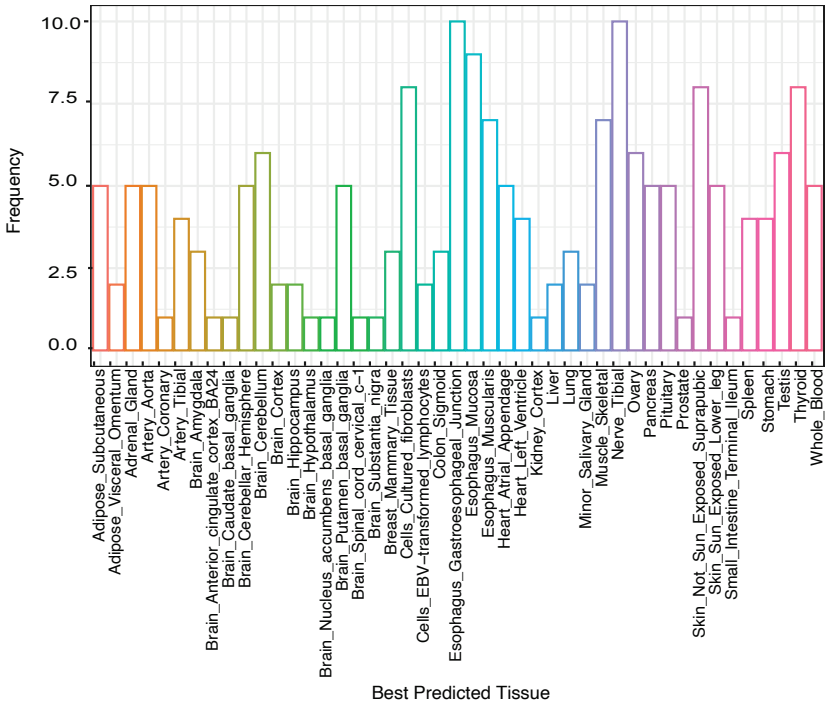


Figure S4

Title: Best Predicted Tissue Distribution of Genes Associated between Depression and CAD

Legend: (a) Tissue distribution of Best Predicted Tissue for genes associated with both Depression and CAD using SMultiXcan Results of CAD
(b) Tissue distribution of Best Predicted Tissue for genes associated with both Depression and CAD using SMultiXcan Results of Depression