Supplementary documents

netCRS: Network-based comorbidity risk score for prediction of myocardial infarction using biobank-scaled PheWAS data

Yonghyun Nam[1,§], Sang-Hyuk Jung[1,2,§], Anurag Verma[3], Vivek Sriram[1], Hong-Hee Won[2], Jae-Seung Yun[1,4], Regeneron Genetics Center, and Dokyoon Kim[1,5,*]

[1]*Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA*
[2]*Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul 06351, Republic of Korea*
[3]*Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4]*Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Republic of Korea*
[5]*Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA*

[§]Y.Nam and S-H.Jung equally contributed for co-first author

*Corresponding author: Dokyoon Kim, PhD
Department of Biostatistics, Epidemiology & Informatics, The Perelman School of Medicine, University of Pennsylvania, B304 Richards Building, 3700 Hamilton Walk, University of Pennsylvania, Philadelphia, PA 19104-6116, United States.
Tel.: +1-215-573-5336; Fax: +1-215-573-3111; E-mail: dokyoon.kim@pennmedicine.upenn.edu

Supplementary Figure 1. Pseudo-code for netCRS of section 2.2

INPUT:
1. **Similarity matrix for DS-Net**
   Disease-layer ($W_{Disease}$), SNP-layer (W_{SNP}), Disease-SNP association (C)
2. **Initialization**
   y_{SNP}: Set of initial label vector for SNP-layer, $f_{Disease}$: Set of resulting candidate comorbidity scores
   N: The number of participants, z: outcome vector for index disease of interest (case: 1 / control: 0)
3. **Hyper-parameter**
   μ: Hyper-parameter for graph-based SSL

PROCEDURE:

Calculating graph Laplacian matrix for disease-layer
$$\mathbf{L}_{Disease} = \boldsymbol{D}_{Disease} - \boldsymbol{W}_{Disease}\text{ , where }\mathbf{D}\text{ is diagonal degree matrix of }\mathbf{W}$$

For i = 1: N,

Initializing labels for SNP-layer: $\boldsymbol{y}_{SNP} = (y_1, \dots, y_k)^{\mathbf{T}}$, where $y_i = \begin{cases} 0 & \text{if genotype CC} \\ 0.5 & \text{if genotype CT} \\ 1 & \text{if genotype TT} \end{cases}$

Applying closed form solution of SSL in Eq. (6): $\boldsymbol{f}_{Disease} = \mu(\boldsymbol{I} + \mu\boldsymbol{L}_{Disease})^{-1}\boldsymbol{C} \cdot \boldsymbol{y}_{SNP}$

End for

Aggregating predicted scores for netCRS( ) – Logistic model: $\boldsymbol{z}$ (MI case vs. Normal control) ~ $\boldsymbol{f}_{Disease}$

OUTPUT:
Individual comorbidity risk scores: netCRS( ) = $\widehat{\boldsymbol{\beta}}\boldsymbol{f}_{Disease}$

Supplementary Table 1. A list of phenotypes in disease-layer

No.	PheCode	Phenotype description	Phenotype category
1	172	Skin cancer	neoplasms
2	172.2	Other non-epithelial cancer of skin	neoplasms
3	174	Breast cancer	neoplasms
4	198	Secondary malignant neoplasm	neoplasms
5	198.1	Secondary malignancy of lymph nodes	neoplasms
6	208	Benign neoplasm of colon	neoplasms
7	214	Lipoma	neoplasms
8	214.1	Lipoma of skin and subcutaneous tissue	neoplasms
9	242	Thyrotoxicosis with or without goiter	endocrine/metabolic
10	244	Hypothyroidism	endocrine/metabolic
11	244.1	Secondary hypothyroidism	endocrine/metabolic
12	244.4	Hypothyroidism NOS	endocrine/metabolic
13	250.1	Type 1 diabetes	endocrine/metabolic
14	250.2	Type 2 diabetes	endocrine/metabolic
15	250.23	Type 2 diabetes with ophthalmic manifestations	endocrine/metabolic
16	250.7	Diabetic retinopathy	endocrine/metabolic
17	272	Disorders of lipoid metabolism	endocrine/metabolic
18	272.1	Hyperlipidemia	endocrine/metabolic
19	272.11	Hypercholesterolemia	endocrine/metabolic
20	275	Disorders of mineral metabolism	endocrine/metabolic

21	278	Overweight, obesity and other hyperalimentation	endocrine/metabolic
22	281	Other deficiency anemia	hematopoietic
23	281.1	Megaloblastic anemia	hematopoietic
24	287	Purpura and other hemorrhagic conditions	hematopoietic
25	287.3	Thrombocytopenia	hematopoietic
26	290	Delirium dementia and amnestic and other cognitive disorders	mental disorders
27	292	Neurological disorders	mental disorders
28	292.4	Altered mental status	mental disorders
29	296	Mood disorders	mental disorders
30	296.2	Depression	mental disorders
31	306	Other mental disorder	mental disorders
32	318	Tobacco use disorder	mental disorders
33	331	Other cerebral degenerations	neurological
34	332	Parkinson's disease	neurological
35	335	Multiple sclerosis	neurological
36	365	Glaucoma	sense organs
37	365.1	Open-angle glaucoma	sense organs
38	365.11	Primary open angle glaucoma	sense organs
39	366	Cataract	sense organs
40	366.2	Senile cataract	sense organs
41	386.9	Dizziness and giddiness (Light-headedness and vertigo)	sense organs
42	394.3	Aortic valve disease	circulatory system
43	395.6	Heart valve replaced	circulatory system
44	401.1	Essential hypertension	circulatory system
45	411.1	Unstable angina (intermediate coronary syndrome)	circulatory system
46	411.2	Myocardial infarction	circulatory system
47	411.3	Angina pectoris	circulatory system
48	411.4	Coronary atherosclerosis	circulatory system
49	411.8	Other chronic ischemic heart disease, unspecified	circulatory system
50	411.9	Other acute and subacute forms of ischemic heart disease	circulatory system
51	415	Pulmonary heart disease	circulatory system
52	418	Nonspecific chest pain	circulatory system
53	418.1	Precordial pain	circulatory system
54	426	Cardiac conduction disorders	circulatory system
55	426.3	Bundle branch block	circulatory system
56	426.9	Cardiac pacemaker__device in situ	circulatory system
57	426.91	Cardiac pacemaker in situ	circulatory system
58	427	Cardiac dysrhythmias	circulatory system
59	427.2	Atrial fibrillation and flutter	circulatory system
60	428	Congestive heart failure; nonhypertensive	circulatory system
61	428.2	Heart failure NOS	circulatory system
62	433	Cerebrovascular disease	circulatory system

63	433.1	Occlusion and stenosis of precerebral arteries	circulatory system
64	433.21	Cerebral artery occlusion, with cerebral infarction	circulatory system
65	433.3	Cerebral ischemia	circulatory system
66	440	Atherosclerosis	circulatory system
67	442	Other aneurysm	circulatory system
68	442.1	Aortic aneurysm	circulatory system
69	443	Peripheral vascular disease	circulatory system
70	443.9	Peripheral vascular disease, unspecified	circulatory system
71	447	Other disorders of arteries and arterioles	circulatory system
72	451	Phlebitis and thrombophlebitis	circulatory system
73	451.2	Phlebitis and thrombophlebitis of lower extremities	circulatory system
74	454	Varicose veins	circulatory system
75	454.1	Varicose veins of lower extremity	circulatory system
76	455	Hemorrhoids	circulatory system
77	459	Other disorders of circulatory system	circulatory system
78	459.9	Circulatory disease NEC	circulatory system
79	475	Chronic sinusitis	respiratory
80	480.11	Pneumococcal pneumonia	respiratory
81	495	Asthma	respiratory
82	496	Chronic airway obstruction	respiratory
83	496.1	Emphysema	respiratory
84	496.2	Chronic bronchitis	respiratory
85	506	Empyema and pneumothorax	respiratory
86	507	Pleurisy; pleural effusion	respiratory
87	509.2	Respiratory insufficiency	respiratory
88	530	Diseases of esophagus	digestive
89	530.1	Esophagitis, GERD and related diseases	digestive
90	535	Gastritis and duodenitis	digestive
91	535.6	Duodenitis	digestive
92	550.2	Diaphragmatic hernia	digestive
93	557	Intestinal malabsorption (non-celiac)	digestive
94	557.1	Celiac disease	digestive
95	562	Diverticulosis and diverticulitis	digestive
96	562.1	Diverticulosis	digestive
97	564	Functional digestive disorders	digestive
98	564.9	Personal history of diseases of digestive system	digestive
99	565	Anal and rectal conditions	digestive
100	565.1	Anal and rectal polyp	digestive
101	569	Other disorders of intestine	digestive
102	571	Chronic liver disease and cirrhosis	digestive
103	571.5	Other chronic nonalcoholic liver disease	digestive
104	574	Cholelithiasis and cholecystitis	digestive

105	574.1	Cholelithiasis	digestive
106	574.11	Cholelithiasis with acute cholecystitis	digestive
107	574.12	Cholelithiasis with other cholecystitis	digestive
108	574.2	Calculus of bile duct	digestive
109	574.3	Cholecystitis without cholelithiasis	digestive
110	575	Other biliary tract disease	digestive
111	575.7	Other disorders of gallbladder	digestive
112	579	Other symptoms involving abdomen and pelvis	digestive
113	586	Other disorders of the kidney and ureters	genitourinary
114	593	Hematuria	genitourinary
115	594	Urinary calculus	genitourinary
116	596	Other disorders of bladder	genitourinary
117	599	Other symptoms__disorders or the urinary system	genitourinary
118	686.1	Carbuncle and furuncle	dermatologic
119	696	Psoriasis and related disorders	dermatologic
120	696.4	Psoriasis	dermatologic
121	696.41	Psoriasis vulgaris	dermatologic
122	704	Diseases of hair and hair follicles	dermatologic
123	706	Diseases of sebaceous glands	dermatologic
124	706.2	Sebaceous cyst	dermatologic
125	714	Rheumatoid arthritis and other inflammatory polyarthropathies	musculoskeletal
126	714.1	Rheumatoid arthritis	musculoskeletal
127	716	Other arthropathies	musculoskeletal
128	716.9	Arthropathy NOS	musculoskeletal
129	728.71	Contracture of palmar fascia [Dupuytren's disease]	musculoskeletal
130	735.2	Acquired toe deformities	musculoskeletal
131	742	Derangement of joint, non-traumatic	musculoskeletal
132	743	Osteoporosis, osteopenia and pathological fracture	musculoskeletal
133	747	Cardiac and circulatory congenital anomalies	congenital anomalies
134	747.1	Cardiac congenital anomalies	congenital anomalies
135	747.13	Congenital anomalies of great vessels	congenital anomalies

Supplementary Table 2. Nagelkerke's R^2 values of PRS in the classification of myocardial infarction cases and controls at the linkage disequilibrium clumping and p-value thresholds

LD-clump r^2 threshold	p-value threshold	Nagelkerke's R^2	p-value	Beta	Standard error	Number of used SNPs
>0.1	5.00E-08	0.0093	8.00E-14	0.5414	0.0725	32
	1.00E-06	0.0085	8.53E-13	0.4108	0.0574	63
	0.0001	0.0111	3.52E-16	0.2882	0.0353	307
	0.001	0.0111	4.30E-16	0.1609	0.0198	1483
	0.01	0.0099	1.52E-14	0.0708	0.0092	9685
	0.05	0.0085	9.78E-13	0.0366	0.0051	37482
>0.2	5.00E-08	0.0071	6.30E-11	0.3831	0.0586	39
	1.00E-06	0.0070	9.55E-11	0.3143	0.0485	73
	0.0001	0.0095	5.07E-14	0.2287	0.0304	348
	0.001	0.0110	5.66E-16	0.1418	0.0175	1638
	0.01	0.0116	9.44E-17	0.0678	0.0082	10863
	0.05	0.0103	4.74E-15	0.0348	0.0044	44067
>0.3	5.00E-08	0.0070	9.82E-11	0.3126	0.0483	45
	1.00E-06	0.0075	2.31E-11	0.2684	0.0401	85
	0.0001	0.0103	4.42E-15	0.2024	0.0258	388
	0.001	0.0117	7.15E-17	0.1274	0.0153	1805
	0.01	0.0130	1.58E-18	0.0640	0.0073	11911
	0.05	0.0111	4.27E-16	0.0316	0.0039	49675
>0.4	5.00E-08	0.0063	7.37E-10	0.2552	0.0414	50
	1.00E-06	0.0065	4.72E-10	0.2072	0.0333	99
	0.0001	0.0093	1.03E-13	0.1647	0.0221	438
	0.001	0.0115	1.15E-16	0.1123	0.0136	1962
	0.01	0.0144	2.51E-20	0.0603	0.0065	12985
	0.05	0.0121	2.38E-17	0.0292	0.0034	55003
>0.5	5.00E-08	0.0061	1.84E-09	0.2139	0.0356	56
	1.00E-06	0.0063	7.36E-10	0.1746	0.0283	113
	0.0001	0.0088	4.52E-13	0.1361	0.0188	492
	0.001	0.0118	5.29E-17	0.1006	0.0120	2133
	0.01	0.0153	1.42E-21	0.0557	0.0058	14160
	0.05	0.0128	2.60E-18	0.0267	0.0031	60663
>0.6	5.00E-08	0.0057	6.01E-09	0.1700	0.0292	67
	1.00E-06	0.0064	6.34E-10	0.1549	0.0251	124
	0.0001	0.0092	1.23E-13	0.1202	0.0162	561
	0.001	0.0121	1.99E-17	0.0888	0.0105	2353
	0.01	0.0162	1.05E-22	0.0506	0.0052	15489
	0.05	0.0134	4.72E-19	0.0240	0.0027	66890
>0.7	5.00E-08	0.0059	3.34E-09	0.1499	0.0253	76
	1.00E-06	0.0062	1.30E-09	0.1265	0.0208	146
	0.0001	0.0089	3.51E-13	0.0990	0.0136	631
	0.001	0.0118	5.97E-17	0.0733	0.0088	2650
	0.01	0.0170	9.14E-24	0.0447	0.0044	17122
	0.05	0.0148	6.25E-21	0.0219	0.0023	74227
>0.8	5.00E-08	0.0049	7.42E-08	0.0839	0.0156	100
	1.00E-06	0.0055	8.75E-09	0.0791	0.0138	188
	0.0001	0.0084	1.37E-12	0.0702	0.0099	757
	0.001	0.0124	8.29E-18	0.0592	0.0069	3042
	0.01	0.0170	8.31E-24	0.0377	0.0057	19321
	0.05	0.0160	2.19E-22	0.0191	0.0020	84126
>0.9	5.00E-08	0.0036	3.72E-06	0.0454	0.0098	135
	1.00E-06	0.0046	1.50E-07	0.0472	0.0090	247
	0.0001	0.0071	6.79E-11	0.0448	0.0069	969
	0.001	0.0112	3.59E-16	0.0406	0.0050	3747
	0.01	**0.0171**	**7.40E-24**	**0.0282**	**0.0028**	**23106**
	0.05	0.0166	3.65E-23	0.0152	0.0015	100493

Supplementary Table 3. Demographics of participants according to netCRS risk groups

netCRS		Low	Intermediate	High	Very high	*P*-value
Number of samples		(n=1243)	(n=1243)	(n=1243)	(n=1243)	
Age		61.7 ± 15.2	61.8 ± 15.2	62.0 ± 14.7	62.6 ± 14.3	0.116
Sex	Female (%)	778 (62.6%)	771 (62.0%)	769 (61.9%)	800 (64.4%)	0.539
Myocardial infarction	Cases (%)	94 (7.6%)	150 (12.1%)	218 (17.5%)	301 (24.20%)	<0.001
	Controls (%)	1149 (92.4%)	1093 (87.9%)	1025 (82.5%)	942 (75.80%)	