Supplemental Data

Using Principal Components of Genetic Variation

for Robust and Powerful Detection of Gene-Gene

Interactions in Case-Control and Case-Only Studies

Samsiddhi Bhattacharjee, Zhaoming Wang, Julia Ciampa, Peter Kraft, Stephen Chanock, Kai Yu, and Nilanjan Chatterjee

TABLE OF CONTENTS

Table S1. List of allele frequencies of causal SNPs and baseline disease risks for the simulation study of type I error and power (Table 1 and Figure 5).

Stratum	Tertile along PC-s				Common MAF	Baseline Disease Risk	
No.	PC1	PC2	PC3	PC4	of Causal SNPs	Correlated**	Uncorrelated*
1	1	1	1	1	0.1500	0.0050	0.0118
2	1	1	1	2	0.1525	0.0054	0.0121
3	1	1	1	3	0.1544	0.0057	0.0124
4	1	1	2	1	0.1562	0.0057	0.0156
5	1	1	2	2	0.1638	0.0061	0.0159
6	1	1	2	3	0.1694	0.0065	0.0104
7	1	1	3	1	0.1626	0.0065	0.0110
8	1	1	3	2	0.1757	0.0070	0.0171
9	1	1	3	3	0.1856	0.0074	0.0126
10	1	2	1	1	0.1630	0.0070	0.0099
11	1	2	1	2	0.1731	0.0075	0.0103
12	1	2	1	3	0.1806	0.0079	0.0103
13	1	2	2	1	0.1831	0.0079	0.0147
14	1	2	2	2	0.2000	0.0085	0.0141
15	1	2	2	3	0.2127	0.0090	0.0158
16	1	2	3	1	0.2051	0.0090	0.0119
17	1	2	3	2	0.2299	0.0097	0.0122
18	1	2	3	3	0.2487	0.0102	0.0138
19	1	3	1	1	0.1687	0.0080	0.0115
20	1	3	1	2	0.1823	0.0086	0.0139
21	1	3	1	3	0.1924	0.0090	0.0116
22	1	3	2	1	0.1953	0.0091	0.0115
23	1	3	2	2	0.2166	0.0098	0.0134
24	1	3	2	3	0.2328	0.0103	0.0097
25	1	3	3	1	0.2249	0.0103	0.0123
26	1	3	3	2	0.2555	0.0111	0.0104
27	1	3	3	3	0.2788	0.0117	0.0120
28	2	1	1	1	0.1584	0.0061	0.0097
29	2	1	1	2	0.1629	0.0066	0.0134
30	2	1	1	3	0.1662	0.0070	0.0153
31	2	1	2	1	0.1713	0.0070	0.0102
32	2	1	2	2	0.1816	0.0075	0.0120
33	2	1	2	3	0.1891	0.0079	0.0084
34	2	1	3	1	0.1851	0.0080	0.0121
35	2	1	3	2	0.2019	0.0086	0.0213
36	2	1	3	3	0.2144	0.0090	0.0114
37	2	2	1	1	0.1816	0.0086	0.0107
38	2	2	1	2	0.1947	0.0092	0.0144
39	2	2	1	3	0.2044	0.0097	0.0109
40	2	2	2	1	0.2110	0.0097	0.0115
41	2	2	2	2	0.2321	0.0105	0.0095
42	2	2	2	3	0.2480	0.0110	0.0074
43	2	2	3	1	0.2438	0.0111	0.0108
44	2	2	3	2	0.2744	0.0119	0.0119
45	2	2	3	3	0.2975	0.0126	0.0130
46	2	3	1	1	0.1919	0.0098	0.0133

47	2	3	1	2	0.2091	0.0105	0.0150
48	2	3	1	3	0.2220	0.0111	0.0087
49	2	3	2	1	0.2293	0.0112	0.0133
50	2	3	2	2	0.2558	0.0120	0.0087
51	2	3	2	3	0.2757	0.0126	0.0104
52	2	3	3	1	0.2715	0.0127	0.0123
53	2	3	3	2	0.3088	0.0137	0.0123
54	2	3	3	3	0.3369	0.0144	0.0162
55	3	1	1	1	0.1748	0.0090	0.0111
56	3	1	1	2	0.1833	0.0096	0.0145
57	3	1	1	3	0.1896	0.0102	0.0098
58	3	1	2	1	0.2021	0.0102	0.0092
59	3	1	2	2	0.2179	0.0110	0.0148
60	3	1	2	3	0.2296	0.0116	0.0179
61	3	1	3	1	0.2324	0.0116	0.0155
62	3	1	3	2	0.2569	0.0125	0.0099
63	3	1	3	3	0.2752	0.0132	0.0084
64	3	2	1	1	0.2196	0.0125	0.0162
65	3	2	1	2	0.2392	0.0134	0.0124
66	3	2	1	3	0.2538	0.0141	0.0146
67	3	2	2	1	0.2697	0.0142	0.0146
68	3	2	2	2	0.2997	0.0152	0.0117
69	3	2	2	3	0.3222	0.0161	0.0139
70	3	2	3	1	0.3264	0.0161	0.0102
71	3	2	3	2	0.3682	0.0173	0.0135
72	3	2	3	3	0.3991	0.0183	0.0173
73	3	3	1	1	0.2405	0.0143	0.0145
74	3	3	1	2	0.2655	0.0153	0.0103
75	3	3	1	3	0.2842	0.0162	0.0123
76	3	3	2	1	0.3015	0.0162	0.0110
77	3	3	2	2	0.3383	0.0175	0.0145
78	3	3	2	3	0.3657	0.0184	0.0173
79	3	3	3	1	0.3705	0.0185	0.0135
80	3	3	3	2	0.4197	0.0199	0.0180
81	3	3	3	3	0.4557	0.0209	0.0116

** Disease risk (α_S) varies along the same direction as common MAF (p_S) of the two causal SNPs (i.e. monotonically along tertiles of each PC). The following two models were used (here i_1, i_2, i_3 and i_4 denote tertiles of the 4 significant PC-s for the Sth stratum).

$$\log(p_S/p_0) = \sum_{j=1}^{4} a_j \psi_j(i_j - 1) + \sum_{j}\sum_{k<j} a_{jk} \psi_j(i_j - 1)\psi_k(i_k - 1) \text{ and}$$

$$\log(\alpha_S/\alpha_0) = \sum_{j=1}^{4} A_j \psi_j(i_j - 1), \text{ where } a_j > 0,\ a_{jk} > 0, A_j > 0,\ \forall j,k$$

with $\psi_1(x) = x\sqrt{x},\ \psi_2(x) = \sqrt{x},\ \psi_3(x) = x,\ \psi_4(x) = \log(1+x)$

* Disease risk varies randomly independent of the principal components,
$\text{logit}(\alpha_S) \sim iid\ N(\mu,\sigma^2)$, for S = 1, …, 81