Appendix

A. Data Simulation

Simulating a dataset involves (i) generating genotypes for each of n individuals; (ii) calculating the risk value for each individual; and (iii) generating event status for each individual. We use the following notation

i index for individual ; n individuals
g index for gene ; G genes
ρ overall proportion of events in the population
f_g frequency of minor allele variant for gene g
K_{gi} number of minor alleles for gene g that subject i has, K_{gi}=0,1 or 2
α_g log odds ratio for gene g

(i) Genotype

For each gene, g, the two alleles are generated independently as binary random variables with probability f_g of being the minor variant E and probability $1\text{-}f_g$ of being the major variant e.

(ii) Risk

For each subject his risk is calculated as $risk_i$=exp(S_i/(1+exp S_i)) where

$$S_i = \alpha_0 + \sum_g^G \alpha_g K_{gi}$$

We chose α_0 so that the overall prevalence is equal to the specified value for ρ. This value for α_0 was calculated numerically.

(iii) Event Status

For each subject his event status was generated as a random binary variable with probability of being 1 equal to $risk_i$.

B. Program

Programs to simulate datasets and calculate summary statistics shown in Tables 2 and 3 use the freely available R statistical package. Within R, install the package `mgrp` available from the Diagnostics and Biomarkers Statistical Center software website (http://labs.fhcrc.org/pepe/dabs/software.html). The key command is `simulateRiskPerf`.

To generate a single row of Table 2 with 50 genes (`nog=50`), common odds ratio=1.05 (`or=1.05`), risk allele frequency of 5% (`f=.05`), setting the population event rate ρ to 10% (`p=0.1`), the number of subjects to 100,000 (`n=1e+05`), and using a high risk threshold of 20% (`highRisk=.2`), the following command was run to yield the corresponding row of Table 2:

```
simulateRiskPerf(nog=50,or=1.05,f=.05,p=0.1,n=1e+05,highRisk=0.2)
```

Options to vary the output and to set the random number seed are also available. Full documentation is provided with the software.

C. Additional Scenarios

The predictive capacity of genetic profiling.These tables vary the high risk threshold and the proportion of subjects in the population who have an event.

(a) risk allele frequency 5%, high risk threshold = 10%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0.387	0.433	0.382	0.005	0.051	0.001	0.533
	150	0.432	0.496	0.425	0.007	0.071	0.003	0.550
	250	0.449	0.536	0.439	0.010	0.097	0.005	0.568
	350	0.457	0.553	0.447	0.011	0.106	0.007	0.574
OR=1.10	50	0.387	0.469	0.378	0.009	0.091	0.004	0.560
	150	0.432	0.562	0.417	0.014	0.145	0.013	0.602
	250	0.449	0.616	0.430	0.019	0.186	0.021	0.629
	350	0.390	0.578	0.369	0.021	0.209	0.029	0.648
OR=1.25	50	0.387	0.573	0.366	0.021	0.206	0.026	0.639
	150	0.333	0.628	0.301	0.033	0.327	0.075	0.723
	250	0.372	0.727	0.332	0.039	0.395	0.116	0.768
	350	0.327	0.730	0.282	0.045	0.448	0.157	0.801
OR=1.50	50	0.387	0.706	0.351	0.035	0.355	0.089	0.738
	150	0.333	0.803	0.282	0.052	0.521	0.217	0.842
	250	0.233	0.773	0.172	0.060	0.601	0.302	0.885
	350	0.268	0.854	0.203	0.065	0.651	0.364	0.908

(b) risk allele frequency 10%, high risk threshold = 10%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0.418	0.472	0.412	0.006	0.060	0.002	0.541
	150	0.452	0.536	0.442	0.009	0.093	0.006	0.568
	250	0.463	0.572	0.451	0.012	0.121	0.010	0.588
	350	0.420	0.545	0.406	0.014	0.139	0.013	0.602
OR=1.10	50	0.418	0.524	0.406	0.012	0.117	0.008	0.582
	150	0.376	0.547	0.357	0.019	0.190	0.024	0.634
	250	0.405	0.620	0.381	0.024	0.239	0.039	0.667
	350	0.372	0.620	0.344	0.028	0.276	0.054	0.694
OR=1.25	50	0.418	0.657	0.391	0.027	0.266	0.047	0.684
	150	0.307	0.687	0.265	0.042	0.422	0.130	0.781
	250	0.297	0.748	0.247	0.050	0.502	0.197	0.829
	350	0.282	0.783	0.227	0.056	0.557	0.253	0.862
OR=1.50	50	0.299	0.702	0.253	0.045	0.449	0.150	0.796
	150	0.245	0.817	0.182	0.064	0.635	0.328	0.897
	250	0.249	0.880	0.179	0.070	0.701	0.420	0.929
	350	0.206	0.884	0.131	0.075	0.753	0.490	0.948

(c) risk allele frequency 30%, high risk threshold = 10%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0.453	0.534	0.443	0.009	0.090	0.005	0.563
	150	0.418	0.557	0.403	0.015	0.154	0.014	0.607
	250	0.403	0.585	0.382	0.020	0.202	0.024	0.639
	350	0.418	0.618	0.395	0.022	0.223	0.032	0.655
OR=1.10	50	0.453	0.608	0.435	0.017	0.173	0.018	0.620
	150	0.371	0.634	0.342	0.029	0.292	0.055	0.700
	250	0.366	0.686	0.329	0.036	0.357	0.089	0.744
	350	0.324	0.693	0.283	0.041	0.411	0.121	0.777
OR=1.25	50	0.367	0.707	0.328	0.038	0.378	0.100	0.756
	150	0.282	0.788	0.227	0.056	0.561	0.246	0.862
	250	0.261	0.835	0.197	0.064	0.639	0.340	0.904
	350	0.241	0.863	0.171	0.069	0.692	0.407	0.926
OR=1.50	50	0.289	0.817	0.230	0.059	0.587	0.268	0.874
	150	0.205	0.879	0.131	0.075	0.748	0.480	0.947
	250	0.202	0.929	0.122	0.081	0.808	0.576	0.967
	350	0.171	0.923	0.087	0.084	0.836	0.631	0.976

(d) risk allele frequency 5%, high risk threshold = 30%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0.001	0.533
	150	0	0	0	0	0	0.003	0.550
	250	0	0	0	0	0	0.005	0.568
	350	0	0	0	0	0	0.007	0.574
OR=1.10	50	0	0	0	0	0	0.004	0.560
	150	0	0.001	0	0	0.001	0.013	0.602
	250	0.001	0.005	0.001	0	0.004	0.021	0.629
	350	0.005	0.019	0.004	0.001	0.016	0.029	0.648
OR=1.25	50	0.004	0.016	0.003	0	0.013	0.026	0.639
	150	0.029	0.116	0.019	0.004	0.097	0.075	0.723
	250	0.045	0.200	0.028	0.009	0.172	0.116	0.768
	350	0.074	0.317	0.046	0.014	0.271	0.157	0.801
OR=1.50	50	0.028	0.125	0.017	0.006	0.107	0.089	0.738
	150	0.078	0.388	0.044	0.022	0.344	0.217	0.842
	250	0.093	0.500	0.048	0.032	0.452	0.302	0.885
	350	0.100	0.567	0.048	0.038	0.519	0.364	0.908

(e) risk allele frequency 10%, high risk threshold = 30%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0.002	0.541
	150	0	0	0	0	0	0.006	0.568
	250	0	0	0	0	0	0.010	0.588
	350	0	0.001	0	0	0.001	0.013	0.602
OR=1.10	50	0	0	0	0	0	0.008	0.582
	150	0.002	0.008	0.002	0	0.006	0.024	0.634
	250	0.013	0.044	0.009	0.001	0.035	0.039	0.667
	350	0.021	0.075	0.015	0.002	0.060	0.054	0.694
OR=1.25	50	0.020	0.073	0.014	0.002	0.059	0.047	0.684
	150	0.054	0.235	0.035	0.010	0.201	0.130	0.781
	250	0.080	0.373	0.048	0.019	0.325	0.197	0.829
	350	0.096	0.472	0.054	0.026	0.418	0.253	0.862
OR=1.50	50	0.074	0.311	0.047	0.013	0.264	0.150	0.796
	150	0.107	0.569	0.056	0.035	0.513	0.328	0.897
	250	0.104	0.626	0.046	0.045	0.581	0.420	0.929
	350	0.118	0.727	0.051	0.053	0.676	0.490	0.948

(f) risk allele frequency 30%, high risk threshold = 30%, ρ = 10%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0.005	0.563
	150	0	0.001	0	0	0	0.014	0.607
	250	0.002	0.009	0.002	0	0.007	0.024	0.639
	350	0.007	0.021	0.005	0	0.016	0.032	0.655
OR=1.10	50	0.001	0.004	0.001	0	0.003	0.018	0.620
	150	0.020	0.073	0.014	0.002	0.059	0.055	0.700
	250	0.045	0.171	0.030	0.005	0.141	0.089	0.744
	350	0.054	0.225	0.035	0.009	0.190	0.121	0.777
OR=1.25	50	0.053	0.203	0.036	0.006	0.167	0.100	0.756
	150	0.094	0.459	0.054	0.025	0.405	0.246	0.862
	250	0.111	0.585	0.057	0.036	0.528	0.340	0.904
	350	0.116	0.651	0.056	0.044	0.596	0.407	0.926
OR=1.50	50	0.080	0.431	0.040	0.027	0.390	0.268	0.874
	150	0.116	0.714	0.050	0.052	0.664	0.480	0.947
	250	0.111	0.766	0.038	0.062	0.728	0.576	0.967
	350	0.116	0.818	0.038	0.067	0.780	0.631	0.976

(g) risk allele frequency 5%, high risk threshold = 20%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.519
	150	0	0	0	0	0	0	0.558
	250	0	0	0	0	0	0.001	0.572
	350	0	0	0	0	0	0.001	0.566
OR=1.10	50	0	0	0	0	0	0	0.551
	150	0	0	0	0	0	0.002	0.612
	250	0	0	0	0	0	0.003	0.637
	350	0	0	0	0	0	0.003	0.640
OR=1.25	50	0	0	0	0	0	0.003	0.631
	150	0	0.001	0	0	0.001	0.013	0.734
	250	0	0.013	0	0	0.013	0.027	0.790
	350	0.001	0.031	0.001	0	0.030	0.038	0.818
OR=1.50	50	0	0.009	0	0	0.009	0.016	0.733
	150	0.005	0.158	0.004	0.001	0.154	0.077	0.870
	250	0.008	0.246	0.005	0.001	0.240	0.133	0.921
	350	0.008	0.324	0.005	0.002	0.319	0.192	0.944

(h) risk allele frequency 10%, high risk threshold = 20%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.534
	150	0	0	0	0	0	0.001	0.563
	250	0	0	0	0	0	0.001	0.593
	350	0	0	0	0	0	0.001	0.596
OR=1.10	50	0	0	0	0	0	0.001	0.575
	150	0	0	0	0	0	0.003	0.635
	250	0	0	0	0	0	0.006	0.682
	350	0	0	0	0	0	0.008	0.696
OR=1.25	50	0	0	0	0	0	0.006	0.680
	150	0.001	0.022	0	0	0.022	0.029	0.800
	250	0.003	0.089	0.002	0	0.087	0.062	0.859
	350	0.004	0.147	0.003	0.001	0.144	0.090	0.896
OR=1.50	50	0.002	0.053	0.002	0	0.051	0.037	0.805
	150	0.007	0.259	0.004	0.002	0.255	0.153	0.927
	250	0.013	0.479	0.008	0.003	0.471	0.255	0.962
	350	0.012	0.534	0.007	0.004	0.527	0.321	0.974

(i) risk allele frequency 30%, high risk threshold = 20%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.549
	150	0	0	0	0	0	0.002	0.610
	250	0	0	0	0	0	0.003	0.641
	350	0	0	0	0	0	0.004	0.658
OR=1.10	50	0	0	0	0	0	0.002	0.617
	150	0	0	0	0	0	0.007	0.702
	250	0	0.006	0	0	0.006	0.017	0.761
	350	0	0.015	0	0	0.014	0.023	0.782
OR=1.25	50	0	0.010	0	0	0.010	0.018	0.759
	150	0.004	0.157	0.003	0.001	0.155	0.094	0.888
	250	0.009	0.317	0.006	0.002	0.311	0.165	0.936
	350	0.010	0.408	0.007	0.002	0.401	0.232	0.961
OR=1.50	50	0.007	0.214	0.005	0.001	0.209	0.106	0.900
	150	0.011	0.541	0.006	0.004	0.535	0.331	0.976
	250	0.015	0.668	0.008	0.005	0.660	0.419	0.986
	350	0.015	0.775	0.008	0.006	0.767	0.515	0.992

(j) risk allele frequency 5%, high risk threshold = 10%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.519
	150	0	0	0	0	0	0	0.558
	250	0	0	0	0	0	0.001	0.572
	350	0	0	0	0	0	0.001	0.566
OR=1.10	50	0	0	0	0	0	0	0.551
	150	0	0	0	0	0	0.002	0.612
	250	0	0	0	0	0	0.003	0.637
	350	0	0	0	0	0	0.003	0.640
OR=1.25	50	0	0	0	0	0	0.003	0.631
	150	0.001	0.018	0.001	0	0.017	0.013	0.734
	250	0.005	0.070	0.004	0	0.067	0.027	0.790
	350	0.008	0.129	0.007	0.001	0.122	0.038	0.818
OR=1.50	50	0.004	0.053	0.004	0	0.049	0.016	0.733
	150	0.017	0.283	0.014	0.001	0.269	0.077	0.870
	250	0.020	0.423	0.016	0.002	0.407	0.133	0.921
	350	0.018	0.493	0.014	0.003	0.479	0.192	0.944

(k) risk allele frequency 10%, high risk threshold = 10%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.534
	150	0	0	0	0	0	0.001	0.563
	250	0	0	0	0	0	0.001	0.593
	350	0	0	0	0	0	0.001	0.596
OR=1.10	50	0	0	0	0	0	0.001	0.575
	150	0	0	0	0	0	0.003	0.635
	250	0	0	0	0	0	0.006	0.682
	350	0	0.003	0	0	0.003	0.008	0.696
OR=1.25	50	0	0.007	0	0	0.007	0.006	0.680
	150	0.007	0.101	0.006	0	0.095	0.029	0.800
	250	0.013	0.225	0.010	0.001	0.215	0.062	0.859
	350	0.016	0.293	0.013	0.002	0.280	0.090	0.896
OR=1.50	50	0.010	0.138	0.009	0	0.129	0.037	0.805
	150	0.017	0.433	0.013	0.003	0.420	0.153	0.927
	250	0.025	0.626	0.019	0.004	0.607	0.255	0.962
	350	0.021	0.664	0.015	0.005	0.649	0.321	0.974

(l) risk allele frequency 30%, high risk threshold = 10%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.549
	150	0	0	0	0	0	0.002	0.610
	250	0	0	0	0	0	0.003	0.641
	350	0	0	0	0	0	0.004	0.658
OR=1.10	50	0	0	0	0	0	0.002	0.617
	150	0	0.002	0	0	0.002	0.007	0.702
	250	0.002	0.024	0.001	0	0.023	0.017	0.761
	350	0.006	0.075	0.005	0	0.070	0.023	0.782
OR=1.25	50	0.002	0.034	0.002	0	0.032	0.018	0.759
	150	0.015	0.309	0.012	0.002	0.297	0.094	0.888
	250	0.019	0.445	0.015	0.003	0.431	0.165	0.936
	350	0.022	0.571	0.017	0.004	0.554	0.232	0.961
OR=1.50	50	0.021	0.370	0.017	0.002	0.353	0.106	0.900
	150	0.020	0.678	0.013	0.005	0.665	0.331	0.976
	250	0.024	0.773	0.016	0.006	0.757	0.419	0.986
	350	0.022	0.851	0.014	0.007	0.837	0.515	0.992

(m) risk allele frequency 5%, high risk threshold = 30%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.519
	150	0	0	0	0	0	0	0.558
	250	0	0	0	0	0	0.001	0.572
	350	0	0	0	0	0	0.001	0.566
OR=1.10	50	0	0	0	0	0	0	0.551
	150	0	0	0	0	0	0.002	0.612
	250	0	0	0	0	0	0.003	0.637
	350	0	0	0	0	0	0.003	0.640
OR=1.25	50	0	0	0	0	0	0.003	0.631
	150	0	0	0	0	0	0.013	0.734
	250	0	0.010	0	0	0.010	0.027	0.790
	350	0	0.012	0	0	0.012	0.038	0.818
OR=1.50	50	0	0.002	0	0	0.002	0.016	0.733
	150	0.001	0.061	0.001	0	0.060	0.077	0.870
	250	0.005	0.179	0.003	0.001	0.176	0.133	0.921
	350	0.005	0.262	0.003	0.002	0.260	0.192	0.944

(n) risk allele frequency 10%, high risk threshold = 30%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.534
	150	0	0	0	0	0	0.001	0.563
	250	0	0	0	0	0	0.001	0.593
	350	0	0	0	0	0	0.001	0.596
OR=1.10	50	0	0	0	0	0	0.001	0.575
	150	0	0	0	0	0	0.003	0.635
	250	0	0	0	0	0	0.006	0.682
	350	0	0	0	0	0	0.008	0.696
OR=1.25	50	0	0	0	0	0	0.006	0.680
	150	0	0.008	0	0	0.008	0.029	0.800
	250	0.001	0.048	0.001	0	0.047	0.062	0.859
	350	0.002	0.080	0.001	0	0.078	0.090	0.896
OR=1.50	50	0	0.015	0	0	0.015	0.037	0.805
	150	0.004	0.179	0.002	0.001	0.176	0.153	0.927
	250	0.006	0.323	0.003	0.002	0.320	0.255	0.962
	350	0.008	0.462	0.004	0.003	0.458	0.321	0.974

(o) risk allele frequency 30%, high risk threshold = 30%, ρ = 1%

	Number of genes	Total at high risk	Cases at high risk (TPR)	Controls at high risk (FPR)	Expected Net Benefit	Net reclassification index	R^2	AUC
OR=1.05	50	0	0	0	0	0	0	0.549
	150	0	0	0	0	0	0.002	0.610
	250	0	0	0	0	0	0.003	0.641
	350	0	0	0	0	0	0.004	0.658
OR=1.10	50	0	0	0	0	0	0.002	0.617
	150	0	0	0	0	0	0.007	0.702
	250	0	0.003	0	0	0.003	0.017	0.761
	350	0	0.004	0	0	0.004	0.023	0.782
OR=1.25	50	0	0.002	0	0	0.002	0.018	0.759
	150	0.002	0.090	0.001	0	0.088	0.094	0.888
	250	0.004	0.205	0.002	0.001	0.203	0.165	0.936
	350	0.007	0.331	0.004	0.002	0.327	0.232	0.961
OR=1.50	50	0.002	0.106	0.001	0.001	0.105	0.106	0.900
	150	0.008	0.451	0.004	0.003	0.447	0.331	0.976
	250	0.009	0.549	0.004	0.004	0.545	0.419	0.986
	350	0.010	0.655	0.004	0.005	0.652	0.515	0.992