Sample Size, Power and Effect Size

Below we provide estimates of power and effect size for various genetic tests and subgroups within the Framingham Heart Study. Estimates are provided for the following situations:

1) linkage analyses with the 330 pedigrees included in the genome scan with microsatellite markers, conducted by the Mammalian Genotyping Service,
2) association analyses for unrelated individuals with varying sample sizes
3) association analyses for family-based association tests

Linkage Analyses in 330 Pedigrees

All power calculations for linkage were performed using SOLAR v1.6.6 (Almasy and Blangero, 1998). We evaluated the power of a likelihood-ratio test for linkage for a range of QTL heritabilities (5% to 40%) and various sample sizes. The different sample sizes will reflect the variation in the availability of data for different phenotypes. Three sample sizes, ranging from the largest possible size (e.g. for a trait measured in all Framingham participants such as blood pressure) to smaller sizes (e.g. traits measured only in the Offspring Cohort or a subset of Offspring) were considered within the existing 330 pedigrees. For these samples we computed power for three critical values (LOD scores 1.0, 2.0 and 3.0). These critical values encompass LOD scores that are suggestive of linkage (LOD $\geq$ 1.0 or 1.5) or establish significant linkage (LOD $\geq$ 3). Given a selected critical value, an approximate value of the power of the likelihood-ratio test is computed using the non-central χ^2 distribution whose non-centrality parameter is twice the difference between the expected log-likelihoods of the null and alternative hypotheses of linkage (Sham et al., 2000). In this analysis the non-centrality parameters were estimated by means of simulation using our pedigrees.

For each QTL heritability, 100 data sets were simulated using the 330 pedigree structure and a genetic model containing a marker with 10 alleles of equal frequency (0.90 heterozygosity) that was linked to a diallelic QTL with 10% allelic frequency. We evaluated other allele frequencies, but found that all were essentially equivalent since the primary determinant of power here is the proportion of variance that the QTL explains. Simulation was performed using the "simqtl" command in SOLAR. First, the marker and the phenotypic data (under the assumed QTL variance) were generated and the IBD probabilities were computed for all observations in the pedigrees. For a given sample the non-centrality parameter of the χ^2 distribution was estimated from a twopoint linkage analysis of that sample.

Tables 1 presents the expected lod score (ELOD, the average twopoint LOD score from 100 simulated data sets) and power computed for the various samples and levels of QTL heritability in the 330 families. In these tables the three columns for power represent different critical values. For example, the estimated power for sample size N=2885 in 330 pedigrees and 20% QTL heritability was 99%, 92% and 76% if linkage is deemed significant for LOD $\geq$ 1.0, LOD $\geq$ 2.0 and LOD $\geq$ 3.0, respectively.

Table 1: Power for Linkage Analyses in 330 Pedigrees

QTL Heritability (%)	Original Cohort and Offspring [N=2885, Ped=330] ELOD	LOD≥1	LOD≥2	LOD≥3	Offspring [N=1672, Ped=330] ELOD	LOD≥1	LOD≥2	LOD≥3	Offspring Subset [N=1228, Ped=326] ELOD	LOD≥1	LOD≥2	LOD≥3
5	0.39	21	4	1	0.30	16	3	1	0.21	12	2	0
10	1.20	58	25	9	0.83	43	14	4	0.51	27	7	1
15	2.47	89	63	36	1.68	74	40	17	0.98	49	18	6
20	4.22	99	92	76	2.84	93	72	46	1.63	72	38	16
25	6.50	100	99	96	4.36	99	93	78	2.47	89	63	37
30	9.35	100	100	100	6.26	100	99	95	3.53	97	84	62
35	12.87	100	100	100	8.60	100	100	99	4.84	99	95	84

***For 40% QTL heritability, power greater than 95%**

Association Studies with Biologically Unrelated Subjects

Table 2 displays detectable effect sizes (ΔR^2 or percent variation explained by a QTL SNP) for a quantitative trait measured on unrelated subjects for power of 80% or 90% and alpha levels of 0.01 0.001 and 0.0001. These calculations are performed for varying samples sizes, from 1800 to 800. The numbers for 1800 correspond to genotyping the unrelated plate set and all subjects having the trait. Thus, for 1% significance level, if background (baseline) covariates explain 30% of the variation in the trait (R^2 base), then a sample of size 1800 will have 80% power to detect a QTL variance of 0.0054 and 90% power to detect a QTL variance of 0.0068.

Table 2: QTL variance sufficient for power 0.80 (0.90) in association analyses of unrelated subjects
Analysis with General Linear Model having unrestricted genotype means [u(AA), u(Aa), u(aa)]
Entries are values of QTL variance for stated power and significance level
given fixed sample size, and non-QTL baseline R-squared, with total variance=1.0

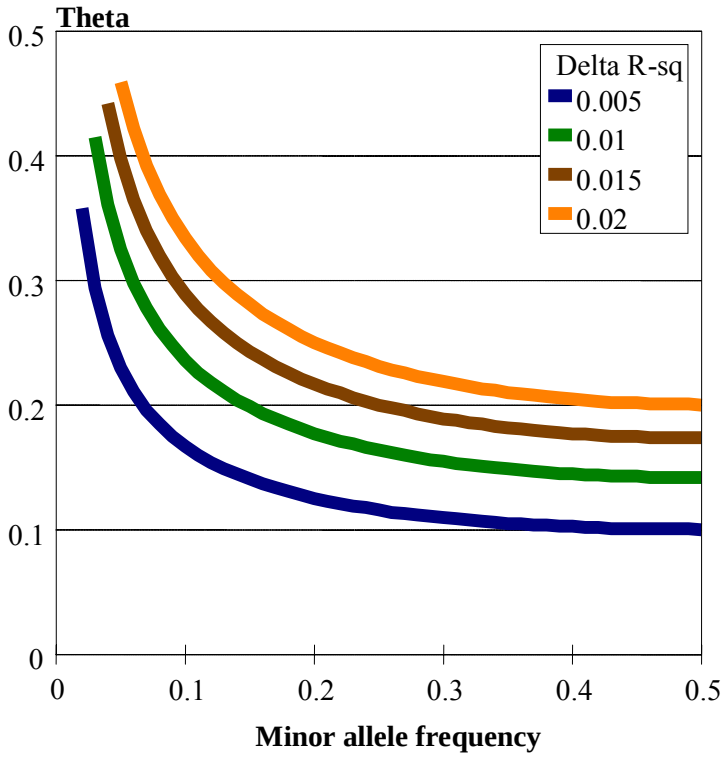
Sample			Significance Level		
Size	Power	R^2 base	0.01	0.001	0.0001
1800	0.8	0.5	0.0039	0.0055	0.0070
1800	0.8	0.3	0.0054	0.0076	0.0098
1800	0.8	0.1	0.0069	0.0098	0.0126
1800	0.9	0.5	0.0048	0.0066	0.0083
1800	0.9	0.3	0.0068	0.0092	0.0116
1800	0.9	0.1	0.0087	0.0119	0.0149
1700	0.8	0.5	0.0041	0.0058	0.0074
1700	0.8	0.3	0.0057	0.0081	0.0104
1700	0.8	0.1	0.0074	0.0104	0.0133
1700	0.9	0.5	0.0051	0.0070	0.0087
1700	0.9	0.3	0.0072	0.0098	0.0122

1700	0.9	0.1	0.0092	0.0126	0.0157
1600	0.8	0.5	0.0043	0.0061	0.0079
1600	0.8	0.3	0.0061	0.0086	0.0110
1600	0.8	0.1	0.0078	0.0110	0.0142
1600	0.9	0.5	0.0054	0.0074	0.0093
1600	0.9	0.3	0.0076	0.0104	0.0130
1600	0.9	0.1	0.0098	0.0133	0.0167
1500	0.8	0.5	0.0046	0.0065	0.0084
1500	0.8	0.3	0.0065	0.0092	0.0117
1500	0.8	0.1	0.0083	0.0118	0.0151
1500	0.9	0.5	0.0058	0.0079	0.0099
1500	0.9	0.3	0.0081	0.0111	0.0139
1500	0.9	0.1	0.0104	0.0142	0.0178
1400	0.8	0.5	0.0050	0.0070	0.0090
1400	0.8	0.3	0.0069	0.0098	0.0126
1400	0.8	0.1	0.0089	0.0126	0.0162
1400	0.9	0.5	0.0062	0.0085	0.0106
1400	0.9	0.3	0.0087	0.0119	0.0148
1400	0.9	0.1	0.0112	0.0152	0.0191
1300	0.8	0.5	0.0053	0.0075	0.0097
1300	0.8	0.3	0.0075	0.0106	0.0135
1300	0.8	0.1	0.0096	0.0136	0.0174
1300	0.9	0.5	0.0067	0.0091	0.0114
1300	0.9	0.3	0.0094	0.0128	0.0160
1300	0.9	0.1	0.0120	0.0164	0.0205
1200	0.8	0.5	0.0058	0.0082	0.0105
1200	0.8	0.3	0.0081	0.0114	0.0147
1200	0.8	0.1	0.0104	0.0147	0.0188
1200	0.9	0.5	0.0072	0.0099	0.0124
1200	0.9	0.3	0.0101	0.0138	0.0173
1200	0.9	0.1	0.0130	0.0178	0.0222
1100	0.8	0.5	0.0063	0.0089	0.0114
1100	0.8	0.3	0.0088	0.0125	0.0160
1100	0.8	0.1	0.0114	0.0160	0.0206
1100	0.9	0.5	0.0079	0.0108	0.0135
1100	0.9	0.3	0.0111	0.0151	0.0189
1100	0.9	0.1	0.0142	0.0194	0.0242
1000	0.8	0.5	0.0069	0.0098	0.0126
1000	0.8	0.3	0.0097	0.0137	0.0176
1000	0.8	0.1	0.0125	0.0177	0.0226
1000	0.9	0.5	0.0087	0.0118	0.0148
1000	0.9	0.3	0.0122	0.0166	0.0207
1000	0.9	0.1	0.0156	0.0213	0.0266
900	0.8	0.5	0.0077	0.0109	0.0139
900	0.8	0.3	0.0108	0.0152	0.0195
900	0.8	0.1	0.0139	0.0196	0.0251
900	0.9	0.5	0.0097	0.0131	0.0164
900	0.9	0.3	0.0135	0.0184	0.0230

900	0.9	0.1	0.0174	0.0236	0.0296
800	0.8	0.5	0.0087	0.0122	0.0157
800	0.8	0.3	0.0122	0.0171	0.0219
800	0.8	0.1	0.0156	0.0220	0.0282
800	0.9	0.5	0.0109	0.0148	0.0185
800	0.9	0.3	0.0152	0.0207	0.0258
800	0.9	0.1	0.0195	0.0266	0.0332

Differences between SNP genotypes relate to ΔR^2 through allele frequencies (let f denote minor allele frequency) — which determine expected genotype frequencies (g_1, g_2 and g_3) — and through covariate-adjusted genotype-specific means (μ_1, μ_2 and μ_3). Without loss of generality, we assume trait values are standardized to variance 1. Then $\Delta R^2 = \Sigma\, g_i(\mu_i - \mu)^2$ where $\mu = \Sigma\, g_i\mu_i$ is the overall mean. With an additive genetic model, $\mu_2 - \mu_1 = \theta$ and $\mu_3 - \mu_1 = 2\theta$. Figure 1 displays the relation between the increment ΔR^2, the minor allele frequency, f, and the additive difference between means, θ. For example, suppose we determine that we have power 0.80 to detect a SNP that contributes 2% of the variance (Δ=0.02). If the minor allele frequency is f=0.10 then $\theta \approx 0.33$; note that θ is in standard deviation units. If the minor allele frequency is f=0.30, then $\theta \approx 0.22$. In conjunction with the power table, this Figure helps us to determine which combinations of allele frequencies and genotype mean differences will be compatible with the value of ΔR^2 from power calculations for specified power and significance level.

Figure 1. Delta R-square related to Allele Frequency
Theta = difference between genotype means



Association Studies using Family-Based Association Test

Table 3 displays power calculated using PBAT for the family-based association test with the family plate set. Two sample sizes were evaluated: n=1398 from 400 nuclear families using essentially all subjects on the plate set and n=1061 from 291 nuclear families, a subset of the total. Power was computed for a range of QTL and marker minor allele frequencies. The QTL and the marker are assumed to be in linkage disequilibrium (D'~1). Thus, when the minor allele for the QTL is equal to that of the marker, it is assumed that the marker and the QTL are the same and power is at its maximum. Although the D' remains constant, as the differences between the minor alleles for the QTL and the marker increase, the LD correlation (R^2) and the study power decrease.

For example, if a study used 1400 subjects on the family plate set for a trait with allele frequency for the QTL of 5% and allele frequency of the nearby marker of 10%, the power to detect a QTL variance (heritability) of 10% would be 52% for an alpha of 0.001 and 93% for an alpha of 0.05. When the allele frequencies of the marker and the underlying functional variant of the QTL are the same, power in these examples increase to a maximum of 98% and 100%.

Table 3: Power for FBAT analyses

Sample Size	Allele Freq for Functional Variant	Allele Freq for Typed Marker	QTL H^2	Alpha	Power
1400	5	5	5	0.001	66
			5	0.05	98
			10	0.001	98
			10	0.05	100
1400	5	10	5	0.001	21
			5	0.05	72
			10	0.001	52
			10	0.05	93
1400	5	20	5	0.001	6
			5	0.05	41
			10	0.001	16
			10	0.05	65
1400	5	30	5	0.001	2
			5	0.05	26
			10	0.001	7
			10	0.05	44
1400	10	5	5	0.001	24
			5	0.05	77

			10	0.001	65
			10	0.05	97
1400	10	10	5	0.001	71
			5	0.05	98
			10	0.001	99
			10	0.05	100
1400	10	20	5	0.001	22
			5	0.05	72
			10	0.001	57
			10	0.05	94
1400	10	30	5	0.001	9
			5	0.05	50
			10	0.001	27
			10	0.05	77
1400	20	5	5	0.001	6
			5	0.05	43
			10	0.001	21
			10	0.05	73
1400	20	10	5	0.001	24
			5	0.05	75
			10	0.001	66
			10	0.05	97
1400	20	20	5	0.001	74
			5	0.05	98
			10	0.001	99
			10	0.05	100
1400	20	30	5	0.001	36
			5	0.05	85
			10	0.001	80
			10	0.05	99
1400	30	5	5	0.001	3
			5	0.05	28
			10	0.001	9
			10	0.05	51
1400	30	10	5	0.001	10
			5	0.05	53
			10	0.001	33
			10	0.05	84

1400	30	20	5	0.001	39
			5	0.05	87
			10	0.001	86
			10	0.05	99
1400	30	30	5	0.001	75
			5	0.05	98
			10	0.001	99
			10	0.05	100
1061	5	5	5	0.001	29
			5	0.05	85
			10	0.001	72
			10	0.05	99
1061	5	10	5	0.001	8
			5	0.05	50
			10	0.001	22
			10	0.05	76
1061	5	20	5	0.001	2
			5	0.05	27
			10	0.001	7
			10	0.05	45
1061	5	30	5	0.001	1
			5	0.05	18
			10	0.001	3
			10	0.05	30
1061	10	5	5	0.001	9
			5	0.05	54
			10	0.001	28
			10	0.05	84
1061	10	10	5	0.001	36
			5	0.05	87
			10	0.001	82
			10	0.05	100
1061	10	20	5	0.001	9
			5	0.05	51
			10	0.001	27
			10	0.05	78
1061	10	30	5	0.001	4
			5	0.05	33
			10	0.001	11

			10	0.05	55
1061	20	5	5	0.001	2
			5	0.05	28
			10	0.001	7
			10	0.05	50
1061	20	10	5	0.001	9
			5	0.05	54
			10	0.001	31
			10	0.05	84
1061	20	20	5	0.001	40
			5	0.05	88
			10	0.001	86
			10	0.05	100
1061	20	30	5	0.001	15
			5	0.05	64
			10	0.001	46
			10	0.05	91
1061	30	5	5	0.001	1
			5	0.05	19
			10	0.001	3
			10	0.05	33
1061	30	10	5	0.001	4
			5	0.05	35
			10	0.001	13
			10	0.05	62
1061	30	20	5	0.001	17
			5	0.05	66
			10	0.001	53
			10	0.05	93
1061	30	30	5	0.001	41
			5	0.05	88
			10	0.001	88
			10	0.05	100