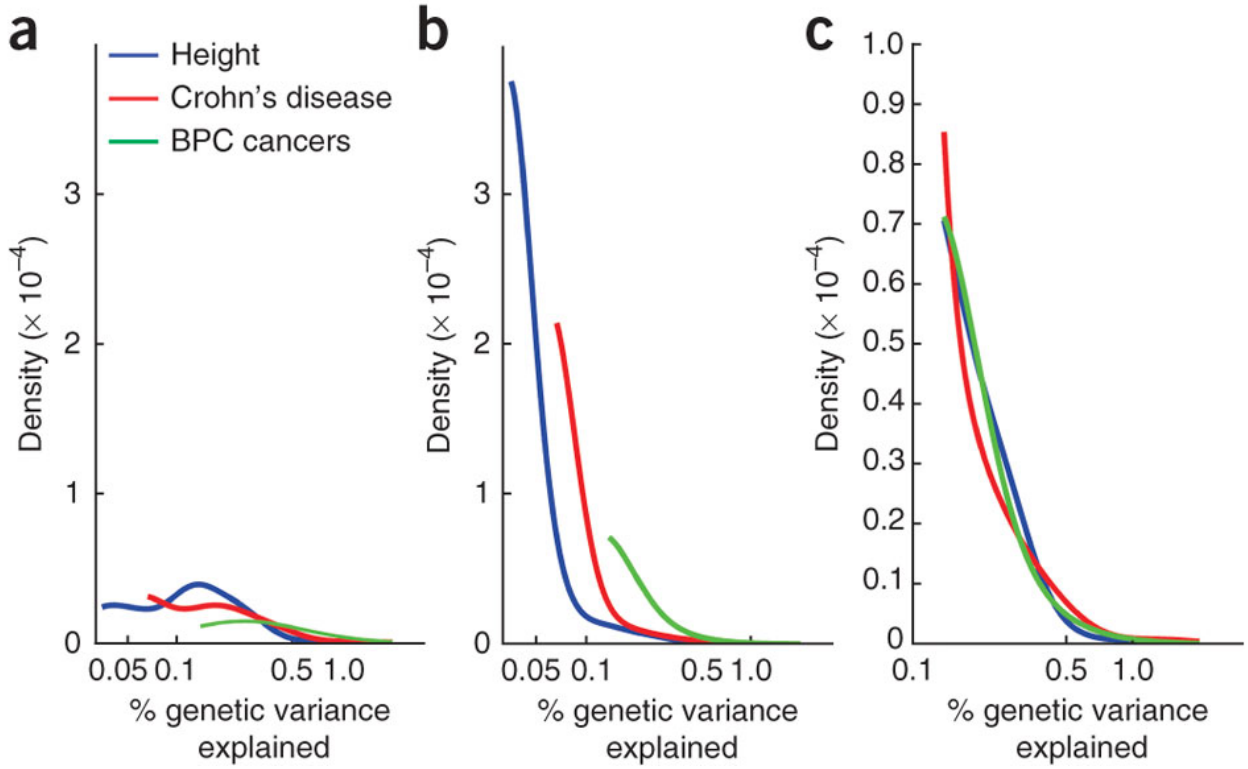
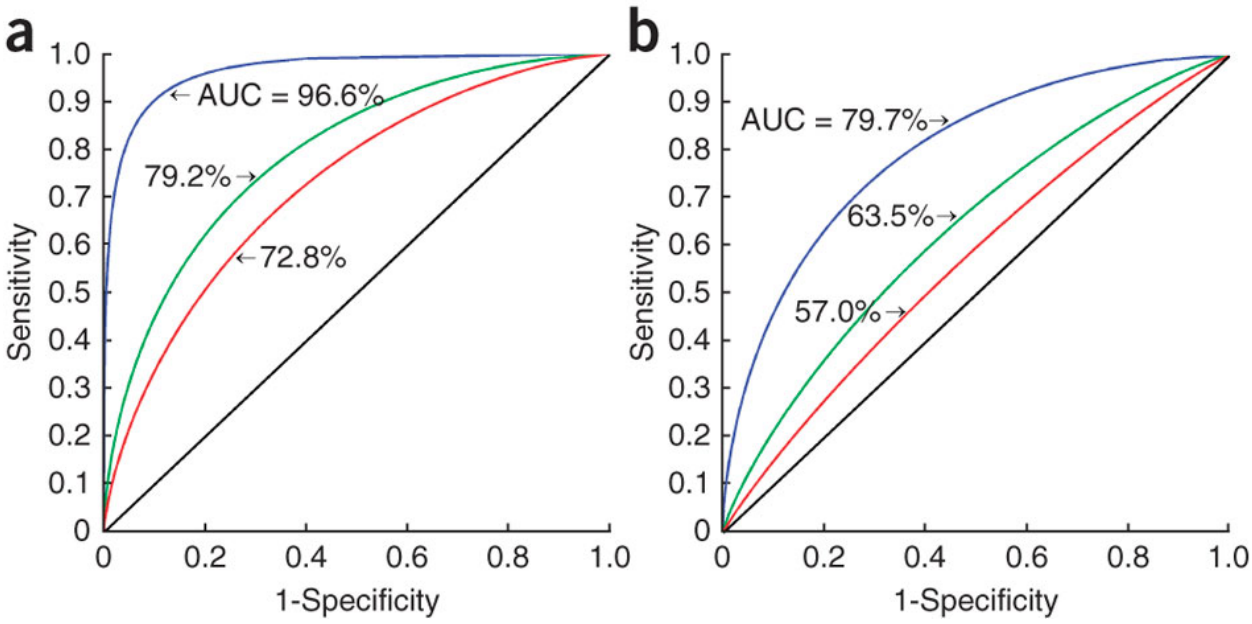
Figure 1



(a) Curves based only on observed susceptibility loci; these curves are distorted because loci with larger effect sizes are more likely to have been detected. (**b**) Curves based on estimated susceptibility loci, representative of the population of all susceptibility loci. (**c**) Estimated nonparametric distributions after normalization over the common observed range for the three traits.

Figure 2



(a,b) Curves for Crohn's disease (a) and BPC cancers (b). AUC is a measure of the discriminatory power of the risk model. Blue, a theoretical genetic risk model that explains all of the known familial risk of the trait. Green, a risk model that includes all of the susceptibility loci (142 for Crohn's disease and 67 on average for BPC cancers) estimated to exist within the range of effect sizes seen in the current GWASs. Red, a risk model that includes only known susceptibility loci (~30 for Crohn's disease and ~7 on average for each of the BPC cancers), which we used to estimate the distribution of effect sizes of these traits. Black, reference line corresponding to a model without discriminatory power in which cases have the same distribution of risk as controls.

Table 1: Estimated numbers of common susceptibility SNPs, and associated genetic variances explained, for three complex traits

	Estimated number of total loci (95% CI)	Total GV[a] explained by estimated loci (95% CI)	Observed range of effect sizes (% GV)
Height	201 (75, 494)	16.4 (10.6, 30.6)	0.04–1.13
Crohn's disease	142 (71, 244)	20.0 (15.7, 28.0)	0.07–1.96
BPC[b] cancers	67 (31, 173)	17.1 (11.6, 35.8)	0.14–1.82

All the projections were performed using a nonparametric method and are restricted to the range of observed effect sizes for known susceptibility SNPs (shown in the last column).

[a]All genetic variances (GV) are shown as a percentage of the total variance of the trait attributable to heritability. For Crohn's disease and BPC cancers, the variance due to heritability is computed from estimates of sibling relative risk using a log-normal model for risk[5].

[b]All estimates should be interpreted as averages over the three cancers.

Table 2: Cumulative number of susceptibility loci expected to be discovered from a single-stage GWAS with increasing sample sizes

Height			Crohn's disease			BPC[c] cancers		
Sample size	Expected number of discoveries	Expected GV[a] explained	Sample size[b]	Expected number of discoveries	Expected GV explained	Sample size[b]	Expected number of discoveries	Expected GV explained
25,000	27.4	6.6	10,000	26.0	11.1	10,000	2.8	2.8
50,000	74.6	10.3	20,000	64.4	14.6	20,000	10.1	5.8
75,000	125.7	13.2	30,000	108.2	17.7	30,000	21.2	8.7
100,000	161.6	14.9	40,000	132.7	19.3	40,000	33.6	11.4
125,000	182.9	15.7	50,000	140.1	19.8	50,000	44.5	13.5

The projections were obtained by accounting for the estimated distribution of effect sizes for the traits. All calculations are based on a significance level for discovery of 10^{-7}.

[a]All genetic variances (GV) are shown as a percentage of the total variance of the trait attributable to heritability. For Crohn's disease and BPC cancers, the variance due to heritability is computed from estimates of sibling relative risk using a log-normal model for risk[5].

[b]Sample size assumes 50% affected individuals and 50% controls.

[c]All estimates should be interpreted as averages over the three cancers.

Table 3: Sample size required for detecting novel loci in first or later studies

Height		Crohn's disease		BPC cancers	
No. of novel loci	Sample size (first study/later study)	No. of novel loci	Sample size (first study/later study)	No. of novel loci	Sample size (first study/later study)
25	24,800/40,100	15	6,700/14,000	5	15,500/25,100
50	39,800/53,000	30	11,900/18,300	10	21,700/31,000
75	52,100/65,300	45	16,300/21,900	15	26,600/35,900
100	63,900/79,100	60	19,800/25,300	20	31,000/40,600
125	77,000/96,800	75	23,100/28,900	25	35,100/45,400

Shown are sample sizes required for a single-stage GWAS to have 80% probability of detecting the specified number of novel loci (or more), when it is either the first study (all loci will be novel) or a later study ('novel' loci exclude known susceptibility loci detected in earlier studies), with a significance level of 10^{-7}. For 'later studies', only SNPs used for the estimation of the effect size distribution were excluded. For a number of these traits, there are known additional loci, and thus the sample-size requirement for later studies is expected to increase when all known susceptibility loci are accounted for.

Table 4: Expected and observed numbers of discoveries in external data sets

	Expected number of discoveries	Observed number of discoveries
Cancer		
Total number of discoveries in CGEMS prostate two-stage study	2.7	5
Total number of discoveries in CGEMS breast two-stage study	3.0	3
Number of additional discoveries in the latest CRUK prostate study[b]	9.5	7–9[c]
Height		
Number of additional discoveries in ref. 13 after inclusion of stage 2[d]	9.3	11
GIANT consortium[e]	186	Not available

Data sets used for this validation exercise were not used in selection of loci for estimating effect size distribution. All calculations are based on a genome-wide significance of 10^{-7}.

[a]Obtained using the externally estimated distributions of effect sizes, along with sample size and study design of the specified studies.

[b]Data from only five prostate cancer loci discovered from the original CRUK prostate study contributed to the estimation of the distribution of effect sizes of BPC cancers. Here, expected number of additional discoveries is calculated as the difference between expected number of discoveries with and without the third-stage data.

[c]Study reported discovery of nine independent susceptibility SNPs from seven different chromosomal regions.

[d]Data from only 20 loci discovered in the first stage of this study contributed to estimation of the distribution of effect size for height.

[e]Prospective projection for a meta-analysis of GWAS data for 130,000 subjects.