Genome-wide detection of intervals of genetic heterogeneity underlying complex traits

Felipe Llinares-López, Dominik G. Grimm, Dean A. Bodenham
Udo Gieraths, Mahito Sugiyama, Beth Rowan, Karsten Borgwardt

S1 Supplementary Tables

Phenotype Name	**#Samples**	**%Cases**	**#SNPs**	$\boldsymbol{\delta^*_{\text{Bonferroni}}}$	$\boldsymbol{\delta_{\texttt{FAIS}}}$	$\boldsymbol{\delta_{\texttt{FAIS-WY}}}$
Chlorosis 16	176	47.73	214 051	2.18e-12	2.09e-08	1.70e-07
Chlorosis 10	177	15.82	214 051	2.18e-12	1.52e-08	1.96e-07
Leaf roll 22	176	17.61	214 051	2.18e-12	1.52e-08	1.85e-07
Emco5	86	80.23	214 044	2.18e-12	2.41e-08	3.81e-07
Emoy	76	53.95	213 981	2.18e-12	3.26e-08	3.13e-07
Hiks1	84	60.71	213 988	2.18e-12	2.65e-08	2.56e-07
Noco2	87	55.17	214 043	2.18e-12	2.96e-08	3.18e-07
Anthocyanin 16	176	39.77	214 051	2.18e-12	1.90e-08	1.54e-07
Anthocyanin 10	177	18.64	214 051	2.18e-12	1.52e-08	1.86e-07
Anthocyanin 22	177	36.16	214 051	2.18e-12	1.85e-08	1.56e-07
Emwa1	85	62.35	214 042	2.18e-12	2.64e-08	2.71e-07
avrRpt2	89	80.9	214 032	2.18e-12	2.40e-08	3.44e-07
avrB	87	63.22	214 032	2.18e-12	2.48e-08	2.68e-07
Leaf roll 16	176	21.02	214 051	2.18e-12	1.57e-08	1.77e-07
avrRpm1	84	66.67	214 022	2.18e-12	2.42e-08	2.99e-07
Chlorosis 22	176	62.5	214 051	2.18e-12	1.85e-08	1.55e-07
Leaf roll 10	177	55.93	214 051	2.18e-12	1.99e-08	1.74e-07
avrPphB	90	51.11	214 032	2.18e-12	3.16e-08	2.86e-07
LES	95	22.11	214 051	2.18e-12	2.04e-08	3.32e-07
LY	95	30.53	214 051	2.18e-12	2.17e-08	2.80e-07
YEL	95	8.42	214 051	2.18e-12	4.64e-08	9.67e-07

Table S1: Corrected significance threshold of `FAIS` ($\boldsymbol{\delta_{\texttt{FAIS}}}$), `FAIS-WY` ($\boldsymbol{\delta_{\texttt{FAIS-WY}}}$) and the exhaustive Bonferroni significance threshold ($\boldsymbol{\delta^*_{\text{Bonferroni}}}$).

Description	Phenotype	Chr	Position	Ref	Alt	MAF	snpEFF Effect	Codon Change	AA Change	Gene ID	P-Value
Hypersensitive Response	*avrB*	3	2288913	G	T	0.046	5 Prime UTR Variant			AT3G07195	2.92E-01
		3	2289178	T	A	0.23	Intron Variant			AT3G07195	3.05E-05
		3	2289559	T	A	0.356	Synonymous Variant	cgT/cgA	Arg58Arg	AT3G07195	1.94E-02
	avrPphB	1	4139802	T	C	0.167	Synonymous Variant	agT/agC	Ser455Ser	AT1G12200	1.62E-05
		1	4140044	A	T	0.256	Intergenic				2.88E-07
	avrRpm1	3	2310055	T	G	0.155	Missense Variant	Tgg/Ggg	Trp118Gly	AT3G07260	2.78E-06
		3	2311035	T	G	0.202	Intergenic				8.26E-02
		3	2311574	C	T	0.179	Intergenic				1.28E-01
	avrRpt2	4	13231162	T	C	0.247	Intergenic				1.02E-04
		4	13231999	G	T	0.056	Intergenic				1.49E-04
Leaf roll presence	Leaf roll 10	2	14018996	A	C	0.079	Synonymous Variant	gcT/gcG	Ala320Ala	AT2G33040	4.73E-02
		2	14019038	A	G	0.085	Synonymous Variant	tcT/tcC	Ser306Ser	AT2G33040	5.88E-01
		2	14019977	G	C	0.096	Missense Variant	atC/atG	Ile120Met	AT2G33040	3.71E-02
		2	14020898	G	T	0.124	Intergenic				3.60E-01
		2	14021726	G	T	0.079	Intergenic				1.60E-01
		2	14022076	T	A	0.124	Synonymous Variant	ctT/ctA	Leu69Leu	AT2G33050	5.26E-03
		2	14023462	A	T	0.079	Synonymous Variant	tcA/tcT	Ser531Ser	AT2G33050	1.60E-01
		2	14023726	A	T	0.294	Synonymous Variant	atA/atT	Ile619Ile	AT2G33050	8.19E-03
		2	14023855	C	G	0.09	Missense Variant	agC/agG	Ser662Arg	AT2G33050	6.18E-02
		2	14023969	C	A	0.085	Synonymous Variant	ctC/ctA	Leu700Leu	AT2G33050	1.01E-01
		2	14024162	C	T	0.09	Missense Variant	Ctt/Ttt	Leu765Phe	AT2G33050	6.18E-02
		2	14024485	C	A	0.09	Intergenic				6.18E-02
		2	14025154	T	G	0.085	Intergenic				1.01E-01
		2	14025341	A	G	0.147	Intergenic				5.76E-02
		2	14025518	T	G	0.356	5 Prime UTR Variant			AT2G33060	4.88E-04
	Leaf roll 16	5	18040355	A	G	0.136	Intergenic				6.14E-04
		5	18040400	C	T	0.188	Intergenic				1.75E-03
Presence or absence of lesioning	LES	4	8307440	T	G	0.389	Missense Variant	Aaa/Caa	Lys75Gln	AT4G14440	2.26E-04
		4	8307761	A	T	0.147	Intergenic				1.82E-01
		4	8307910	T	A	0.232	Intergenic				1.42E-01
		4	8308076	A	G	0.368	Intergenic				4.02E-03
		4	8308306	T	A	0.074	Intergenic				3.42E-01
		4	8308768	T	A	0.042	Intergenic				5.72E-01
		4	8308977	C	A	0.295	Intergenic				2.88E-01
Presence or absence of either lesioning or yellowing	LY	1	18192753	G	A	0.253	Missense Variant	gGc/gAc	Gly406Asp	AT1G49190	3.87E-02
		1	18192956	C	A	0.263	Intron Variant			AT1G49190	7.99E-02
		1	18193152	A	C	0.211	Intron Variant			AT1G49190	2.83E-04
		1	18196934	C	G	0.284	Intergenic				4.80E-02
Presence or absence of yellowing	YEL	3	20137884	C	A	0.495	3 Prime UTR Variant			AT3G54390	5.69E-03
		3	20139860	T	G	0.358	Intergenic				2.52E-01
		3	20142484	A	G	0.211	Missense Variant	cTa/cCa	Leu39Pro	AT3G54400	1.97E-01
		3	20143997	T	G	0.421	Intergenic				1.33E-01
		3	20144047	C	G	0.053	Intergenic				1.00E+00
		3	20144480	A	C	0.411	Intergenic				1.93E-02
Visual anthocyanin presence	Anthocyanin 10	1	5951857	G	T	0.085	Synonymous Variant	ctC/ctA	Leu415Leu	AT1G17370	7.71E-02
		1	5952292	A	G	0.362	Synonymous Variant	ttT/ttC	Phe301Phe	AT1G17370	3.21E-05
		1	5954613	T	A	0.119	Intron Variant			AT1G17370	7.69E-01
	Anthocyanin 16	2	1702970	A	G	0.057	Synonymous Variant	gtT/gtC	Val106Val	AT2G04845	6.46E-03
		2	1703341	A	T	0.091	Intron Variant			AT2G04845	2.96E-03
		2	1703427	G	A	0.136	Synonymous Variant	gtC/gtT	Val70Val	AT2G04845	1.35E-02
		2	1703930	T	A	0.091	Missense Variant	Atg/Ttg	Met22Leu	AT2G04845	2.96E-03
	Anthocyanin 22	1	14037978	A	G	0.254	Intergenic				1.14E-02
		1	14039321	T	C	0.35	Intergenic				7.43E-01
		1	14040254	C	A	0.232	Intergenic				1.52E-02
		1	14040529	G	A	0.237	Intergenic				9.68E-03
Visual Chlorosis precense	Chlorosis 22	3	9825219	T	G	0.176	Missense Variant	Ttg/Gtg	Leu400Val	AT3G26730	1.98E-04
		3	9825306	T	C	0.17	Synonymous Variant	Ttg/Ctg	Leu429Leu	AT3G26730	6.86E-04
		3	9825506	A	T	0.034	Synonymous Variant	tcA/tcT	Ser495Ser	AT3G26730	1.99E-01
		3	9825646	G	A	0.176	Intron Variant			AT3G26730	1.98E-04
		3	9825765	A	T	0.188	Intron Variant			AT3G26730	1.15E-03
		3	9826159	A	C	0.176	Intron Variant			AT3G26730	1.98E-04
		3	9826344	G	C	0.244	Intron Variant			AT3G26730	2.05E-01
		3	9827153	T	A	0.239	Missense Variant	ttT/ttA	Phe734Leu	AT3G26730	1.45E-01
		3	9829739	A	T	0.25	Intergenic				1.10E-01
		3	9829837	G	A	0.415	Intergenic				6.38E-01
		3	9831240	A	T	0.29	5 Prime UTR Variant			AT3G26742	2.53E-02
Visual Chlorosis precense	Emwa1	1	5779265	T	C	0.071	Synonymous Variant	aaA/aaG	Lys570Lys	AT1G16900	2.07E-03
		1	5779815	C	T	0.118	Synonymous Variant	tcG/tcA	Ser447Ser	AT1G16900	3.66E-02
		1	5780043	C	T	0.141	Missense Variant	Gca/Aca	Ala429Thr	AT1G16900	5.37E-05

Table S2: `snpEFF` annotation of the most significant intervals that do not contain a significant SNP found by an univariate Fisher exact test. For each individual SNP, the p-value obtained from the univariate Fisher exact test is displayed in the right-most column.

Description	Phenotype	Chr	Position	Ref	Alt	MAF	snpEFF Effect	Codon Change	AA Change	Gene ID	P-Value
Hypersensitive Response	*avrB*	3	2288913	G	T	0.046	5 Prime UTR Variant			AT3G07195	2.92E-01
		3	2289178	T	A	0.23	Intron Variant			AT3G07195	3.05E-05
		3	2289559	T	A	0.356	Synonymous Variant	cgT/cgA	Arg58Arg	AT3G07195	1.94E-02
	avrPphB	1	4129510	T	C	0.189	Synonymous Variant	tcA/tcG	Ser64Ser	AT1G12170	1.00E-06
		1	4130384	A	T	0.078	Intergenic				5.13E-03
	avrRpm1	3	2310055	T	G	0.155	Missense Variant	Tgg/Ggg	Trp118Gly	AT3G07260	2.78E-06
		3	2311035	T	G	0.202	Intergenic				8.26E-02
		3	2311574	C	T	0.179	Intergenic				1.28E-01
	avrRpt2	4	12279427	T	C	0.133	Missense Variant	aTg/aCg	Met86Thr	AT4G23530	9.40E-03
		4	12280837	G	C	0.133	3 Prime UTR Variant			AT4G23530	1.12E-04
Leaf roll presence	Leaf roll 10	2	14018996	A	C	0.079	Synonymous Variant	gcT/gcG	Ala320Ala	AT2G33040	4.73E-02
		2	14019038	A	G	0.085	Synonymous Variant	tcT/tcC	Ser306Ser	AT2G33040	5.88E-01
		2	14019977	G	C	0.096	Missense Variant	atC/atG	Ile120Met	AT2G33040	3.71E-02
		2	14020898	G	T	0.124	Intergenic				3.60E-01
		2	14021726	G	T	0.079	Intergenic				1.60E-01
		2	14022076	T	A	0.124	Synonymous Variant	ctT/ctA	Leu69Leu	AT2G33050	5.26E-03
		2	14023462	A	T	0.079	Synonymous Variant	tcA/tcT	Ser531Ser	AT2G33050	1.60E-01
		2	14023726	A	T	0.294	Synonymous Variant	atA/atT	Ile619Ile	AT2G33050	8.19E-03
		2	14023855	C	G	0.09	Missense Variant	agC/agG	Ser662Arg	AT2G33050	6.18E-02
		2	14023969	C	A	0.085	Synonymous Variant	ctC/ctA	Leu700Leu	AT2G33050	1.01E-01
		2	14024162	C	T	0.09	Missense Variant	Ctt/Ttt	Leu765Phe	AT2G33050	6.18E-02
		2	14024485	C	A	0.09	Intergenic				6.18E-02
		2	14025154	T	G	0.085	Intergenic				1.01E-01
		2	14025341	A	G	0.147	Intergenic				5.76E-02
		2	14025518	T	G	0.356	5 Prime UTR Variant			AT2G33060	4.88E-04
	Leaf roll 16	5	18040355	A	G	0.136	Intergenic				6.14E-04
		5	18040400	C	T	0.188	Intergenic				1.75E-03
Presence or absence of lesioning	LES	5	6528991	T	G	0.105	Missense Variant	Aaa/Caa	Lys164Gln	AT5G19380	5.70E-05
		5	6529733	G	C	0.053	Intron Variant				3.05E-01
		5	6531656	C	A	0.105	5 Prime UTR Variant			AT5G19390	6.98E-03
Presence or absence of either lesioning or yellowing	LY	1	18192753	G	A	0.253	Missense Variant	gGc/gAc	Gly406Asp	AT1G49190	3.87E-02
		1	18192956	C	A	0.263	Intron Variant			AT1G49190	7.99E-02
		1	18193152	A	C	0.211	Intron Variant			AT1G49190	2.83E-04
		1	18196934	C	G	0.284	Intergenic				4.80E-02
Presence or absence of yellowing	YEL	3	20137884	C	A	0.495	3 Prime UTR Variant			AT3G54390	5.69E-03
		3	20139860	T	G	0.358	Intergenic				2.52E-01
		3	20142484	A	G	0.211	Missense Variant	cTa/cCa	Leu39Pro	AT3G54400	1.97E-01
		3	20143997	T	G	0.421	Intergenic				1.33E-01
		3	20144047	C	G	0.053	Intergenic				1.00E+00
		3	20144480	A	C	0.411	Intergenic				1.93E-02
Visual anthocyanin presence	Anthocyanin 10	1	5951857	G	T	0.085	Synonymous Variant	ctC/ctA	Leu415Leu	AT1G17370	7.71E-02
		1	5952292	A	G	0.362	Synonymous Variant	ttT/ttC	Phe301Phe	AT1G17370	3.21E-05
		1	5954613	T	A	0.119	Intron Variant			AT1G17370	7.69E-01
	Anthocyanin 16	2	1702970	A	G	0.057	Synonymous Variant	gtT/gtC	Val106Val	AT2G04845	6.46E-03
		2	1703341	A	T	0.091	Intron Variant			AT2G04845	2.96E-03
		2	1703427	G	A	0.136	Synonymous Variant	gtC/gtT	Val70Val	AT2G04845	1.35E-02
		2	1703930	T	A	0.091	Missense Variant	Atg/Ttg	Met22Leu	AT2G04845	2.96E-03
	Anthocyanin 22	1	14037978	A	G	0.254	Intergenic				1.14E-02
		1	14039321	T	C	0.35	Intergenic				7.43E-01
		1	14040254	C	A	0.232	Intergenic				1.52E-02
		1	14040529	G	A	0.237	Intergenic				9.68E-03
Visual Chlorosis precense	Chlorosis 22	3	9825219	T	G	0.176	Missense Variant	Ttg/Gtg	Leu400Val	AT3G26730	1.98E-04
		3	9825306	T	C	0.17	Synonymous Variant	Ttg/Ctg	Leu429Leu	AT3G26730	6.86E-04
		3	9825506	A	T	0.034	Synonymous Variant	tcA/tcT	Ser495Ser	AT3G26730	1.99E-01
		3	9825646	G	A	0.176	Intron Variant			AT3G26730	1.98E-04
		3	9825765	A	T	0.188	Intron Variant			AT3G26730	1.15E-03
		3	9826159	A	C	0.176	Intron Variant			AT3G26730	1.98E-04
		3	9826344	G	C	0.244	Intron Variant			AT3G26730	2.05E-01
		3	9827153	T	A	0.239	Missense Variant	ttT/ttA	Phe734Leu	AT3G26730	1.45E-01
		3	9829739	A	T	0.25	Intergenic				1.10E-01
		3	9829837	G	A	0.415	Intergenic				6.38E-01
		3	9831240	A	T	0.29	5 Prime UTR Variant			AT3G26742	2.53E-02
Visual Chlorosis precense	Emwa1	1	5779265	T	C	0.071	Synonymous Variant	aaA/aaG	Lys570Lys	AT1G16900	2.07E-03
		1	5779815	C	T	0.118	Synonymous Variant	tcG/tcA	Ser447Ser	AT1G16900	3.66E-02
		1	5780043	C	T	0.141	Missense Variant	Gca/Aca	Ala429Thr	AT1G16900	5.37E-05

Table S3: `snpEFF` annotation of the most significant intervals that do not contain or are in close proximity ($\pm 10kb$) to a significant SNP found by an univariate Fisher exact test or a linear mixed model. For each individual SNP, the p-value obtained from the univariate Fisher exact test is displayed in the right-most column. Rows in blue are different intervals then identified using the filtering Scenario 2.

S2 Supplementary Figures

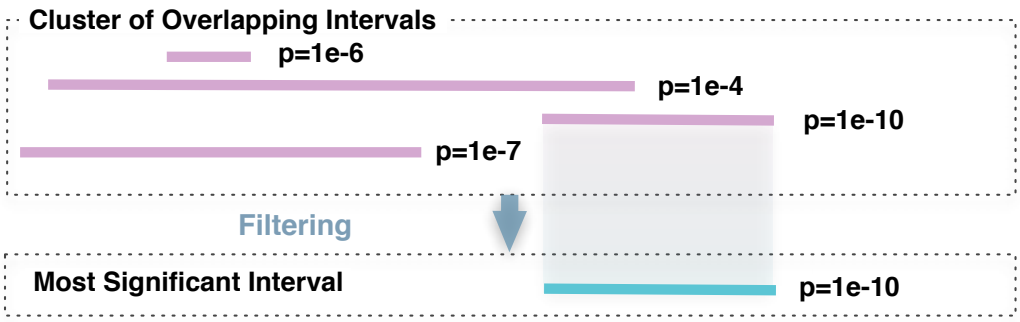

Figure S1: A schematic representation of the filtering procedure. In each cluster of overlapping intervals, the interval with smallest p-value is the result.

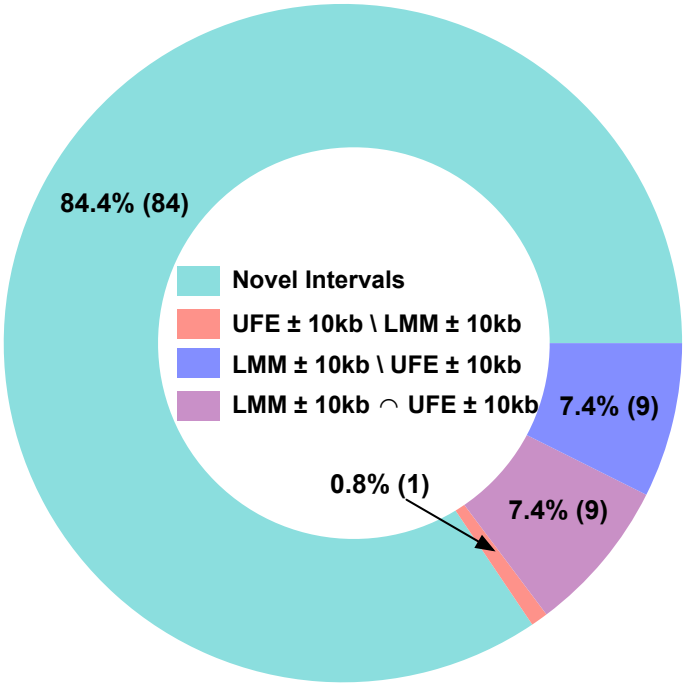

Figure S2: Proportion of novel intervals among all intervals found by `FAIS-WY` for the phenotype YEL with the highest degree of population structure. The green part shows the proportion of novel intervals found by `FAIS-WY`. The red part (UFE ± 10kb \LMM ± 10kb) are intervals containing an UFE hit or are in close proximity (± 10kb) to one and the hit could not be found with a LMM. The blue part (LMM ± 10kb \UFE ± 10kb) are intervals containing a LMM hit or are in close proximity (± 10kb) to one and the hit could not be found with an UFE. The purple part (LMM ± 10kb ∩ UFE ± 10kb) are intervals that contain both, a hit (± 10kb) found with an UFE and a LMM.

S3 Simulation study: additional results

The supplementary figures in this section may aid in our discussion of the performance of `FAIS` and `FAIS-WY`. In all figures in this section, $\alpha = 0.05$.

S3.1 Speed, as the number of permutations varies

Figure S3 shows how the speed of `FAIS-WY` varies as the number of permutations J increases. Of course, `FAIS` does not vary, since it does not use Westfall-Young permutations, but is displayed for comparison purposes. The relationship appears to be approximately linear (speed vs number of permutations, J). This shows that if the simulation that produced Figure 4 (in the main text) used a larger number of permutations, there would be a larger difference in the (raw) times between `FAIS` and `FAIS-WY`.

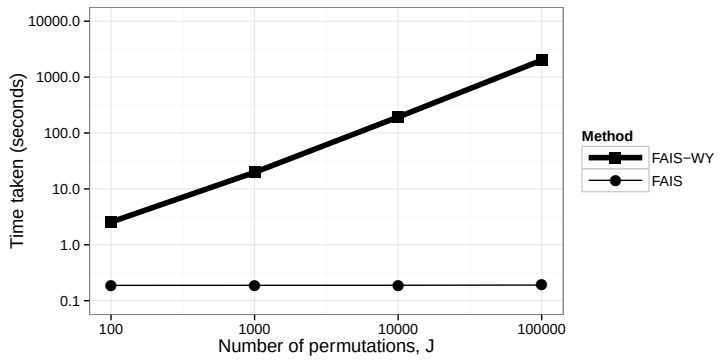

Figure S3: A figure comparing the speed of `FAIS` and `FAIS-WY` as the number of permutations varies. Of course, `FAIS` does not vary, but is displayed for comparison purposes. Note that the axes are log-scaled (base 10). Other parameter values: n=100, L=11000.

S3.2 Speed, as the number of individuals varies

Figure S4 shows how the speeds of `FAIS` and `FAIS-WY` vary as n, the number of individuals, increases. Their speeds both appear to increase (approximately) linearly as n increases (but with different slopes).

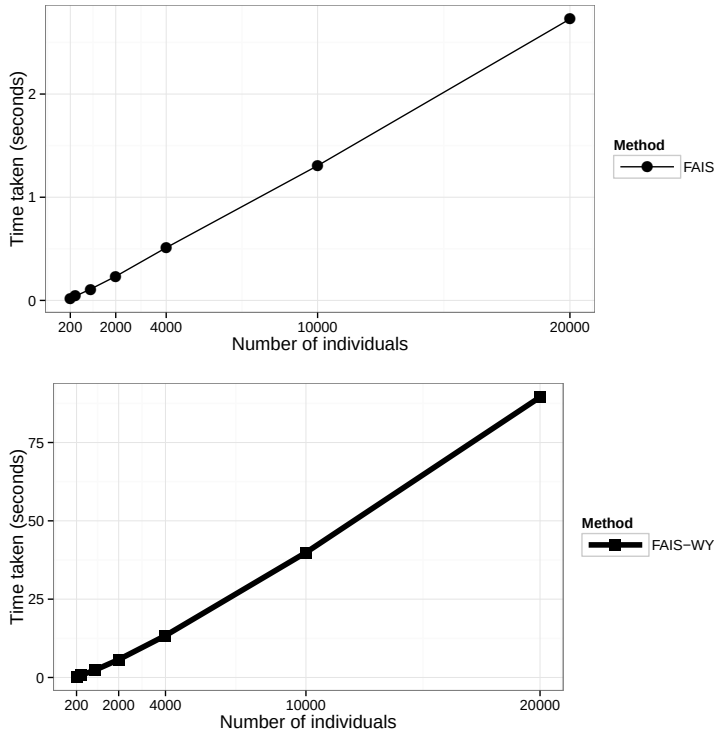

Figure S4: A figure comparing the speed of `FAIS` and `FAIS-WY` as the number of individuals varies. Note that the axes are NOT log-scaled. Other parameter values: L=1100, J=100.

S3.3 Accuracy

Figure S5 shows the proportion of correctly detected significant intervals for `FAIS`, `FAISWY`, `BRUTE` and `UFE`, for different values of p_{case}, as the length of the significant interval increases, averaged over $M = 1000$ trials. It appears that the length of the significant interval does not appear to play too much of a role for the first three methods - for example, see panel $p_{case} = 0.5$, where the detection rates are fairly constant per method for `FAIS`, `FAISWY` and `BRUTE`. However, note that `UFE` performs poorly for longer intervals, for $p_{case} < 0.9$.

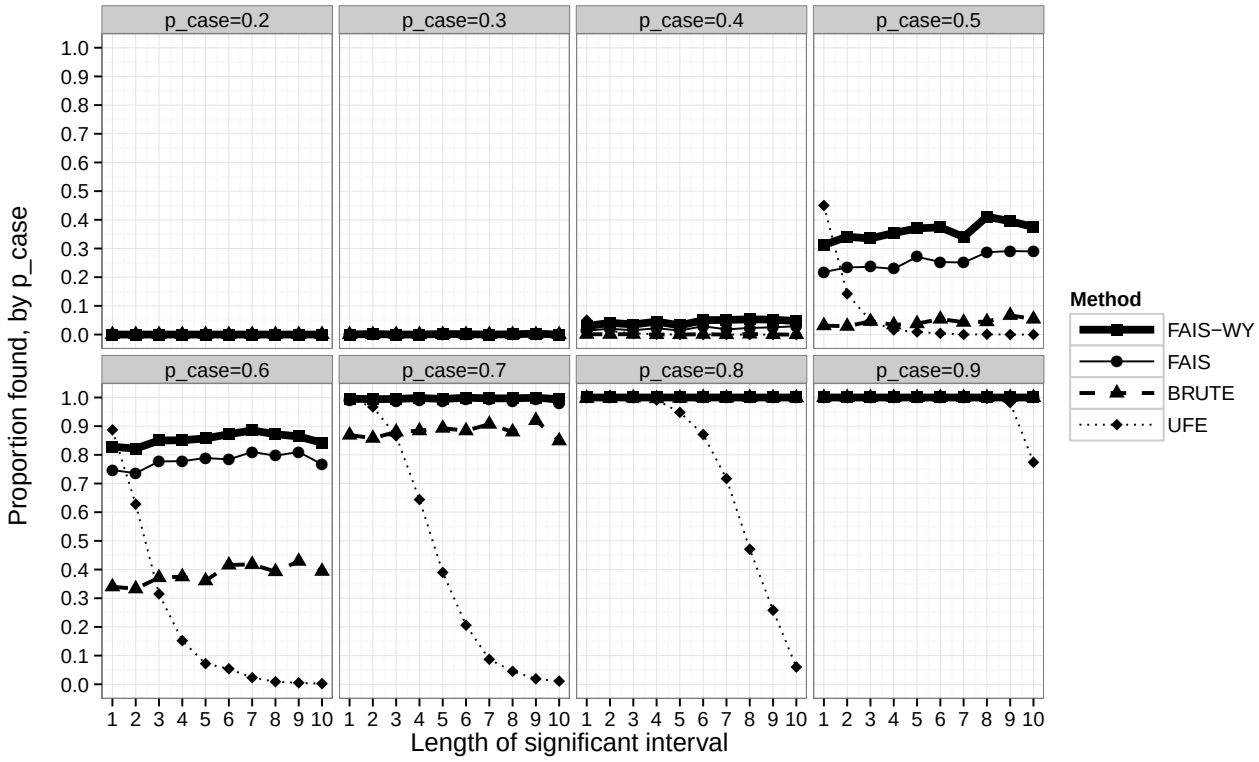

Figure S5: A figure comparing the accuracies of `FAIS`, `FAIS-WY`, `BRUTE` and `UFE` for different values of p_{case}, as the length of the significant interval increases. Other parameter values: n=100, J=1000, number of M=1000.

Figure S6 (on the next page) shows the same information, with the roles of p_{case} and l, the length of the significant interval, exchanged. This figure more clearly shows that `FAIS` and `FAISWY` are significantly more accurate than `BRUTE`, and that `UFE` performs progressively worse as the length of significant intervals increases. This is to be expected, since `UFE` is only attempting to detect a significant difference in single SNPs. However, this figure decisively shows that, for detecting significant *intervals*, `UFE` is far inferior to interval-search methods.

Finally, Figure S7 shows the resulting False Discovery Rate (FDR) when `FAIS` and `FAIS-WY` are set to control the FWER at $\alpha = 0.05$. As expected, the FDR decreases with increasing p_{case}. This figure should be compared to Figure 4 in the main text which illustrates the difference in power between these four methods.

All simulations were performed on an iMac with CPU 2.9 Ghz Intel Core i5, and 16 GB of RAM (1600 MHz DDR3). Figures were produced in R using ggplot2.

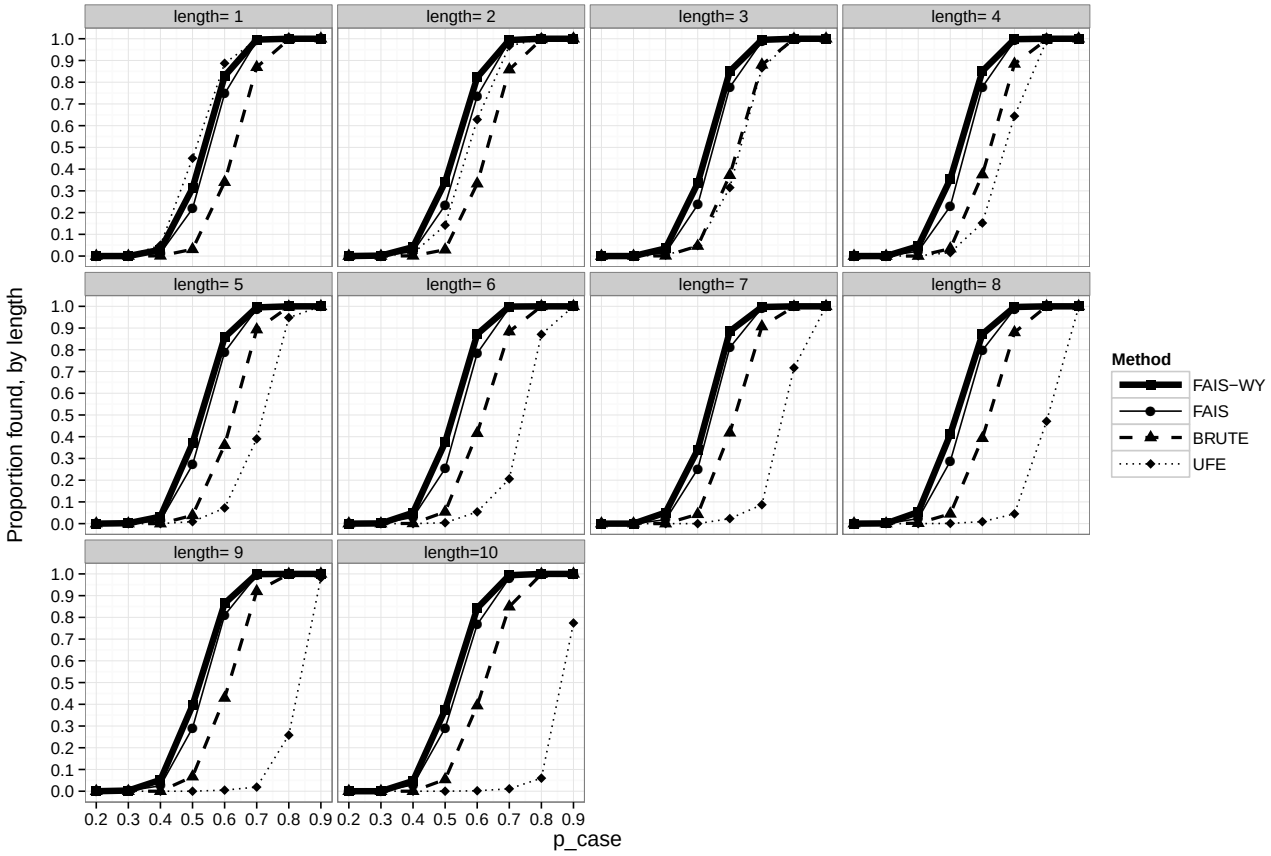

Figure S6: A figure comparing the accuracies of `FAIS` and `FAIS-WY` `BRUTE` and `UFE` for significant intervals of different lengths, as $caseP$ varies. Other parameter values: n=100, J=1000, M=1000.

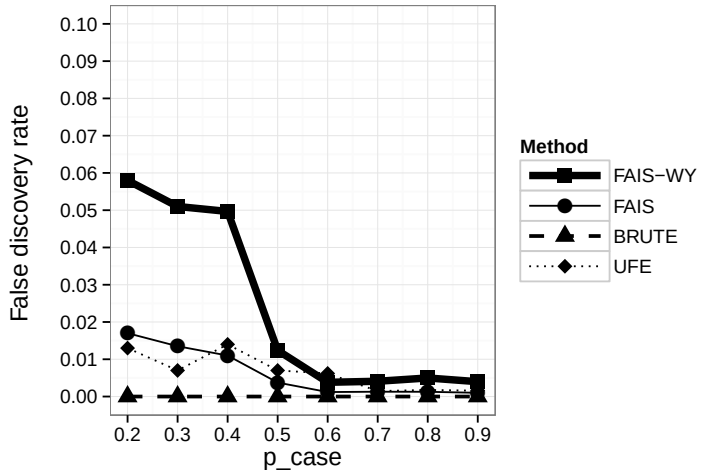

Figure S7: A figure showing the FDR of FAIS, FAISWY and BRUTE, as p_{case} varies. Other parameter values: n=100, J=1000, M=1000.