Supplemental Material

Supplemental Methods

BMPR2 genomic imprinting

We evaluated the potential existence of *BMPR2* genomic imprinting based on the approach presented by Strauch *et al*. (Am J Hum Genet, 66:1945–57, 2000). To that end, the parent-of-origin allele of p.Arg491Gln *BMPR2* mutation was followed in the family. We discarded for the analysis all those individuals whose phenotype was unknown. In addition, two main assumptions were made at the *BMPR2* mutation site. First, all healthy non-carriers, including founders, were considered wild-type homozygous. Second, we assumed that all carriers were heterozygous regarding the p.Arg491Gln mutation. Finally, we compared the penetrance of the heterozygous individuals that had a maternal transmission of the *BMPR2* allele, with the penetrance of those that had a paternal transmission. The observation of significant differences between them may indicate the presence of genomic imprinting in this gene.

Segregation of the BMPR2 mutation in the family

As a quality control, we checked the agreement between the clinical record, the genotyping data and the reported disease causing mutation. In this test, we checked whether the "known gene" *BMPR2* showed evidence of segregation in all carriers. *Parameters: high penetrance (90%), low disease AF (d=10^{-3}, very rare in population), dominant mode of inheritance (MOI) and no phenocopies* (Figure S3).

The known gene test was run with all the linkage programs. In all cases, significant LOD scores (LOD > 3.3) were detected in a wide region of chromosome 2, reaching a maximum value of 6.36 and 7.12 LOD units (Figure S5) using Pseudomarker and Mendel, respectively. Those two-point linkage analysis programs produced nearly identical score profiles within a >30 Mb region of significant linkage (174.7-213.9 Mb; q31.1-q34). As for multipoint linkage, the region was circumscribed to closer boundaries, around 10 Mb in both Morgan (196.1-208.9 Mb; q32.3-q33.3) and Merlin (199.5-209 Mb; q33.1-q33.3) (Figure S6). Remarkably, the local maximum of these regions corresponded to a variant in *BMPR2* (rs2228545) that is located in exon 12 (Figure S7, Figure S8). This variant is only 3,215 bp downstream the p.Arg491Gln mutation, located in exon 11. We obtained similar results with Superlink (Figure S9).

Independent gene contribution

We explored the hypothesis of an independent genetic contribution, apart from *BMPR2*, to HPAH. In this "unknown-gene" test, only clinically affected carriers were marked as affected. *Parameters: low disease AF ($d=10^{-3}$), low penetrance (30%, emulating the observed penetrance) and recessive MOI* (Figure S3).

This "unknown-gene" test did not provide any signal of linkage in the vicinity of *BMPR2* (Figure S10). Negative results were also observed when allowing for 1, 2, 5 and 10 % of phenocopies rate in Mendel for that same model (Figure S11).

Choice of allele frequency in parametric linkage analysis

The linkage analysis technique is specifically oriented towards the detection of rare variants with a strong effect on a particular trait or disease. Accordingly, the statistical power to detect significant linkage is usually limited to low disease frequencies (i.e., d=0.001) and high penetrance, particularly with rare diseases. As one can switch from susceptibility to protection in a linkage model -by changing the mode of inheritance, the penetrance for each genotype and the disease allele frequency- we can also test a high disease frequency (d=0.999) under a recessive model conferring susceptibility, as it is equivalent to a rare disease frequency (p=0.001) conferring protection under a dominant model.

The "rare" (d=0.001) and "common" (d=0.999) disease frequency dichotomy choice that we use is constrained by such limitations on statistical power. In agreement with that, we only observed significant linkage with Merlin multipoint analysis in the vicinity of *FIGN* with the high disease frequency under a susceptibility model. On the contrary, we were unable to detect significant LOD scores in a genome-wide multipoint linkage analysis with intermediate allele frequencies (d=0.22, d=0.4, d=0.6, d=0.8; see Table below). In another approach, we applied the GENEHUNTER MOD-Score functionality on chromosome 2 (data not shown), maximizing the LOD score over different models. The best model outputs a MOD score of 2.927 at 169.5 cM (≈163.15 Mb), also in the vicinity of *FIGN*, with the same disease allele frequency (d=0.999) and a slightly different penetrance vector, {0, 0.33, 1.0}.

Disease allele frequency	Maximum LOD	Maximum LOD chromosomal coordinates	*FIGN* vicinity maximum LOD**
0.22	1.185	17:79,237,900	-0.705
0.4	1.092	22:43,485,385	-0,746
0.6	1.108	3:194,703,666	-0.291
0.8	1.088	3:194,703,666	0.649
0.999	4.09	2:163,738,883-165,107,298**	4.09

** Region of maximum linkage in *FIGN* vicinity

The SNPs in the LD block found within the candidate region and with the strongest functional evidence, present an European MAF of 0.22 according to the 1000 Genomes Project. This intermediate frequency, although considered common in terms of population genetics, it does not match the disease allele frequency used in the linkage parametric model (p=0.999). However, we did use population allele frequencies of SNPs to inform the linkage analysis model, which increases the statistical power to detect linkage. Moreover, the additional prioritization within the candidate region was done using functional genomics data, thus without considering the disease allele frequency of p=0.999 from the linkage model.

Enrichment analyses of EFO terms among candidate regions

Data on the association between SNPs, traits and phenotypes, and their systematic annotation using the Experimental Factor Ontology (EFO) were downloaded from the GWAS Catalog (accessed September 2017), exclusively considering those mapping to GRCh37. To account for the SNPs in linkage disequilibrium (LD) with GWAS Catalog annotations, we searched for genotypes in LD within the 1000 Genomes Project Phase 3. The associated SNPs were identified by using PLINK (R^2>0.8, maximum distance among SNPs=1000 Kb) and then imputed with the same EFO term annotated to the corresponding SNP in LD. EFO terms were

also propagated throughout the hierarchy of the ontology tree using the R package ontologyIndex. A conditional hypergeometric test for EFO term association, applying a one-tailed Fisher's exact test, was used for a functional enrichment analysis with the Bioconductor package GOstats. The resulting list of enriched EFO terms was filtered by considering only those with odds ratio (OR) > 2, minimum EFO term (size) > 5, minimum number of enriching SNPs (count) > 5 and adjusted P-value<10^{-3} using Holm correction.

eQTL analysis

The eQTL analysis of candidate regulatory SNPs was done using GTEx data release V7, downloaded from the dbGaP web site, under phs000424.v7.p2. We first searched for significant *FIGN* cis-eQTLs on the GTEx Portal (see Web Resources in the main text). Then, using the genotype and expression data downloaded from dbGaP, and covariates downloaded from the GTEx Portal, we verified the significant associations between the reported cis-eQTLs and the expression data from corresponding tissues. To show the estimated genotype effect on gene expression in Figure 4C, we removed covariate effects, as provided by GTEx, from the GTEx normalized expression data.

FIGN expression analysis

We downloaded raw Affymetrix CEL files from GEO under accession number GSE53408 and pre-process them using standard procedures. After normalization and filtering, we obtained a gene expression data matrix of 22,144 genes by 23 samples, where 12 were derived from lung tissue of PAH patients and 11 of normal lung tissue. We conducted a differential expression analysis using the R/Bioconductor package limma, comparing PAH patients and controls, adjusting for surrogate variables with the R/Bioconductor package SVA. Co-expression analysis between *FIGN* and *BMPR2* was done using an ANCOVA model where *FIGN* expression was the response variable, *BMPR2* the predictive one and PAH status a factor variable modeling a different intercept term for PAH and control samples.

Haplotype prediction

Haplotypes were predicted between the region of significant linkage and *BMPR2*, using the pruned version of the pedigree employed for Merlin multipoint linkage analysis. Haplotype estimation was performed using the `--best` option, which outputs the most likely pattern of segregation.

Supplemental Figures

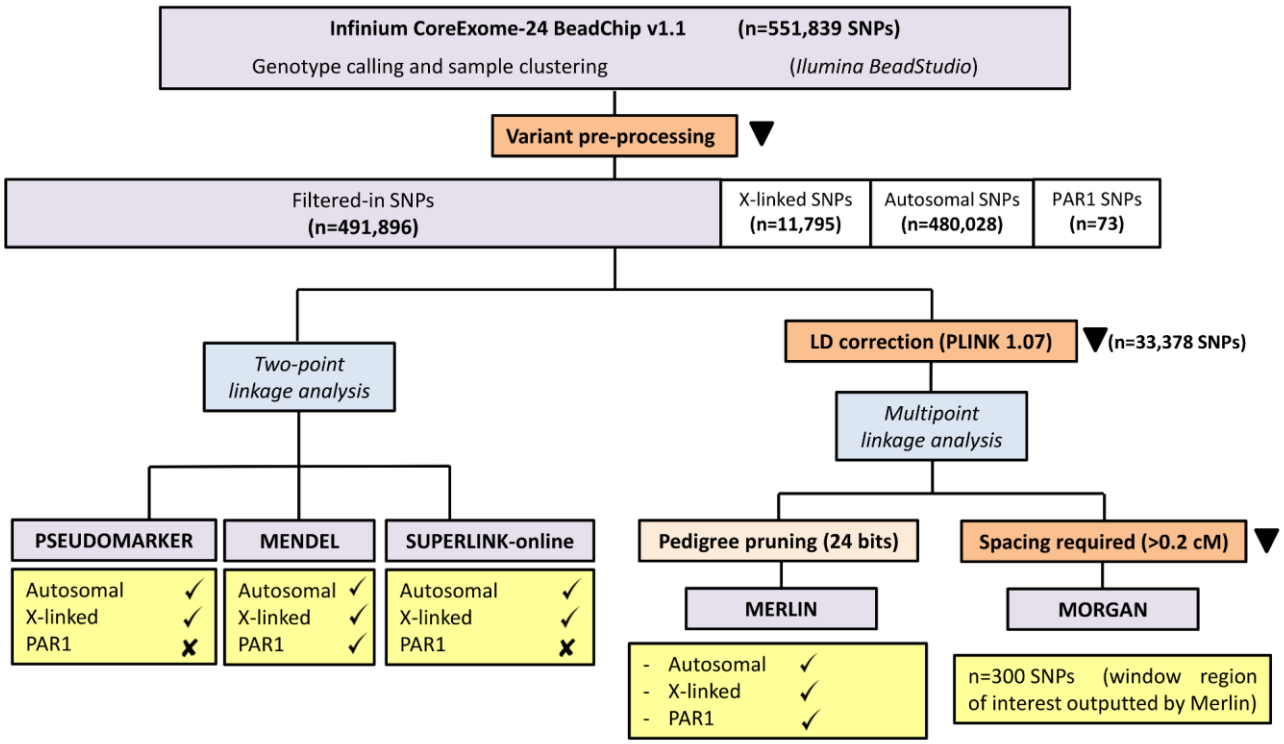

Figure S1. Variant pre-processing pipeline. Data pre-processing steps filtered out 59,943 SNPs that contained Mendelian errors, multiallelic inconsistencies and could not be re-annotated in unique positions. Variants with missing genotypes, missing population allele frequencies in ExAC or 1000 Human Genomes Project, and missing physical or genetic coordinates in hg19/GRCh37, were also discarded. Remaining SNPs were classified in three groups: X-linked, autosomal and the pseudoautosomal region 1 (PAR1). Two different approaches were followed according to the linkage analysis type. In two-point linkage, pre-processed variants were all used for analysis in Mendel, while in Pseudomarker, we discarded PAR1 regions. We also used Superlink-Online, which internally filters out a large fraction of SNPs. As for multi-point linkage, we used PLINK 1.07 to perform linkage disequilibrium (LD) correction to avoid false-positives. This step strongly reduced the number of SNPs considered for the linkage analysis. In Merlin, we additionally pruned the pedigree by creating a sub-pedigree that met the complexity constraint (24 bits) of the Lander-Green algorithm and maximized the number of genotyped affected and healthy carriers. In Morgan, this trimming step was not required, although further SNP pruning was applied by forcing a 0.2 cM genetic map spacing. With that program we only ran 300 SNPs window in the region of interest previously highlighted by Merlin.

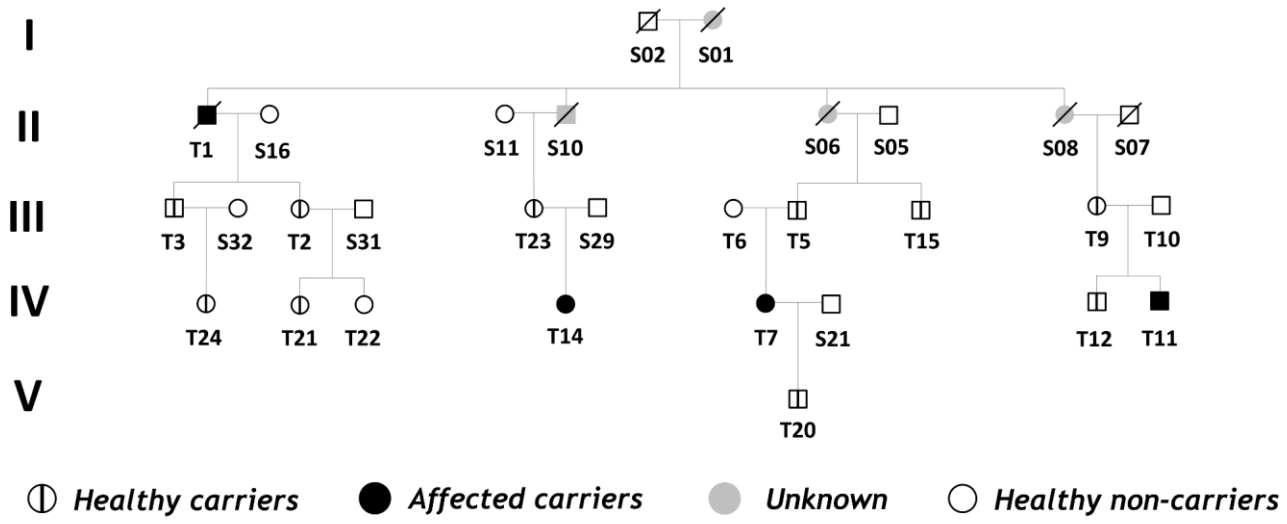

Figure S2. Pruned pedigree with maximum number of genotyped carriers. Sub-pedigree created to meet the complexity upper bound of the Lander-Green algorithm (maximum number of 24 bits in Merlin). It contains 30 individuals, including 4 genotyped affected carriers and 10 genotyped healthy carriers.

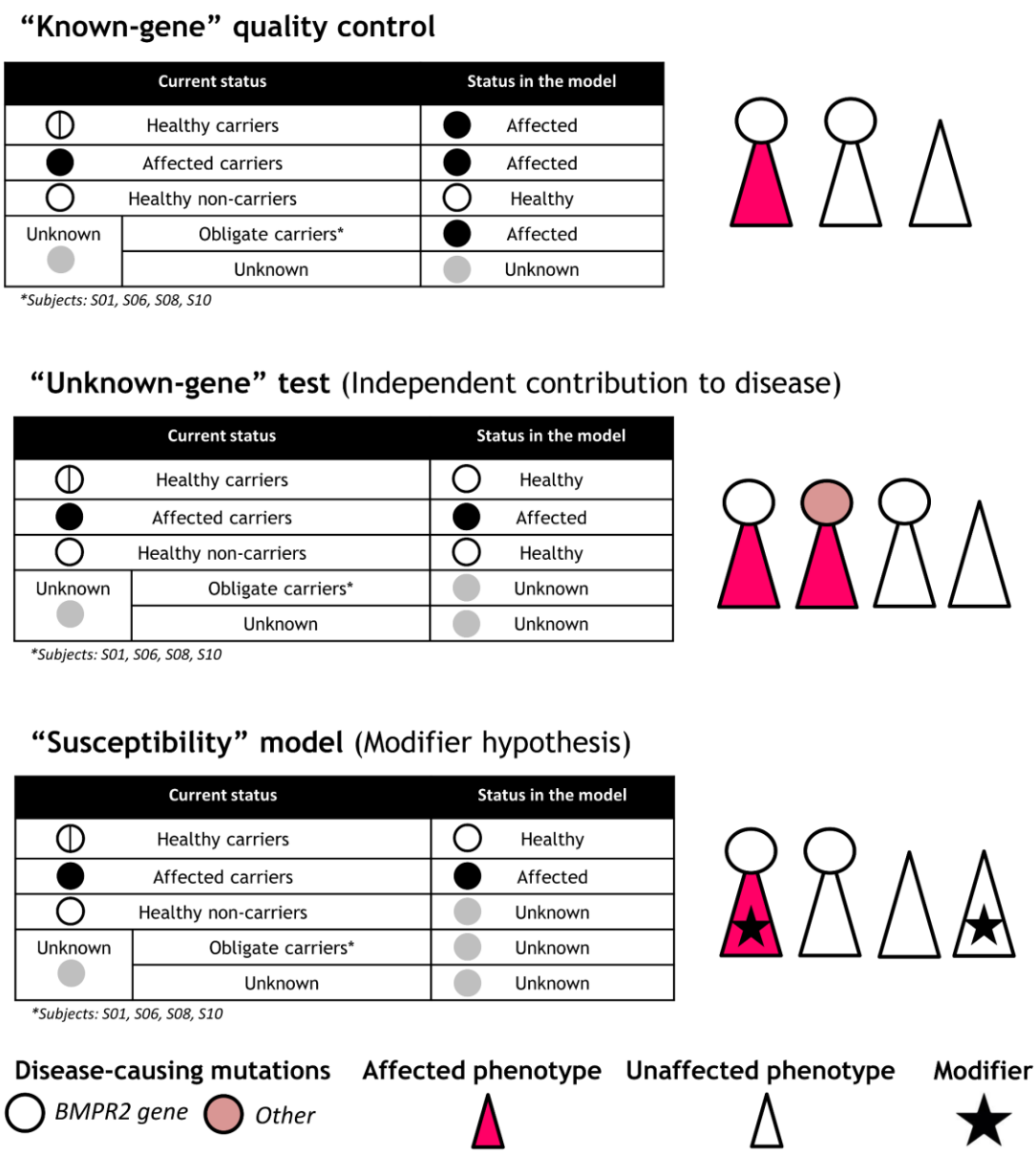

Figure S3. Quality control, unknown gene test and the susceptibility model. As an initial quality control for linkage, we checked the segregation of the *BMPR2* carrier mutation with the *BMPR2* carrier status. Consequently, healthy and affected *BMPR2* mutation carriers were marked as affected in the model. With the unknown gene test, the status of carrier was omitted and only clinical affected individuals were marked as affected. In this test, we also explored the results using different phenocopy rates (0%, 1%, 2%, 5% and 10%) to search for an independent *BMPR2* contribution to HPAH. Finally, the susceptibility model looks for a modifier present in affected carriers and absent in healthy carriers to explain the disease onset in a digenic mechanism. Healthy non-carriers were marked as unknown, as the modifier could be present in these individuals without compromising their clinical status.

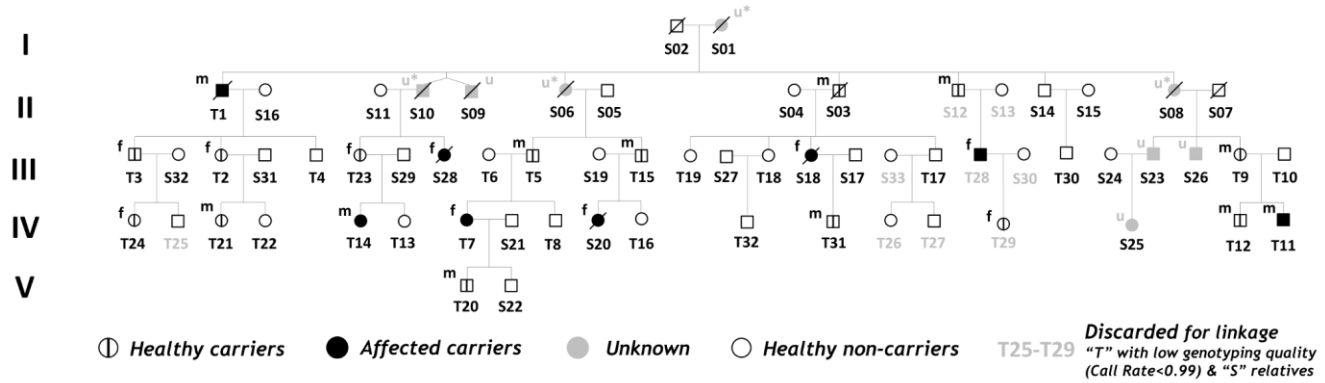

Figure S4. Evaluation of *BMPR2* imprinting as a potential mechanism underlying HPAH reduced penetrance. The parental origin of the *BMPR2* mutation in carrier individuals is described by the letters "f" (father) or "m" (mother). Some individuals are discarded for the analysis as they have an unknown genotype ("u") or they are obligate carriers, but with unknown phenotype ("u*"). Only individuals labeled with black letters "f" or "m" are considered for imprinting evaluation. It is assumed that all carrier individuals are heterozygous. Regarding the parental origin, two heterozygous are possible: the ones with paternal origin (f: mutated/wild-type) and the ones with maternal origin (m: wild-type/mutated). The comparison of the penetrance of each heterozygous, P(f)=5/10=50% and P(m)=3/12=25% yields a 2-fold difference, which is however not statistically significant in this family.

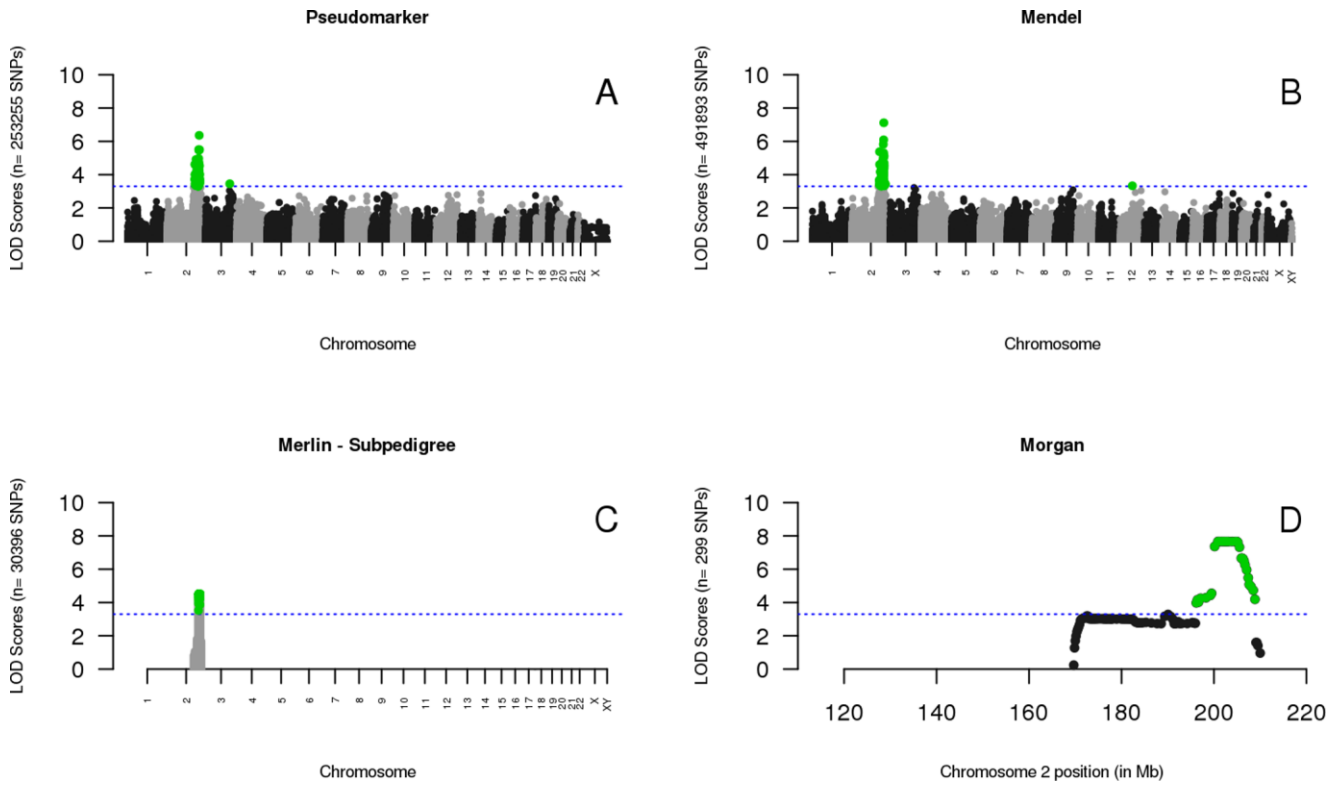

Figure S5. Genome-wide results for the quality control test with the *BMPR2* carrier status. The known gene model was run by four independent genetic linkage programs under dominant mode of inheritance P={0%, 90%, 90%} and rare allele frequency (d=10^{-3}). (A) Pseudomarker: Two-point analysis. (B) Mendel: Two-point analysis. (C) Merlin: Parametric multi-point linkage analysis. (D) Morgan: Multi-point linkage analysis on chromosome 2 window (GRChr37/hg19: 120-220 Mb).

Two-point linkage analysis identified a large region in chromosome 2 (in green) that segregates with the disease (max LOD, 6.36 and 7.12, in Pseudomarker and Mendel, respectively). The trimmed version of the pedigree (Figure S2) also showed significant linkage at this region in Merlin (max LOD = 4.507). The inclusion of the whole pedigree in Morgan boosted the linkage signal up to 7.67.

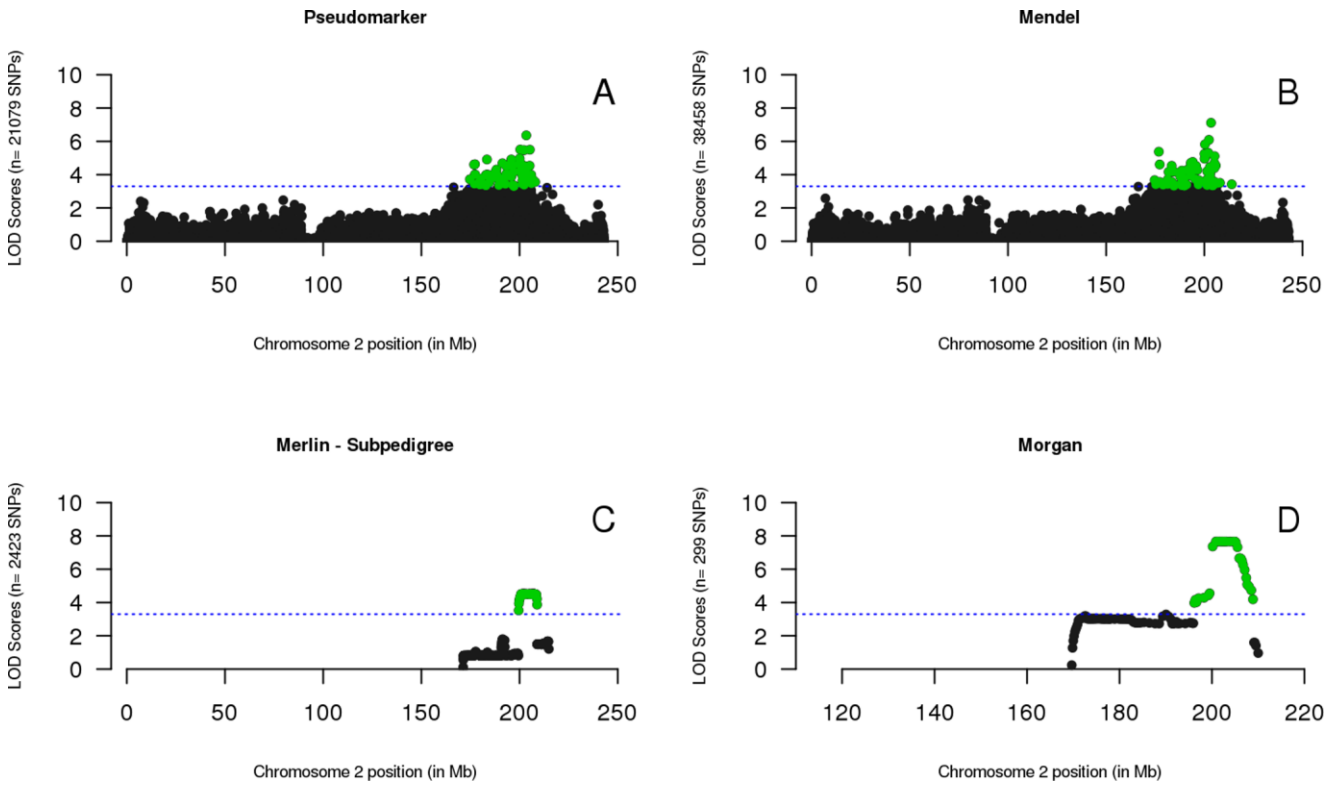

Figure S6. Chromosome 2 results for the quality control test with the *BMPR2* carrier status. The known gene model was run by four independent genetic linkage programs under dominant mode of inheritance P={0%, 90%, 90%} and rare allele frequency (d=10^{-3}). A) Pseudomarker, B) Mendel, C) Merlin (sub-pedigree) and D) Morgan (whole pedigree). Pseudomarker and Mendel identified a >30Mb region around *BMPR2* to segregate together with the disease. Multi-point reported linkage on a shorter region of 10 Mb, in both Morgan (196.1-208.9 Mb) and Merlin (199.5-209 Mb).

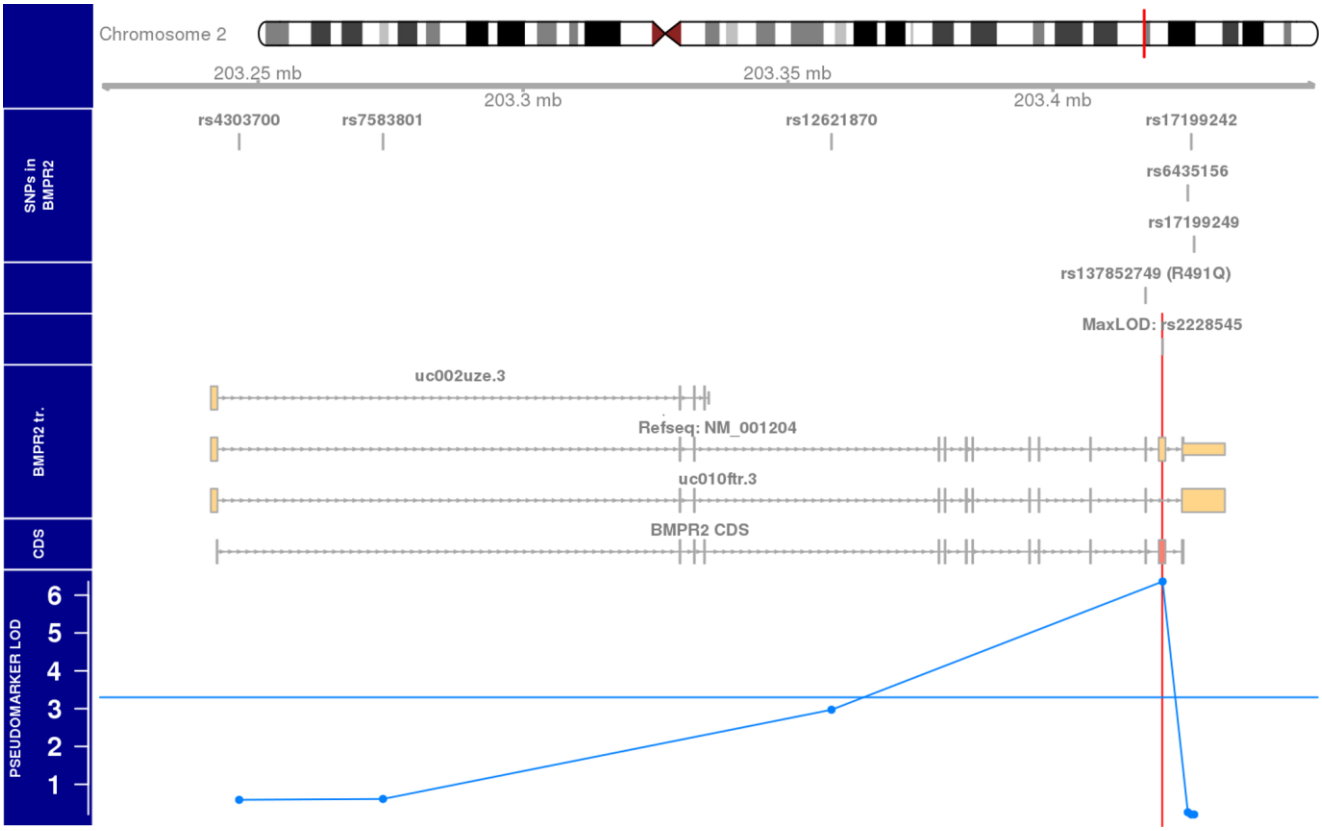

Figure S7. Pseudomarker LOD scores for the quality control test with the *BMPR2* carrier status in chromosome 2 (GRChr37: 203,2-203,5 Mb). Tracks (from top to bottom): 1-SNPs considered for the quality control test in Pseudomarker. 2-p.Arg491Gln variant (rs137852749, exon 11), respect to whom the carrier status is genetically defined. 3-The maximum LOD (6.36) was observed at variant rs2228548 (exon 12). 4-The three *BMPR2* transcripts annotated in UCSC (nomenclature: UCSC and RefSeq ID). 5- *BMPR2* coding DNA sequence (CDS). 6- LOD scores profile. These LOD scores are obtained under a known-gene model (d=10^{-3}, P={0%, 90%, 90%}). Horizontal line: threshold for significant linkage (LOD=3.3). Although the pathogenic variant was not available in the genotyping chip, we observed that the maximum LOD is observed in rs2228545 (exon 12), only 3,215 bp downstream from it.

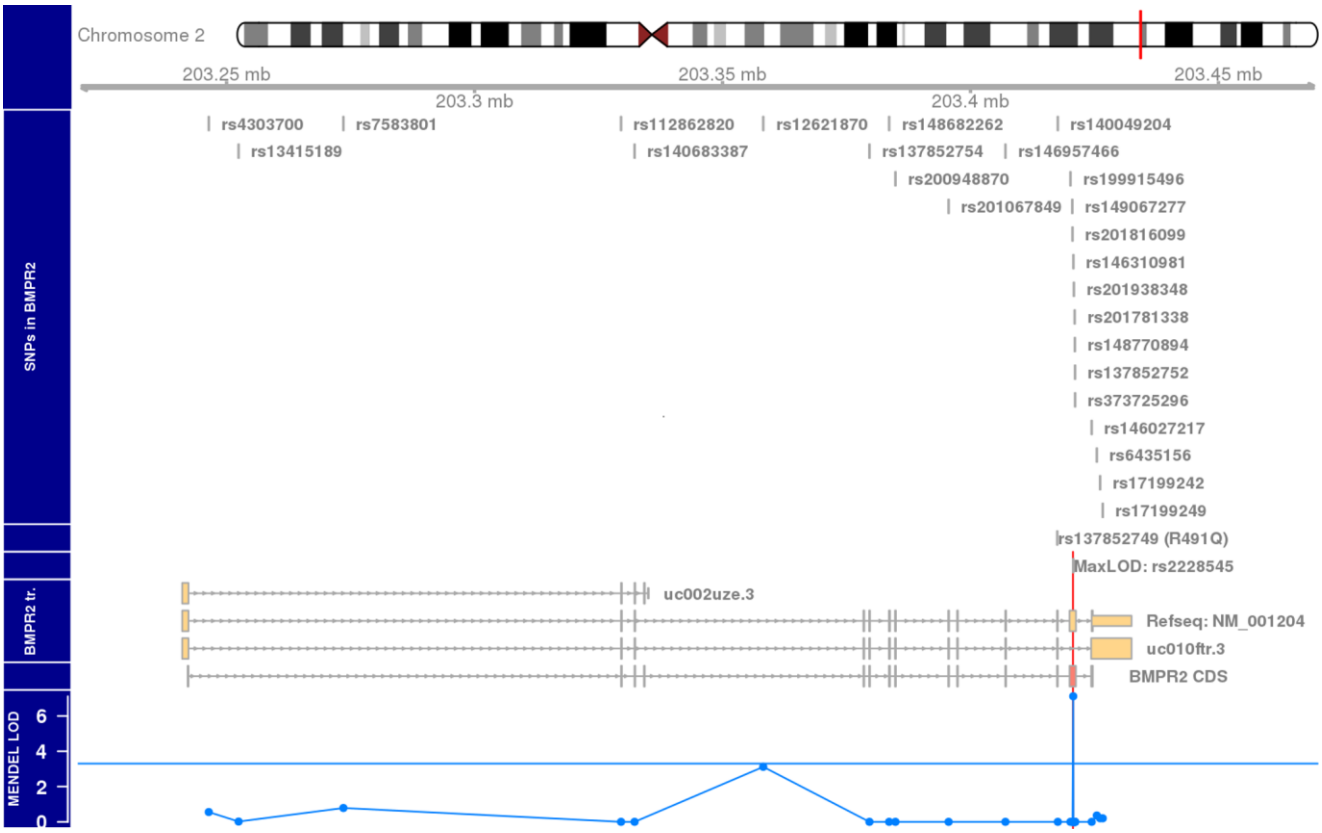

Figure S8. Mendel LOD scores for the quality control test with the *BMPR2* carrier status in chromosome 2 (GRChr37: 203,2-203,5 Mb). Tracks (from top to bottom): 1-SNPs considered for the quality control test in Mendel. 2- p.Arg491Gln variant (rs137852749, exon 11), respect to whom the carrier status is genetically defined. 3-The maximum LOD (7.12) was observed at variant rs2228548 (exon 12). 4-The three *BMPR2* transcripts annotated in UCSC (nomenclature: UCSC and RefSeq ID). 5- *BMPR2* coding DNA sequence (CDS). 6- These LOD scores are obtained under a known-gene model (d=10^{-3}, P={0%, 90%, 90%}). Horizontal line: threshold for significant linkage (LOD=3.3). Although the rs137852749 variant was not available in the genotyping chip, we observed again that the maximum LOD is observed in rs2228545 (exon 12), only 3,215 bp downstream from it.

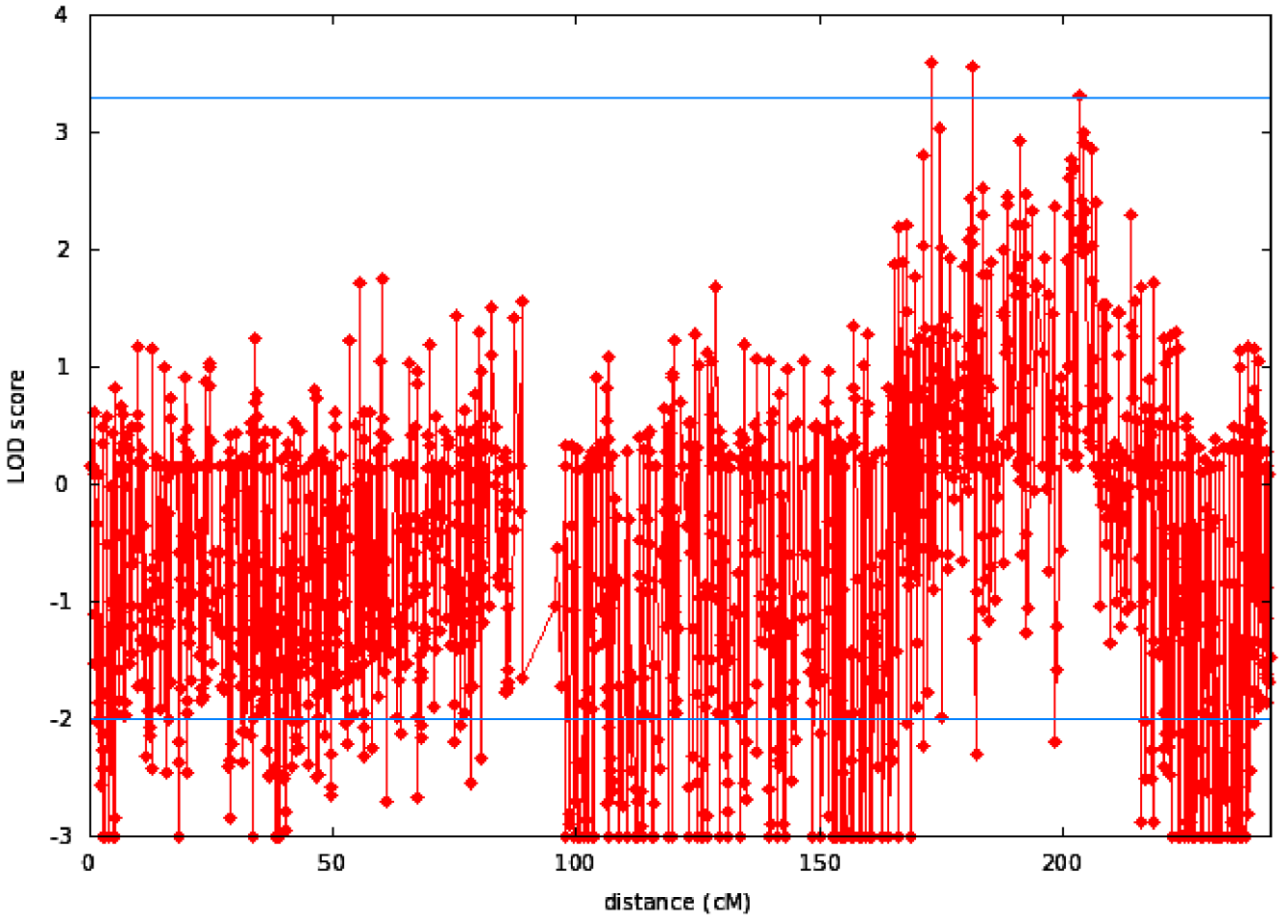

Figure S9. Superlink-Online LOD scores for the quality control test with the *BMPR2* carrier status in chromosome 2. A region with substantially higher LOD scores among 170-205 cM (approximately 170-205 Mb) indicates linkage with the *BMPR2* carrier status using Superlink-Online two-point linkage analysis. Three SNPs (rs4246617, rs16867225, rs12621870) are found to be above the significance threshold (LOD=3.3), being rs12621870 less than 60Kb from rs137852749.

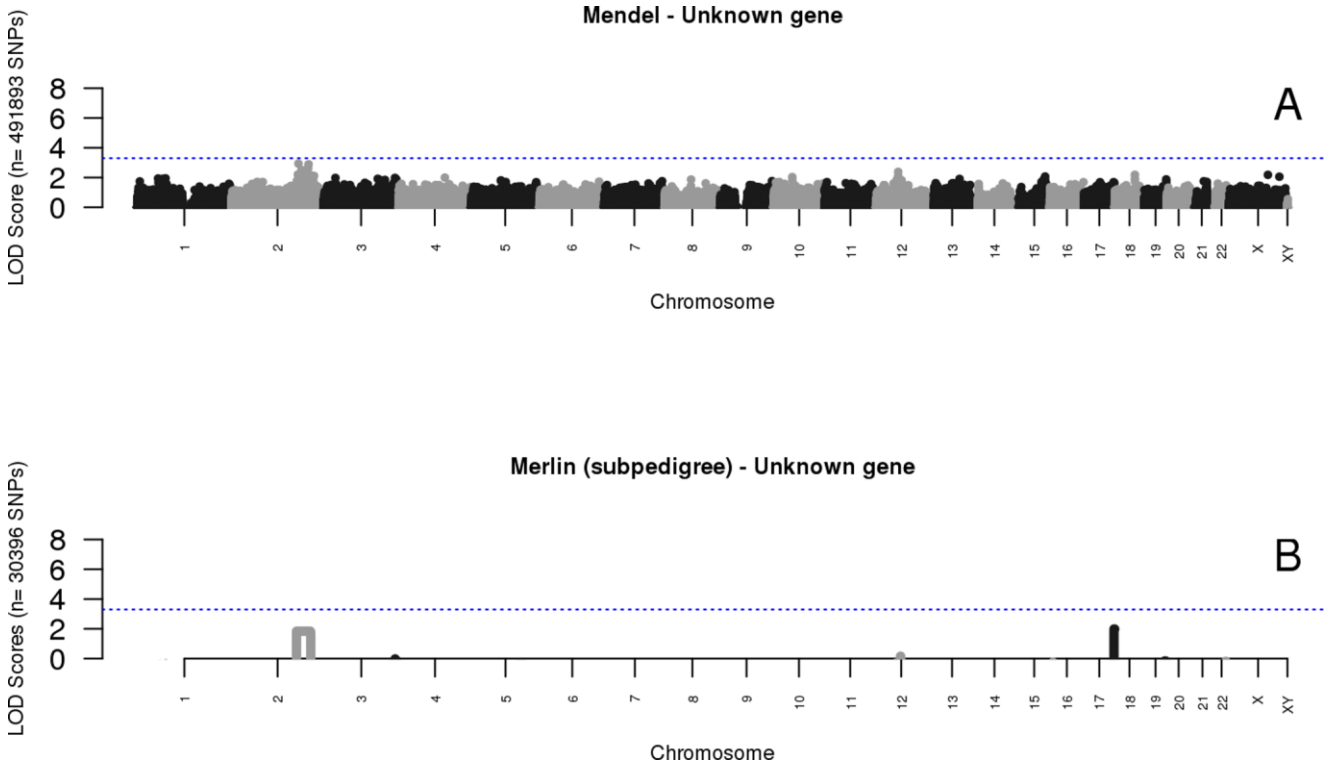

Figure S10. Genome-wide results for the "unknown gene" test without phenocopies. The test was computed by two-point (A, Mendel) and multi-point (B, Merlin) linkage programs under a dominant mode of inheritance, reduced penetrance {0%, 30%, 30%} and rare allele frequency (d=10^{-3}). None of the programs provided signatures of linkage. This shows the difficulty of finding an independent locus of *BMPR2* contributing to HPAH disease with such a low penetrance.

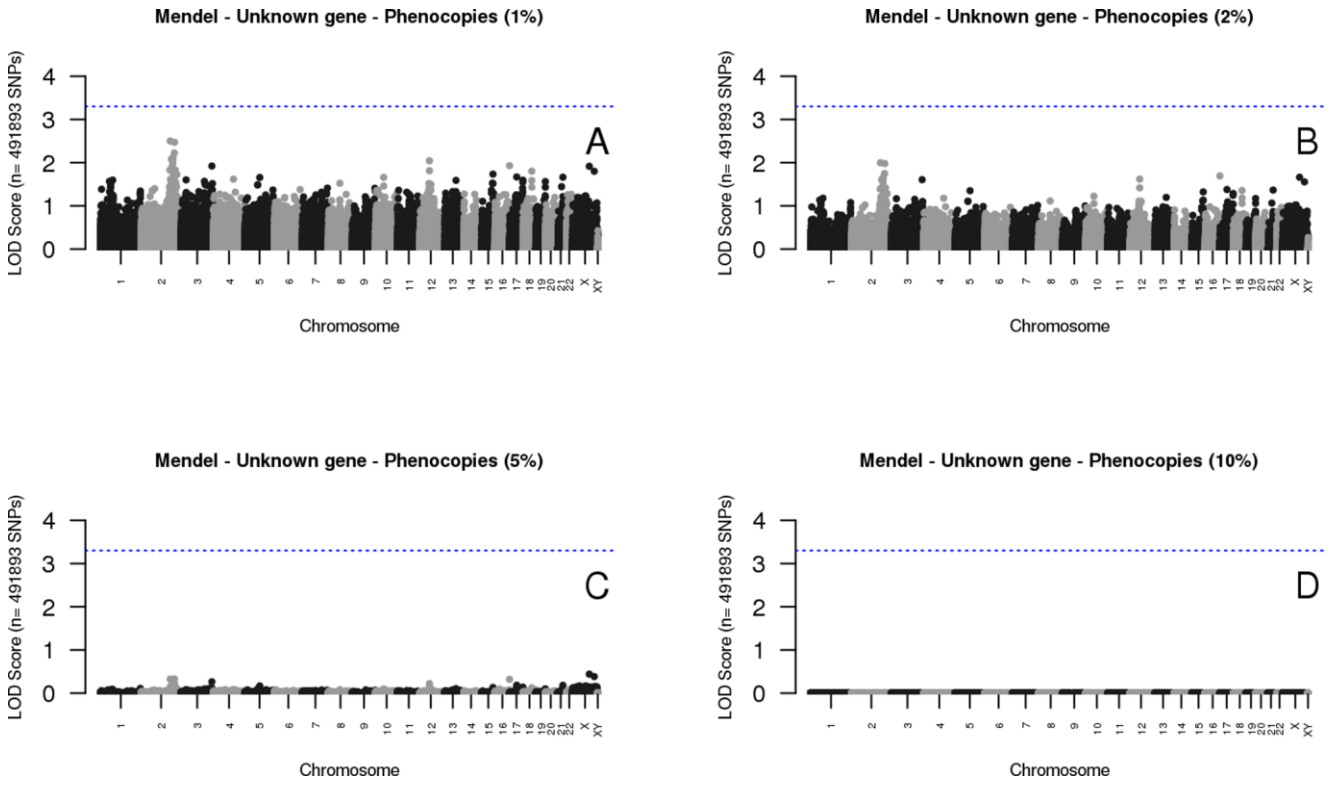

Figure S11. Genome-wide results for the "unknown gene" test with phenocopies. The grid of phenocopies – 1% (A), 2% (B), 5% (C) and 10% (D) – was computed by Mendel two-point linkage. Parameters: Dominant mode of inheritance, reduced penetrance P={phenocopies, 30, 30} and rare allele frequency (d=10^{-3}). None of the phenocopies rates provided signatures of linkage.

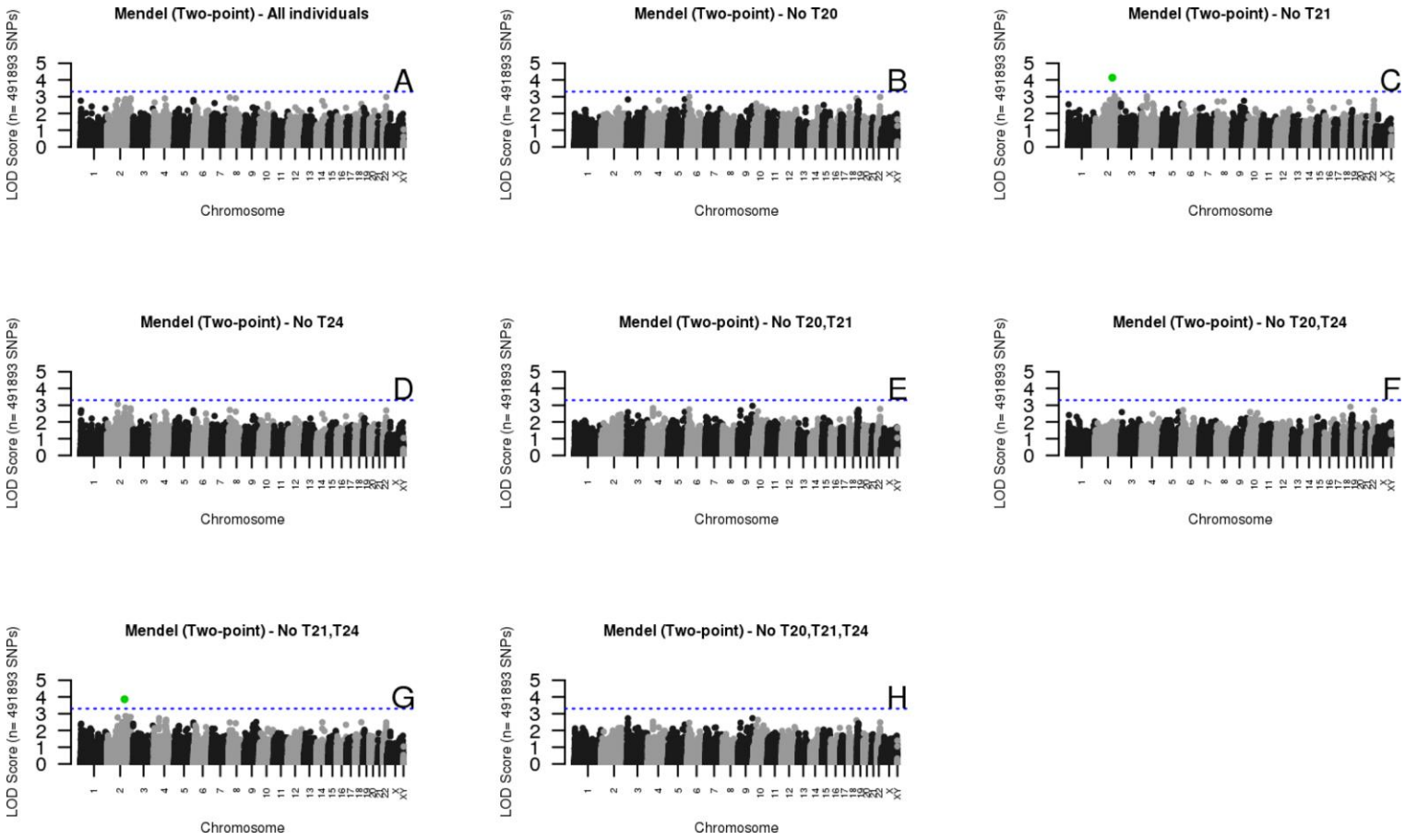

Figure S12. Genome-wide Mendel linkage results for the susceptibility model in 8 different combinations. The above panels show the different combinations of the pedigree removing at a time 1,2 or 3 of the individuals below 10 years (T20,T21,T24). (A): All individuals from the family. (B): T20 removed. (C): T21 removed. (D): T24 removed. (E): T20,T21 removed. (F): T20,T24 removed. (G): T21,T24 removed. (H): T20,T21,T24 removed. Parameters for all models: Common allele frequency (d=0.999), recessive MOI with phenocopies P={2%,2%,100%}. Only the scenario C (no T21, LOD=4.14) and G (no T21,T24; LOD=3.86) provided one SNP with significant linkage signal. In both cases, the signal corresponds to the same SNP (rs17716942, chr2:163260691 Mb, GrCh37/hg19) in chromosome 2.

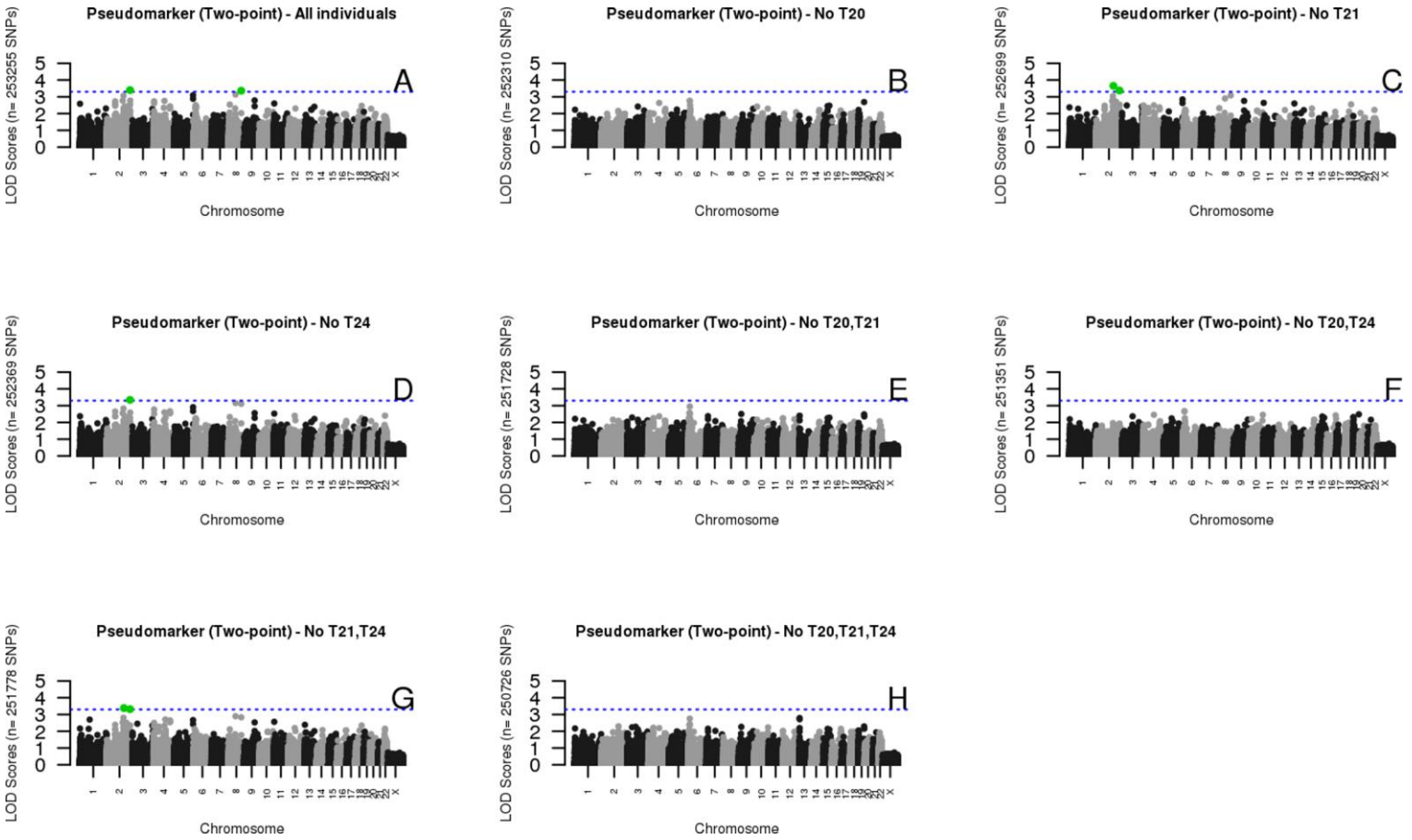

Figure S13. Genome-wide Pseudomarker linkage results for the susceptibility model in 8 different combinations. Four combinations (A, C, D, G) provided linkage signals above LOD>3.3. In almost all cases, they were observed in chromosome 2, with the same two SNPs being identified (rs17716942, chr2:163260691 Mb; rs6436140, chr2:220200242 Mb). The maximum LOD score is observed for SNP rs17716942 in combinations C (No T21: LOD=3.65) and G (No T21&T24: LOD=3.38). This signal reproduced the results observed in Mendel-Two-point. In addition, another SNP (rs6436140) showed signal in combinations A (LOD=3.4), C (LOD=3.37), D (LOD=3.34) and G (LOD=3.31), but in all cases it was weaker than rs17716942.

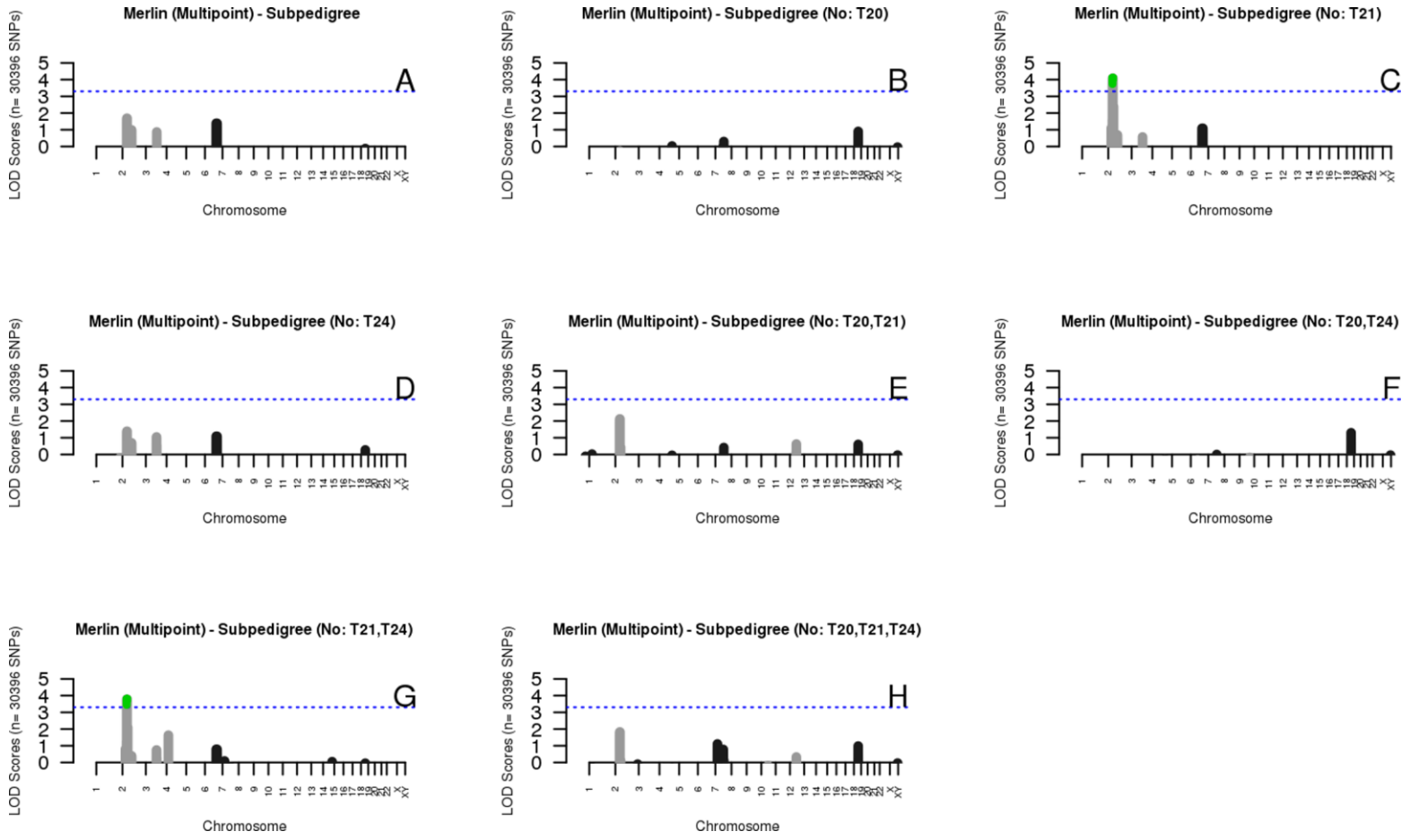

Figure S14. Genome-wide Merlin linkage results for the susceptibility model in 8 different combinations. Only two combinations (C -no T21- and G -no T21,T24) provided significant linkage signals (LOD>3.3). In both cases, the highlighted region was the same (chr2:161503223-165107298) and comprised 27 SNPs. The maximum LOD was higher in combination C (max LOD=4.090) than G (max LOD=3.790).

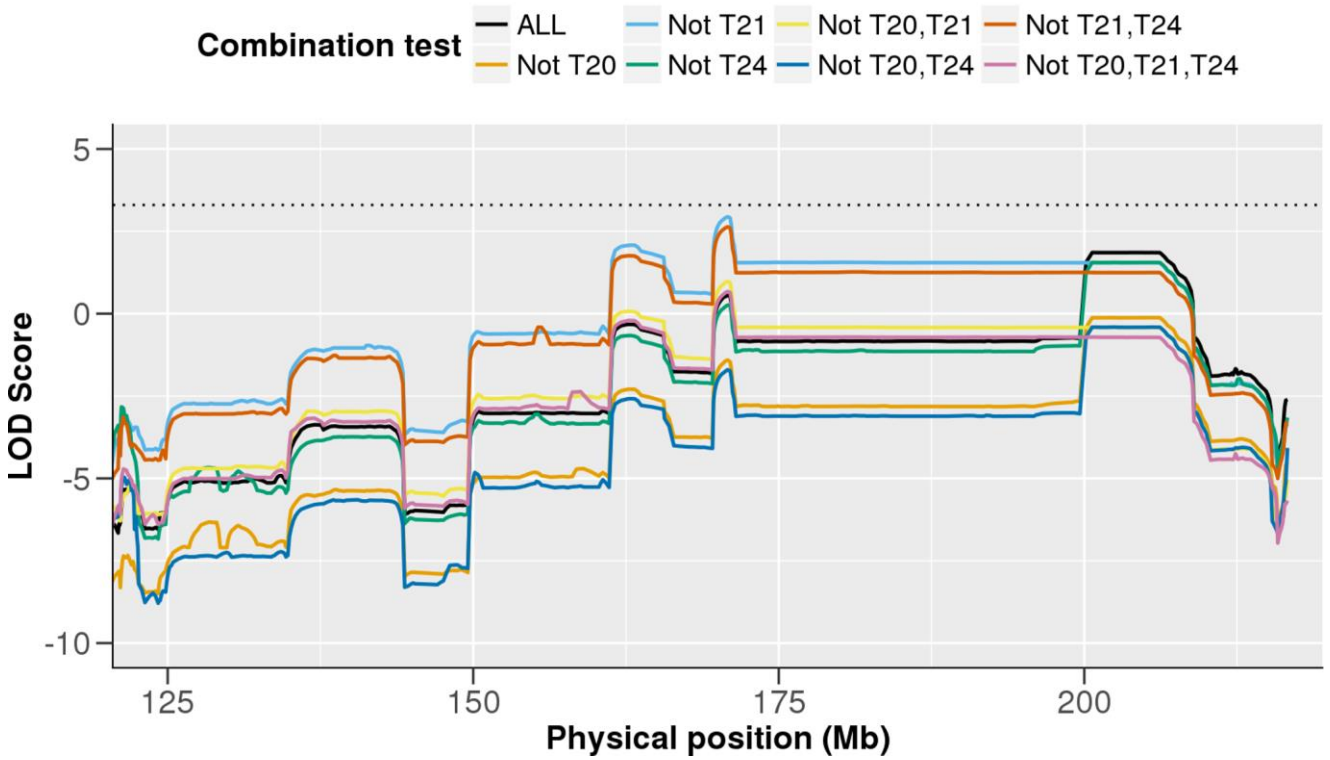

Figure S15. Morgan linkage results in a chromosome 2 window (125-215 Mb) for the susceptibility model in 8 different combinations. None of the combinations provide significant results, although peaks of suggestive linkage (2q24.2 and within 2q24.3-q31.1) are observed for scenarios without T21 (LOD=2.94) and without T21,T24 (LOD=2.6345).

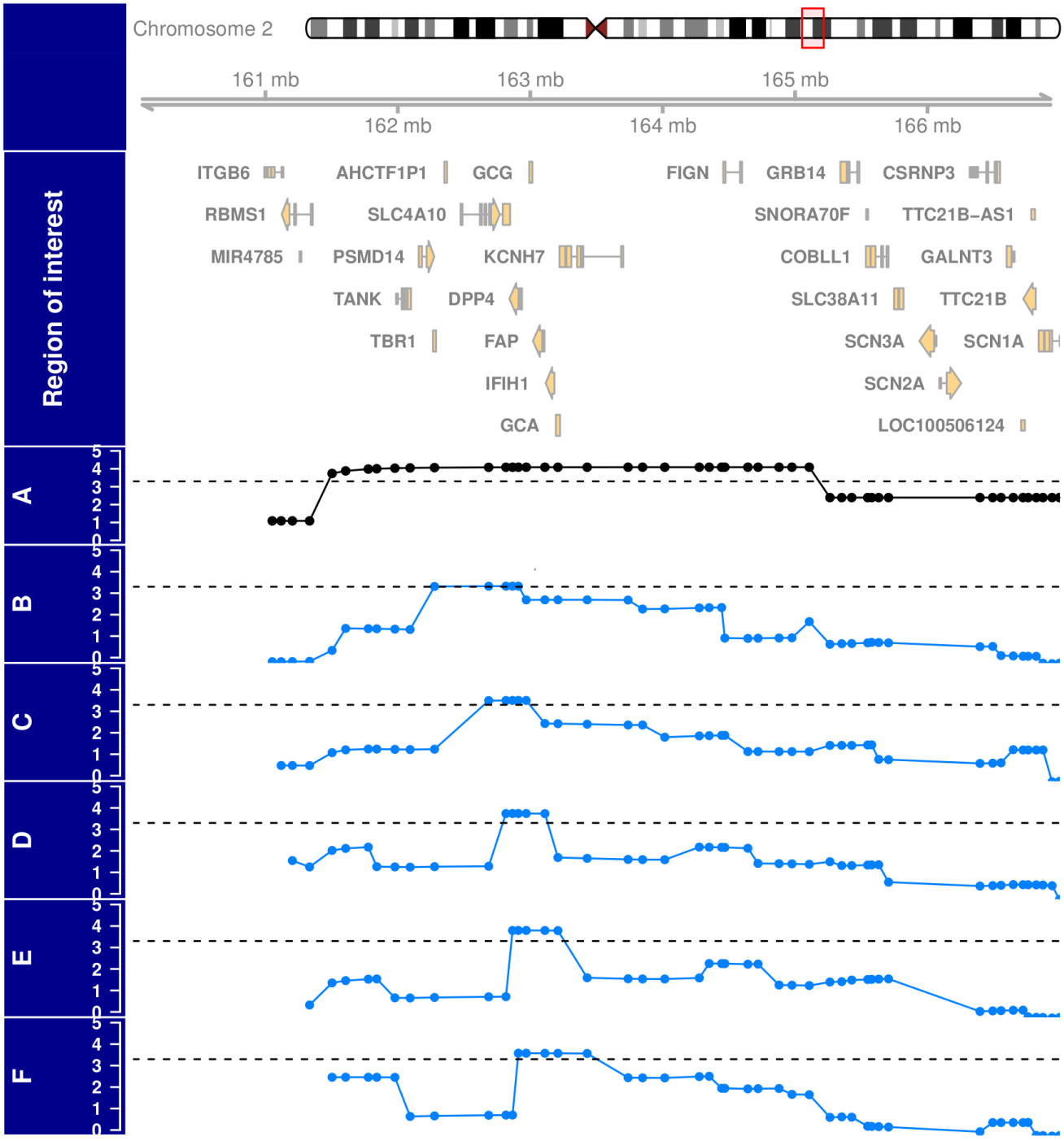

Figure S16. Genome-wide LOD score profiles of the chromosome region (chr2:161-167 Mb) by Merlin and Superlink multi-point, both excluding individual T21. (A) Merlin multi-point linkage under the pruned pedigree version (See Figure S2). (B-F) Superlink multi-point linkage using the whole pedigree and a non-overlapping window size of 5 SNPs. The sliding window starts at each of the 5 possible SNPs (B) rs3111397; (C) rs10170600; (D) rs7425274; (E) rs13020444 and (F) rs6432641.

Query SNP: rs13002573 and variants with r^2 >= 0.8

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
2	164032624	0.98	0.99	rs2892895	G	A	0.05	0.18	0.40	0.22						5 altered motifs				AC092684.1	
2	164035035	0.98	0.99	rs35454965	C	T	0.05	0.18	0.40	0.22										AC092684.1	
2	164036082	0.98	0.99	rs4667725	T	C	0.05	0.18	0.40	0.22						Hoxa10,Pou2f2,Pou5f1				AC092684.1	
2	164037150	0.98	0.99	rs6724342	C	T	0.05	0.18	0.40	0.22						EBF				AC092684.1	
2	164040431	0.98	0.99	rs7588932	C	T	0.04	0.18	0.42	0.22						4 altered motifs				AC092684.1	
2	164040522	0.98	0.99	rs7602967	T	C	0.05	0.18	0.40	0.22										AC092684.1	
2	164041623	0.99	1	rs7593109	G	T	0.04	0.18	0.40	0.22						Pax-5,RFX5				AC092684.1	
2	164043173	0.97	0.99	rs16849211	C	T	0.05	0.19	0.40	0.22						Evi-1,STAT				AC092684.1	
2	164044084	0.98	0.99	rs16849215	C	G	0.05	0.18	0.40	0.22						11 altered motifs				AC092684.1	
2	164044853	0.98	0.99	rs10930112	T	C	0.05	0.18	0.40	0.22				5 tissues	STAT3					AC092684.1	
2	164046371	0.94	0.99	rs13004584	C	T	0.05	0.18	0.40	0.21				BRN	STAT3					AC092684.1	
2	164046633	0.94	0.99	rs13005313	C	T	0.06	0.18	0.40	0.21						4 altered motifs				AC092684.1	
2	164050310	0.98	0.99	rs16849225	C	T	0.08	0.19	0.44	0.22							1 hit			AC092684.1	
2	164051484	0.99	1	rs16849229	T	A	0.09	0.18	0.40	0.22						7 altered motifs				AC092684.1	
2	164051542	0.99	1	rs10497247	G	A	0.09	0.18	0.40	0.22						33 altered motifs				AC092684.1	
2	164052054	0.99	1	rs67453060	C	T	0.05	0.18	0.40	0.22			ESDR, BRST, LNG	KID,LNG		EWSR1-FLI1,STAT				AC092684.1	
2	164053763	0.99	1	rs4667728	C	T	0.05	0.18	0.40	0.22			FAT, STRM, LNG			6 altered motifs				AC092684.1	
2	164055609	0.98	0.99	rs13031230	C	T	0.05	0.18	0.40	0.22			ESDR, BRN			p300				AC092684.1	
2	164058698	1	1	rs13002573	A	G	0.13	0.19	0.40	0.22						SIX5	1 hit	1 hit		AC092684.1	
2	164058769	1	1	rs13002621	A	G	0.13	0.19	0.40	0.22						HMG-IY,Hoxb9,Hoxc10		1 hit		AC092684.1	
2	164062362	0.98	0.99	rs10930113	G	A	0.20	0.19	0.40	0.22						4 altered motifs		1 hit		AC092684.1	
2	164065260	0.96	0.98	rs35303331	A	G	0.08	0.18	0.40	0.22						6 altered motifs				AC092684.1	
2	164067822	0.95	0.98	rs35930173	G	A	0.22	0.20	0.44	0.23			4 tissues			GR				AC092684.1	
2	164068063	0.95	0.98	rs72874178	G	A	0.05	0.17	0.40	0.22			4 tissues							AC092684.1	
2	164071196	0.95	0.98	rs10930114	C	A	0.21	0.20	0.44	0.23						PRDM1				AC092684.1	

Figure S17. HaploReg functional genomics data for the three candidate regulatory SNPs (CRS) associated with systolic blood pressure and pulse pressure measurement (rs13002573, rs16849211, rs16849225). The information is displayed as an LD block. Remarkably, the SNPs from the LD block are predicted to have signs of conservation, alter several motifs and overlap histone marks in fetal lung. Also, the minor allele frequency (MAF) ranges among 0.21-0.23 in European population, consistently enough with the common disease allele frequency proposed in the susceptibility linkage model. Table produced by: HaploReg v4.1.

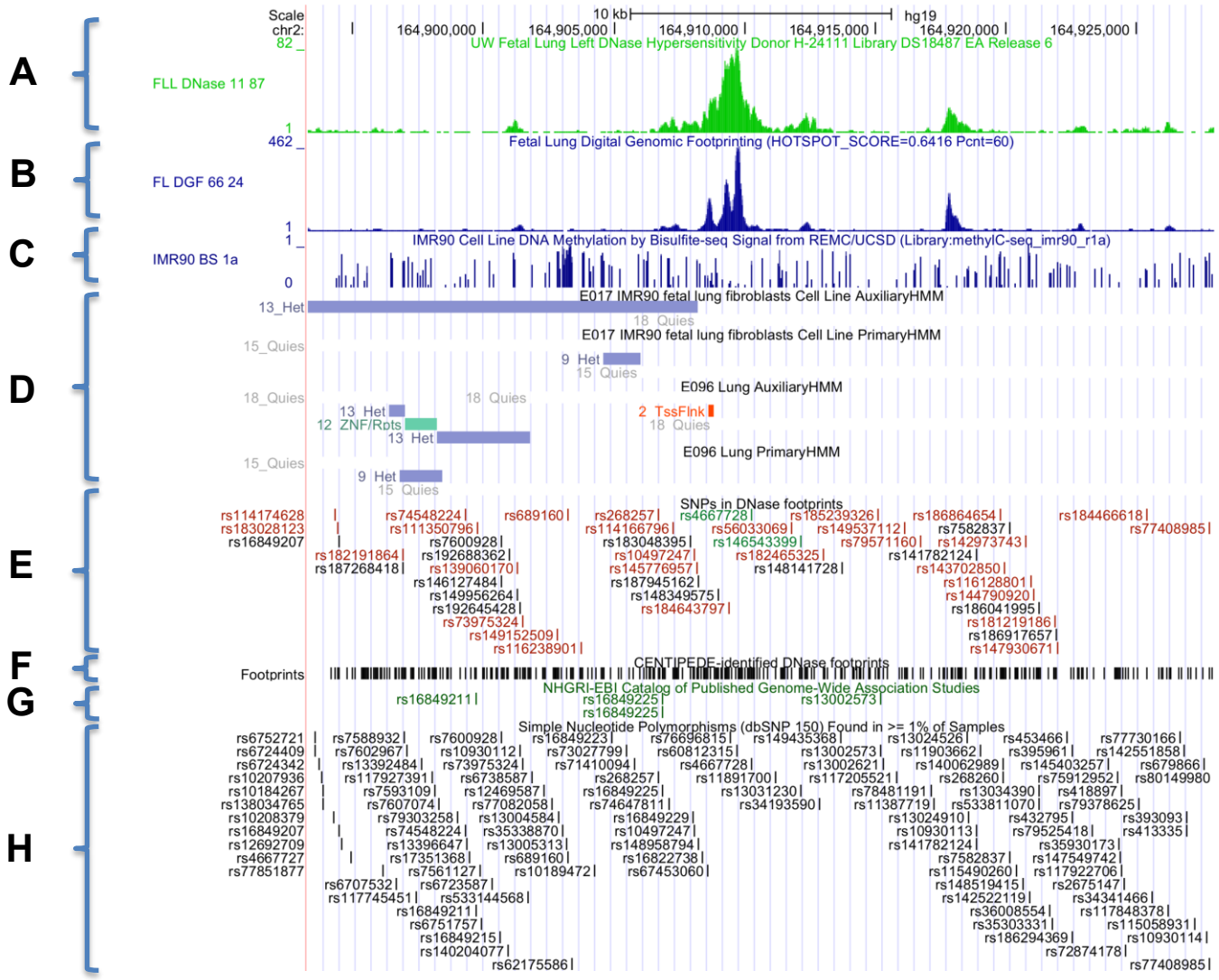

Figure S18. Regulatory and functional elements located within the LD block of the three CRS that are *FIGN* cis-eQTLs. CENTIPEDE integrates ENCODE and Roadmap Epigenomics Mapping Consortium (REMC) DNaseI data (track A, F) with sequence-based motifs from

TRANSFAC and JASPAR to predict the impact of variation on protein DNA-binding. CENTIPEDE generates a catalog of variants and overlapping regulatory DNA-binding sites that are factor and tissue-specific. We assessed the impact of common genetic variants in the region bounded by a linkage disequilibrium block with data from the IMR-90 cell-line derived from fetal lung. We found evidence for open chromatin regions calling three peaks of genomic footprints (track B). All SNPs annotated on CENTIPEDE footprints were evaluated using a logistic sequence hyperprior model (track E). The SNP color indicates its impact on protein binding: silent footprint-SNPs in black; effect-SNPs that alter the prior odds of binding >= 20-fold, in red; and switch-SNPs that alter and flip the prior odds of binding, in green. For illustration purposes, we also include all common variation from the dbSNP release 150 (track H), GWAS SNPs (track G), Bisulfite-Seq methylation signals from the IMR-90 cell-line (track C) and REMC chromatin state segmentation from the same IMR-90 cell-line and lung (track D). The chromatin data shows changes between quiescent, heterochromatic, zinc finger and flanking promoter states across the LD block on lung auxiliary HMM.

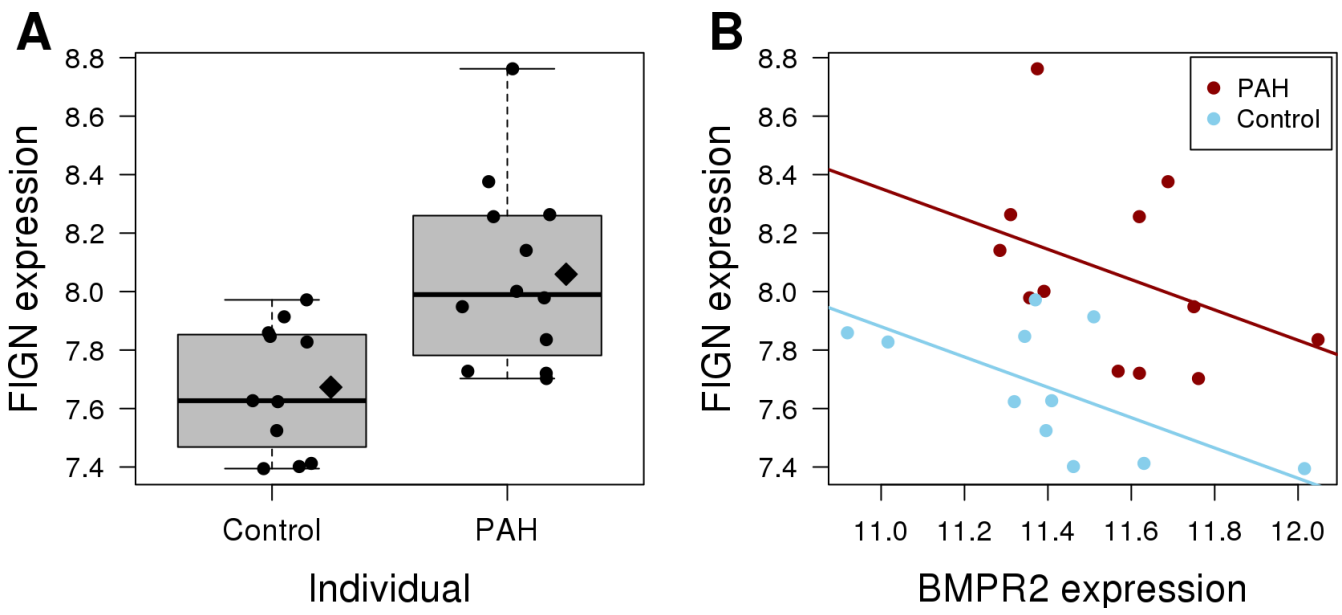

Figure S19. Gene expression of *FIGN* in PAH patients and controls. (A) RNA expression of the *FIGN* gene in lung tissue from PAH patients and controls. (B) Co-expression of *FIGN* and *BMPR2*. Solid lines show the fit to the expression data of a linear model of *FIGN* expression with respect to *BMPR2,* adjusted for PAH status, with one slope and a different intercept term for PAH patients and controls, highlighted with different colors.

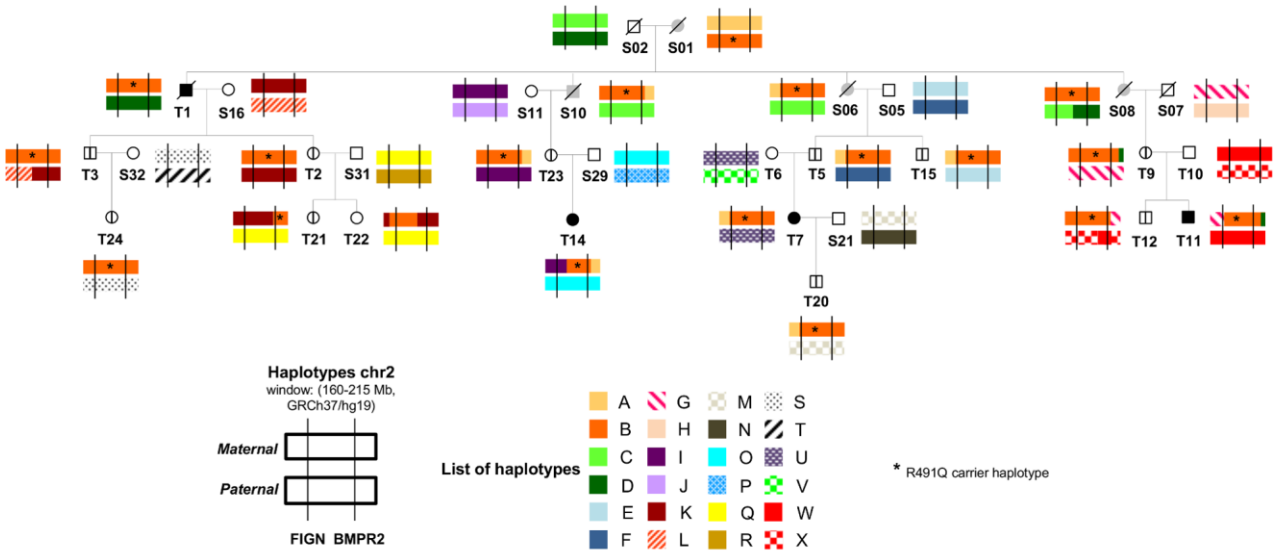

Figure S20. Haplotypes between the putative modifier and the disease-causing gene. Predicted haplotypes for each individual at the physical positions 160-215 Mb of chromosome 2. From left to right, the two vertical bars indicate the position of *FIGN* and *BMPR2*. For each non-founder individual the upper haplotype is maternal and the lower one is paternal. According to the segregation pattern among carriers, haplotype B carries the *BMPR2* c.1472G>A (p.Arg491Gln) pathogenic mutation.

Supplemental Tables

Table S1. Table with the years free of PAH for the 22 *BMPR2* carriers. Three carriers (S03,S18,S12) have their age inferred based on the age of their oldest child plus 18 years. Accordingly, in those cases it should be interpreted as their minimum possible age.

Individual	Generation	Years free of PAH	Current Phenotype	Data source
T1	G2	75.4	Affected carrier	Available
S03	G2	83.0	Healthy carrier	Inferred
S12	G2	70.0	Healthy carrier	Inferred
S28	G3	46.7	Affected carrier	Available
T28	G3	52.5	Affected carrier	Available
S18	G3	42.0	Affected carrier	Inferred
T15	G3	73.3	Healthy carrier	Available
T5	G3	71.9	Healthy carrier	Available
T9	G3	67.9	Healthy carrier	Available
T23	G3	55.2	Healthy carrier	Available
T2	G3	51.5	Healthy carrier	Available
T3	G3	49.3	Healthy carrier	Available
S20	G4	17.0	Affected carrier	Available
T7	G4	35.5	Affected carrier	Available
T11	G4	14.6	Affected carrier	Available
T14	G4	5.5	Affected carrier	Available
T31	G4	24.2	Healthy carrier	Available
T12	G4	41.1	Healthy carrier	Available
T29	G4	24.1	Healthy carrier	Available
T21	G4	9.7	Healthy carrier	Available
T24	G4	7.0	Healthy carrier	Available
T20	G5	6.6	Healthy carrier	Available

Table S2. Functional enrichment analysis on the candidate region defined by Merlin results. EFO terms enriched by 529 SNPs from the region with LOD>3.3 (chr2:161503223-165107298). These SNPs include variants reported by the GWAS Catalog and SNPs in linkage disequilibrium with them, according to 1000 Genomes Project data. EFO terms were selected using one-tailed Fisher's exact tests with Holm adjusted P-value < 10^{-3}, minimum EFO term size > 5, minimum number of enriching SNPs > 5 and are shown below ordered by odds ratio. EFO terms associated with cardiorespiratory traits are highlighted in boldface.

Num	EFO term	P-value	Adjusted P-value	Odds Ratio	Expected Count	Count	Size	Term	Parental term cluster
1	EFO:0007967	2.85e-321	5.34e-318	558.10	0.7	153	330	**blood osmolality measurement**	blood osmolality measurement
2	EFO:0003888	2.49e-163	4.67e-160	39.45	5.5	148	2517	attention deficit hyperactivity disorder	nervous system disease
3	EFO:0005200	4.55e-13	8.44e-10	38.99	0.3	10	130	antiphospholipid antibody measurement	antiphospholipid antibody measurement
4	EFO:0000401	2.08e-49	3.89e-46	32.17	1.6	45	745	diabetic nephropathy	kidney disease, metabolic disease
5	EFO:0000289	1.08e-143	2.03e-140	28.44	7.4	148	3422	bipolar disorder	nervous system disease
6	EFO:0006918	1.11e-13	2.05e-10	22.64	0.6	13	283	female fertility	female fertility
7	EFO:0006923	1.11e-13	2.05e-10	22.64	0.6	13	283	fertility measurement	fertility measurement
8	EFO:0003940	6.86e-21	1.28e-17	22.03	1.0	21	476	physical activity	physical activity
9	EFO:0004247	4.72e-124	8.84e-121	20.43	10.2	148	4680	mood disorder	nervous system disease
10	EFO:0003925	8.89e-35	1.66e-31	15.61	3.1	43	1412	cognition	mental process
11	EFO:0004323	3.11e-31	5.79e-28	12.68	3.8	43	1727	mental process	mental process
12	EFO:0006335	1.41e-40	2.63e-37	11.67	5.9	60	2694	**systolic blood pressure**	vital signs
13	EFO:0003086	9.16e-39	1.71e-35	10.27	6.9	62	3162	kidney disease	kidney disease
14	EFO:0004784	2.90e-40	5.41e-37	9.34	8.5	69	3908	self reported educational attainment	self reported educational attainment
15	EFO:0003884	1.03e-10	1.90e-07	8.36	2.1	17	978	chronic kidney disease	kidney disease
16	EFO:0005763	5.03e-12	9.32e-09	7.53	2.9	21	1348	**pulse pressure measurement**	pulse pressure measurement

Num	EFO term	P-value	Adjusted P-value	Odds Ratio	Expected Count	Count	Size	Term	Parental term cluster
17	EFO:0006995	2.41e-08	4.47e-05	7.25	2.0	14	921	**response to diisocyanate**	response to diisocyanate
18	GO:0097332	3.26e-12	6.05e-09	7.25	3.2	22	1467	response to antipsychotic drug	response to antipsychotic drug
19	EFO:0004325	2.46e-28	4.58e-25	6.79	9.9	60	4553	**blood pressure**	vital signs
20	EFO:0004303	1.89e-23	3.52e-20	5.39	12.4	60	5689	vital signs	vital signs
21	EFO:0000677	1.48e-45	2.76e-42	4.57	46.1	160	21212	mental or behavioural disorder	nervous system disease
22	EFO:0005774	5.26e-32	9.80e-29	3.42	59.8	160	27507	brain disease	nervous system disease
23	EFO:0000589	6.95e-11	1.29e-07	3.21	15.3	46	7037	metabolic disease	metabolic disease
24	EFO:0000618	1.24e-15	2.30e-12	2.22	87.1	161	40092	nervous system disease	nervous system disease

Table S3. Functional enrichment analysis on the *FIGN* gene. EFO terms enriched by 15 SNPs within the boundaries of the *FIGN* gene (chr2:164464118-164592513). These SNPs include variants reported by the GWAS Catalog and SNPs in linkage disequilibrium with them, according to 1000 Genomes Project data. EFO terms were selected using one-tailed Fisher's exact tests with Holm adjusted P-value < 10^{-3}, minimum EFO term size > 5, minimum number of enriching SNPs > 5 and are shown below ordered by odds ratio. No EFO terms associated with cardiorespiratory traits were enriched by the selected SNPs.

Num	EFO term	P-value	Adjusted P-value	Odds Ratio	Expected Count	Count	Size	Term	Parental term cluster
1	EFO:0004729	2.05e-10	3.84e-07	214.6	0.04	5	571	vitamin measurement	vitamin measurement
2	EFO:0004318	2.42e-14	4.54e-11	99.6	0.30	10	4802	smoking behavior	behavior
3	GO:0007610	7.15e-13	1.34e-09	70.1	0.42	10	6760	behavior	behavior
4	EFO:0004340	6.69e-11	1.25e-07	43.5	0.66	10	10721	body mass index	anthropometric measurement
5	EFO:0004324	6.81e-08	1.27e-04	20.2	1.35	10	21907	body weights and measures	anthropometric measurement
6	EFO:0004302	1.13e-07	2.12e-04	19.1	1.42	10	23108	anthropometric measurement	anthropometric measurement

Table S4. Functional enrichment analysis on the region spanning the lincRNA ENSG00000237844. EFO terms enriched by 76 SNPs within the boundaries of the lincRNA ENSG00000237844 (chr2:164606083-165208733). These SNPs include variants reported by the GWAS Catalog and SNPs in linkage disequilibrium with them, according to 1000 Genomes Project data. EFO terms were selected using one-tailed Fisher's exact tests with Holm adjusted P-value < 10^{-3}, minimum EFO term size > 5, minimum number of enriching SNPs > 5 and are shown below ordered by odds ratio. EFO terms associated with cardiorespiratory traits are highlighted in boldface.

Num	EFO term	P-value	Adjusted P-value	Odds Ratio	Expected Count	Count	Size	Term	Parental term cluster
1	EFO:0006335	2.06e-102	3.86e-99	342.76	0.8	60	2694	**systolic blood pressure**	vital signs
2	EFO:0004325	1.13e-88	2.11e-85	199.39	1.4	60	4553	**blood pressure**	vital signs
3	EFO:0004303	7.21e-83	1.35e-79	158.40	1.8	60	5689	vital signs	vital signs
4	EFO:0005763	7.58e-30	1.42e-26	69.65	0.4	21	1348	**pulse pressure measurement**	cardiovascular measurement
5	EFO:0006995	6.16e-20	1.15e-16	60.37	0.3	14	921	**response to diisocyanate**	response to diisocyanate
6	EFO:0000270	1.35e-11	2.52e-08	14.32	1.2	14	3793	**asthma**	bronchial disease
7	EFO:1002018	2.14e-11	4.00e-08	13.81	1.2	14	3929	**bronchial disease**	bronchial disease
8	EFO:0005278	6.92e-11	1.29e-07	7.39	3.7	21	11973	**cardiovascular disease biomarker measurement**	cardiovascular measurement
9	EFO:0004298	5.48e-10	1.02e-06	6.55	4.2	21	13418	**cardiovascular measurement**	cardiovascular measurement