The American Journal of Human Genetics, Volume *108*

Supplemental information

A computational approach for detecting physiological homogeneity in the midst of genetic heterogeneity

Peng Zhang, Aurélie Cobat, Yoon-Seung Lee, Yiming Wu, Cigdem Sevim Bayrak, Clémentine Boccon-Gibod, Daniela Matuozzo, Lazaro Lorenzo, Aayushee Jain, Soraya Boucherit, Louis Vallée, Burkhard Stüve, Stéphane Chabrier, Jean-Laurent Casanova, Laurent Abel, Shen-Ying Zhang, and Yuval Itan

Supplemental Figures

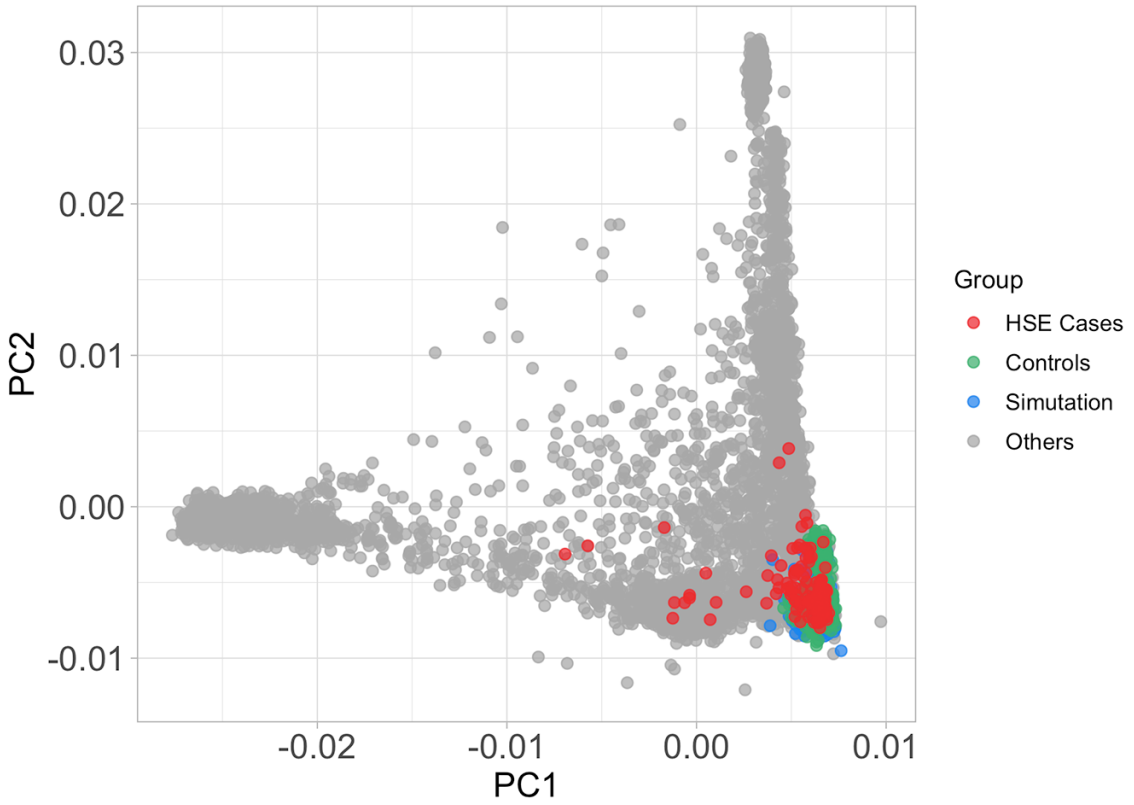

Figure S1. PCA plot of the European individuals studied here. HSE cases: 122 individuals, Controls: 490 individuals, Simulations: 893 individuals, and the other 7,958 individuals in our in-house HGID database.

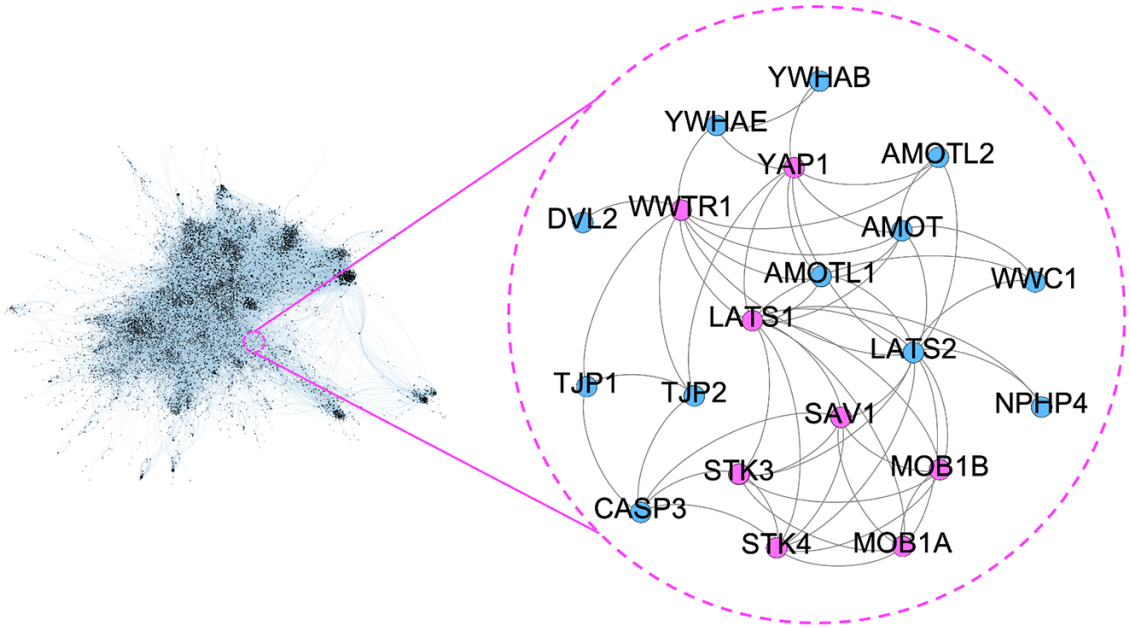

Figure S2. The PPI network of 20 Hippo pathway genes in the background biological interaction network. The eight genes selected for the simulation study are shown in pink. The REACTOME Hippo signaling pathway is accessible via https://reactome.org/content/detail/R-HSA-2028269.

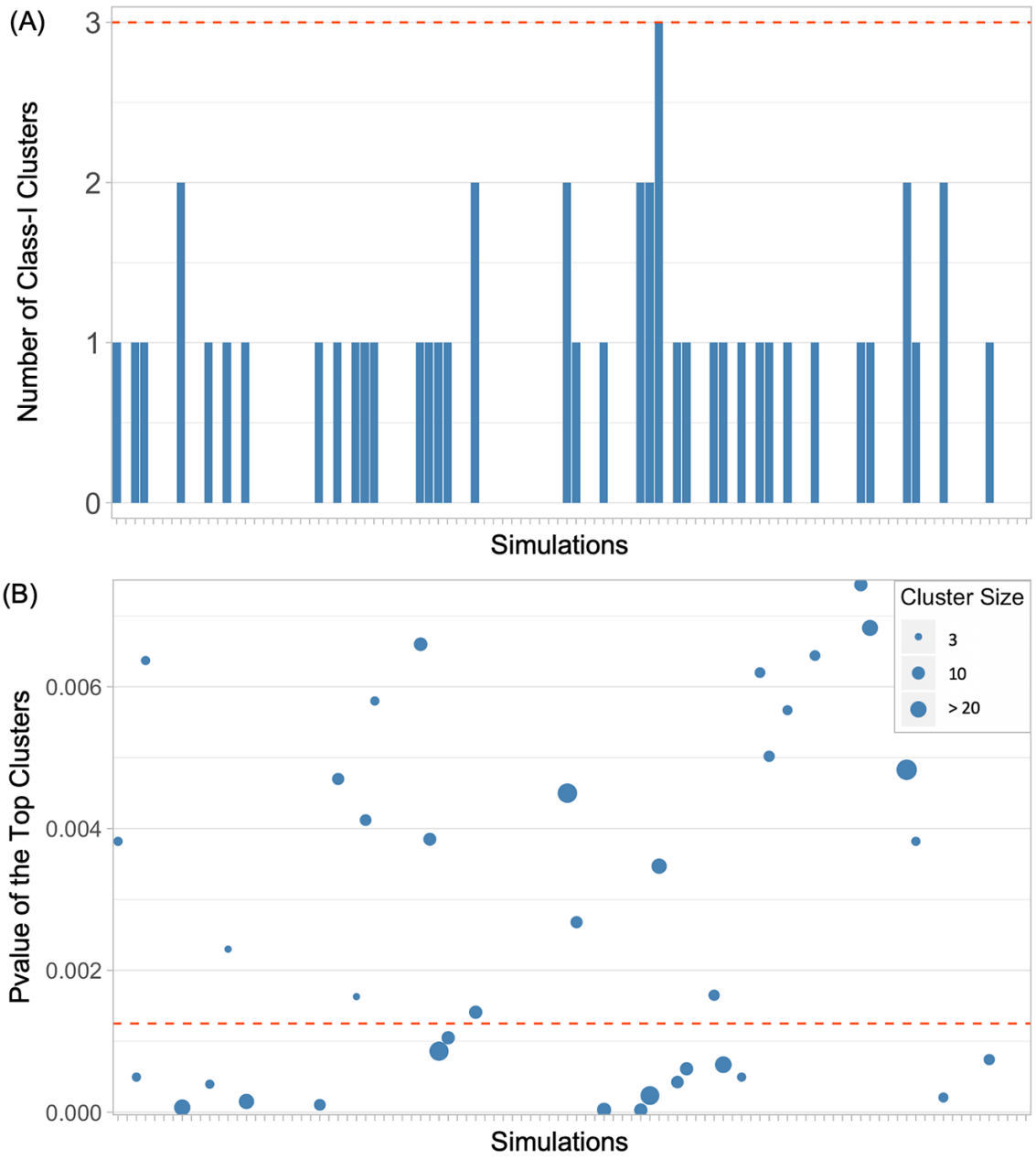

Figure S3. Simulation study for the null hypothesis test. We performed 100 simulations of the NHC method on the randomly sampled 100 individuals with different severe infectious diseases. (A) The number of significant class-I gene clusters obtained in each simulation, where the red line indicates the three class-I gene clusters generated in the HSE study. (B) Cluster-level *p*-value of the top-ranked gene cluster in each simulation, in which the node size represents the number of genes in each top-ranked cluster, and the red line is the *p*-value (=0.00125) of the top-ranked cluster in the HSE study (the HSE study is presented later in the results section.)

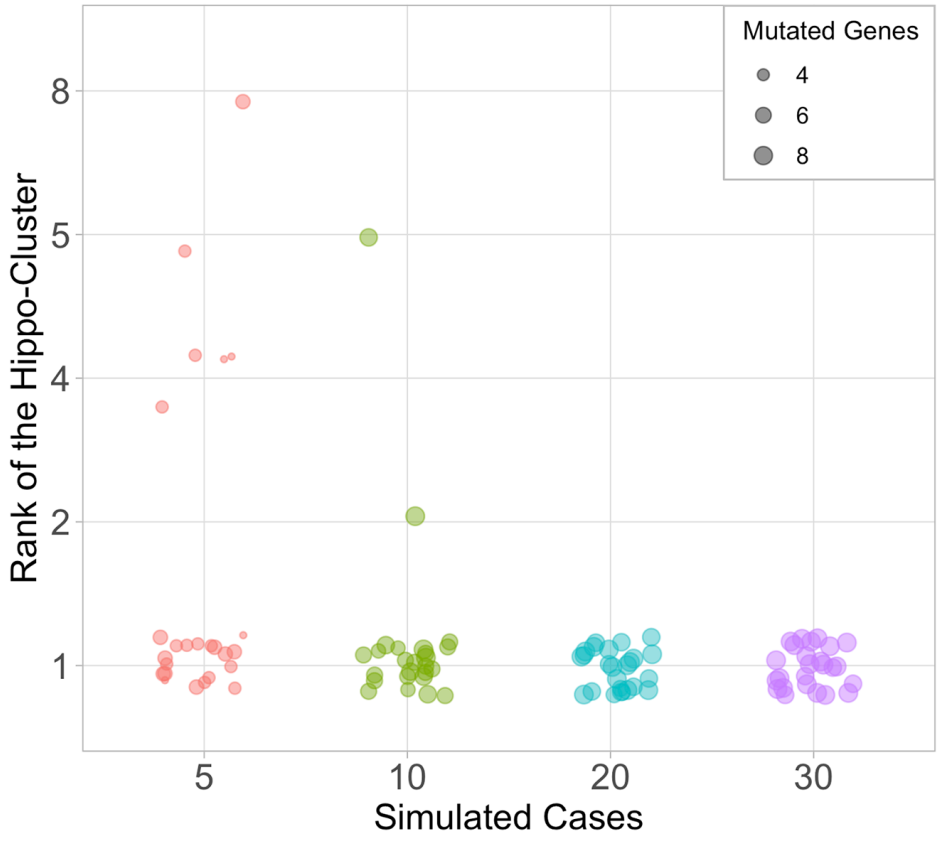

Figure S4. Simulation study for detecting the simulated disease signal in the Hippo pathway. Rank of the gene clusters that are most enriched in the Hippo pathway (Hippo-Cluster), in 100 simulations of 100 random cases with a random subgroup of 5, 10, 20, or 30 cases (25 simulations per subgroup size) being randomly given any of the eight simulated mutations of genes from the Hippo pathway.

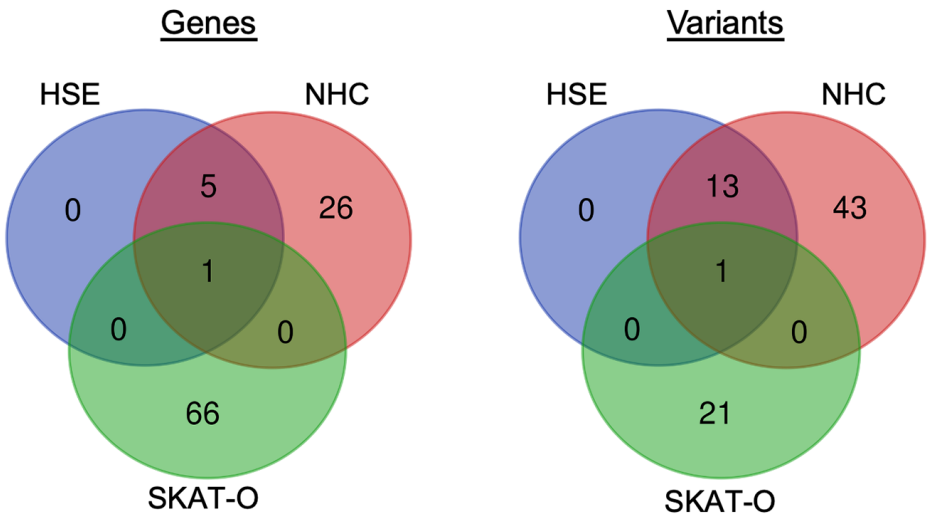

Figure S5. The genes/variants overlapping between different results. The known HSE-causing variants and their genes, the genes/variants identified in the NHC top-ranked gene cluster, and the genes/variants significantly enriched in the SKAT-O test.

Supplemental Tables

#	Control-Embedded Pathways	Occurrence
1	REACTOME_SIGNALING_BY_RECEPTOR_TYROSINE_KINASES	77
2	REACTOME_RRNA_PROCESSING	72
3	REACTOME_CHROMATIN_MODIFYING_ENZYMES	66
4	REACTOME_MITOCHONDRIAL_TRANSLATION	61
5	REACTOME_CELL_CYCLE	58
6	REACTOME_DNA_REPAIR	57
7	REACTOME_MRNA_SPLICING	51
8	KEGG_CELL_CYCLE	49
9	REACTOME_TRANSCRIPTIONAL_REGULATION_OF_WHITE_ADIPOCYTE_DIFFERENTIATION	49
10	REACTOME_METABOLISM_OF_RNA	44
11	KEGG_PATHWAYS_IN_CANCER	41
12	REACTOME_CELL_CYCLE_MITOTIC	39
13	REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	38
14	REACTOME_FORMATION_OF_RNA_POL_II_ELONGATION_COMPLEX	33
15	KEGG_FOCAL_ADHESION	32
16	REACTOME_REGULATION_OF_LIPID_METABOLISM_BY_PPARALPHA	23
17	REACTOME_CELL_CYCLE_CHECKPOINTS	22
18	REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	22
19	REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES	21
20	REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION	21
21	REACTOME_TRANSCRIPTION_OF_THE_HIV_GENOME	20
22	REACTOME_RNA_POLYMERASE_II_TRANSCRIBES_SNRNA_GENES	18
23	REACTOME_TRANSCRIPTIONAL_REGULATION_BY_TP53	18
24	REACTOME_EUKARYOTIC_TRANSLATION_INITIATION	16
25	REACTOME_HATS_ACETYLATE_HISTONES	15
26	KEGG_APOPTOSIS	14
27	REACTOME_NEDDYLATION	14
28	REACTOME_INTRACELLULAR_SIGNALING_BY_SECOND_MESSENGERS	13
29	KEGG_PROTEASOME	12
30	REACTOME_CLATHRIN_MEDIATED_ENDOCYTOSIS	12
31	REACTOME_HIV_TRANSCRIPTION_INITIATION	12
32	REACTOME_EPIGENETIC_REGULATION_OF_GENE_EXPRESSION	11
33	REACTOME_DNA_REPLICATION	10
34	REACTOME_G2_M_CHECKPOINTS	10
35	REACTOME_M_PHASE	10
36	KEGG_ENDOCYTOSIS	9
37	REACTOME_COOPERATION_OF_PDCL_PHLP1_AND_TRIC_CCT_IN_G_PROTEIN_BETA_FOLDING	8
38	REACTOME_COPI_INDEPENDENT_GOLGI_TO_ER_RETROGRADE_TRAFFIC	8
39	REACTOME_MITOTIC_G1_PHASE_AND_G1_S_TRANSITION	8
40	REACTOME_MITOTIC_SPINDLE_CHECKPOINT	8
41	REACTOME_RRNA_MODIFICATION_IN_THE_NUCLEUS_AND_CYTOSOL	8
42	REACTOME_CELLULAR_RESPONSES_TO_EXTERNAL_STIMULI	7
43	REACTOME_SEPARATION_OF_SISTER_CHROMATIDS	7
44	REACTOME_SIGNALING_BY_VEGF	7
45	KEGG_ERBB_SIGNALING_PATHWAY	6
46	REACTOME_CIRCADIAN_CLOCK	6
47	REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	6
48	REACTOME_INTRAFLAGELLAR_TRANSPORT	6
49	KEGG_ENDOMETRIAL_CANCER	5
50	KEGG_MTOR_SIGNALING_PATHWAY	5
51	KEGG_SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	5

52	KEGG_SPLICEOSOME	5
53	REACTOME_APC_C_MEDIATED_DEGRADATION_OF_CELL_CYCLE_PROTEINS	5
54	REACTOME_HIV_INFECTION	5
55	REACTOME_MTOR_SIGNALLING	5
56	REACTOME_PI3K_AKT_SIGNALING_IN_CANCER	5
57	REACTOME_PROCESSING_OF_INTRONLESS_PRE_MRNAS	5
58	REACTOME_REGULATION_OF_TP53_ACTIVITY	5

Table S1: Control-embedded pathways identified by running the NHC method on 100 randomly selected controls versus the remaining 390 controls, for 100 simulations.

Gene	Chrom	Position	Ref Allele	Alt Allele	Consequence	CADD
LATS1	6	150023233	A	T	stop-gained (p.Tyr10*)	35
MOB1A	2	74399864	AG	TC	stop-gained (p.Ser10*)	38
MOB1B	4	71816524	G	T	stop-gained (p.Glu9*)	35
SAV1	14	51134658	C	A	stop-gained (p.Glu10*)	37
STK3	8	99895955	GGT	TTA	stop-gained (p.Thr10*)	22.8
STK4	20	43595237	CC	TA	stop-gained (p.Pro10*)	36
WWTR1	3	149375064	GAG	TTA	stop-gained (p.Leu10*)	36
YAP1	11	101981607	C	T	stop-gained (p.Gln10*)	29.4

Table S2. The artificially created stop-gained mutations of eight genes from the Hippo pathway.

Simulation	# Total Clusters	# Class-I Clusters (p-value ≤ 0.01)	Top Gene Cluster		
			p-value	# Genes	# Cases
simul_1	107	1	3.82E-03	4	7
simul_2	109	1	2.30E-03	3	8
simul_3	104	0	NA	NA	NA
simul_4	106	1	3.85E-03	11	21
simul_5	111	0	NA	NA	NA
simul_6	102	0	NA	NA	NA
simul_7	96	0	NA	NA	NA
simul_8	90	0	NA	NA	NA
simul_9	99	0	NA	NA	NA
simul_10	105	0	NA	NA	NA
simul_11	113	1	6.37E-03	4	10
simul_12	99	0	NA	NA	NA
simul_13	111	0	NA	NA	NA
simul_14	100	0	NA	NA	NA
simul_15	108	2	6.55E-05	23	39
simul_16	110	0	NA	NA	NA
simul_17	110	0	NA	NA	NA
simul_18	109	1	3.95E-04	4	11
simul_19	105	0	NA	NA	NA
simul_20	103	0	NA	NA	NA
simul_21	97	1	1.51E-04	19	37
simul_22	108	0	NA	NA	NA
simul_23	107	0	NA	NA	NA
simul_24	107	0	NA	NA	NA
simul_25	111	0	NA	NA	NA
simul_26	127	0	NA	NA	NA
simul_27	111	0	NA	NA	NA
simul_28	119	0	NA	NA	NA
simul_29	103	1	1.04E-04	8	16
simul_30	110	1	4.70E-03	9	16
simul_31	105	0	NA	NA	NA
simul_32	104	1	1.63E-03	3	10
simul_33	119	1	4.12E-03	8	15
simul_34	105	1	5.80E-03	4	11
simul_35	104	0	NA	NA	NA
simul_36	106	0	NA	NA	NA
simul_37	113	0	NA	NA	NA
simul_38	108	0	NA	NA	NA
simul_39	110	1	6.60E-03	13	22
simul_40	113	1	8.60E-04	38	50
simul_41	109	1	1.05E-03	12	22
simul_42	116	0	NA	NA	NA
simul_43	111	0	NA	NA	NA
simul_44	102	2	1.41E-03	12	30
simul_45	112	0	NA	NA	NA
simul_46	100	0	NA	NA	NA
simul_47	110	0	NA	NA	NA
simul_48	101	0	NA	NA	NA
simul_49	116	0	NA	NA	NA
simul_50	106	0	NA	NA	NA
simul_51	119	0	NA	NA	NA
simul_52	109	0	NA	NA	NA
simul_53	106	2	4.50E-03	40	51
simul_54	109	1	2.68E-03	9	15
simul_55	101	0	NA	NA	NA

simul_56	117	0	NA	NA	NA
simul_57	101	1	3.32E-05	15	29
simul_58	110	0	NA	NA	NA
simul_59	99	0	NA	NA	NA
simul_60	105	2	2.99E-05	12	25
simul_61	102	2	2.35E-04	36	49
simul_62	111	3	3.47E-03	19	32
simul_63	100	0	NA	NA	NA
simul_64	94	1	4.25E-04	10	17
simul_65	106	1	6.12E-04	12	19
simul_66	108	0	NA	NA	NA
simul_67	105	0	NA	NA	NA
simul_68	114	1	1.65E-03	7	12
simul_69	103	1	6.70E-04	25	38
simul_70	102	1	4.95E-04	4	9
simul_71	117	0	NA	NA	NA
simul_72	97	1	6.20E-03	6	15
simul_73	118	1	5.02E-03	7	12
simul_74	98	0	NA	NA	NA
simul_75	115	1	5.67E-03	5	8
simul_76	95	0	NA	NA	NA
simul_77	113	0	NA	NA	NA
simul_78	112	1	6.44E-03	6	9
simul_79	110	0	NA	NA	NA
simul_80	114	0	NA	NA	NA
simul_81	114	0	NA	NA	NA
simul_82	104	1	7.44E-03	12	18
simul_83	113	1	6.83E-03	22	27
simul_84	106	0	NA	NA	NA
simul_85	112	0	NA	NA	NA
simul_86	100	0	NA	NA	NA
simul_87	108	2	4.83E-03	46	54
simul_88	110	1	3.82E-03	4	7
simul_89	106	0	NA	NA	NA
simul_90	116	2	2.07E-04	5	8
simul_91	102	0	NA	NA	NA
simul_92	92	0	NA	NA	NA
simul_93	115	0	NA	NA	NA
simul_94	101	0	NA	NA	NA
simul_95	101	1	7.42E-04	7	14
simul_96	100	0	NA	NA	NA
simul_97	103	0	NA	NA	NA
simul_98	107	0	NA	NA	NA
simul_99	109	0	NA	NA	NA
simul_100	98	1	4.95E-04	4	9

Table S3. Gene clusters generated in the simulation study for the null hypothesis test. We performed the NHC method on 100 randomly sampled individuals with different severe infectious diseases, for 100 times. The number of output gene clusters (after merging the gene clusters from the initial output), the number of significant class-I clusters, and the top-ranked gene clusters are shown.

Simulation	Simulation on Hippo Pathway		Hippo-Cluster (the gene cluster that is most enriched in Hippo pathway)				
	# Simulated Cases	# Mutated Genes	Cluster Rank	# Genes	# Cases	Cluster *p*-value	Hippo Pathway *p*-value
simul 1	5	4	#1	7	17	1.93e-05	9.74e-13
simul 2	5	4	#4	7	11	5.47e-03	1.26e-09
simul 3	5	4	#5	8	12	2.18e-02	3.37e-09
simul 4	5	5	#8	7	10	1.40e-02	9.74e-13
simul 5	5	4	#4	7	12	1.32e-02	1.26e-09
simul 6	5	3	#4	7	13	7.15e-04	9.74e-13
simul 7	5	5	#1	6	11	3.87e-03	1.40e-13
simul 8	5	5	#1	9	14	1.15e-03	1.20e-14
simul 9	5	5	#1	8	20	8.34e-07	2.67e-15
simul 10	5	3	#4	10	15	1.64e-03	3.99e-14
simul 11	5	4	#1	7	16	4.47e-04	1.26e-09
simul 12	5	5	#1	7	16	4.47e-04	9.74e-13
simul 13	5	5	#1	8	16	3.88e-05	2.67e-15
simul 14	5	5	#1	9	19	8.81e-05	1.20e-14
simul 15	5	4	#1	8	18	9.58e-06	2.67e-15
simul 16	5	5	#1	6	19	1.65e-07	1.40e-13
simul 17	5	4	#1	8	20	8.34e-07	2.67e-15
simul 18	5	4	#1	7	16	2.14e-04	9.74e-13
simul 19	5	4	#1	10	27	7.62e-09	3.99e-14
simul 20	5	4	#1	9	17	4.78e-05	1.20e-14
simul 21	5	3	#1	8	21	6.89e-06	3.37e-09
simul 22	5	5	#1	9	17	1.09e-04	1.20e-14
simul 23	5	3	#1	10	21	1.81e-06	3.99e-14
simul 24	5	4	#1	8	23	4.74e-09	2.67e-15
simul 25	5	4	#1	9	23	1.07e-06	1.16e-11
simul 26	10	8	#2	10	17	8.75e-04	3.99e-14
simul 27	10	7	#5	9	14	1.15e-03	1.20e-14
simul 28	10	6	#1	8	21	1.33e-07	2.67e-15
simul 29	10	5	#1	6	16	8.38e-06	4.07e-07
simul 30	10	7	#1	8	22	6.36e-08	2.67e-15
simul 31	10	7	#1	6	17	4.15e-06	1.40e-13
simul 32	10	6	#1	9	21	6.89e-06	1.20e-14
simul 33	10	6	#1	8	23	4.74e-09	2.67e-15
simul 34	10	6	#1	7	25	1.33e-10	3.34e-16
simul 35	10	8	#1	10	24	1.23e-06	3.99e-14
simul 36	10	5	#1	9	21	4.03e-07	1.20e-14
simul 37	10	6	#1	8	26	1.29e-10	3.89e-12
simul 38	10	6	#1	11	31	3.28e-10	1.09e-13
simul 39	10	8	#1	9	31	3.06e-13	1.20e-14
simul 40	10	6	#1	10	31	6.14e-11	3.99e-14
simul 41	10	5	#1	9	29	5.38e-10	1.20e-14
simul 42	10	7	#1	9	36	2.14e-14	1.20e-14
simul 43	10	6	#1	10	38	1.93e-17	3.99e-14
simul 44	10	6	#1	10	29	1.61e-09	3.99e-14
simul 45	10	6	#1	9	36	2.14e-14	1.20e-14
simul 46	10	6	#1	7	30	5.94e-14	3.34e-16
simul 47	10	6	#1	9	39	6.16e-19	1.20e-14
simul 48	10	7	#1	9	46	1.29e-21	1.20e-14
simul 49	10	6	#1	8	33	5.93e-15	2.67e-15
simul 50	10	6	#1	8	38	1.39e-18	2.67e-15
simul 51	20	8	#1	8	21	7.31e-08	2.67e-15
simul 52	20	8	#1	10	26	4.52e-08	3.99e-14

simul_53	20	7	#1	8	24	1.15e-09	2.67e-15
simul_54	20	8	#1	10	26	4.52e-08	3.99e-14
simul_55	20	6	#1	6	24	2.46e-11	1.40e-13
simul_56	20	7	#1	10	30	2.43e-10	3.99e-14
simul_57	20	8	#1	9	29	3.04e-10	1.20e-14
simul_58	20	7	#1	9	29	6.38e-12	1.20e-14
simul_59	20	7	#1	10	37	1.02e-16	3.99e-14
simul_60	20	7	#1	8	33	1.07e-15	2.67e-15
simul_61	20	8	#1	9	38	9.37e-16	1.20e-14
simul_62	20	7	#1	9	37	1.89e-17	1.20e-14
simul_63	20	6	#1	7	37	4.48e-19	3.34e-16
simul_64	20	7	#1	9	37	1.89e-17	1.20e-14
simul_65	20	8	#1	9	45	4.41e-24	1.20e-14
simul_66	20	8	#1	9	41	7.13e-18	1.20e-14
simul_67	20	8	#1	10	43	5.39e-19	3.99e-14
simul_68	20	8	#1	8	49	1.92e-27	2.67e-15
simul_69	20	8	#1	9	51	1.07e-28	1.20e-14
simul_70	20	7	#1	10	51	9.23e-28	3.99e-14
simul_71	20	8	#1	9	42	1.34e-18	1.20e-14
simul_72	20	8	#1	9	50	2.61e-28	1.20e-14
simul_73	20	8	#1	9	48	1.15e-25	1.20e-14
simul_74	20	6	#1	11	54	5.95e-27	1.09e-13
simul_75	20	7	#1	7	46	2.33e-26	3.34e-16
simul_76	30	8	#1	8	29	9.13e-11	2.67e-15
simul_77	30	8	#1	9	34	2.68e-15	1.20e-14
simul_78	30	8	#1	10	33	1.93e-12	3.99e-14
simul_79	30	8	#1	10	36	2.14e-14	3.99e-14
simul_80	30	8	#1	7	35	1.46e-17	3.34e-16
simul_81	30	8	#1	9	46	1.87e-24	1.20e-14
simul_82	30	8	#1	7	40	2.04e-21	3.34e-16
simul_83	30	8	#1	9	39	3.58e-18	1.20e-14
simul_84	30	8	#1	8	48	1.44e-27	2.67e-15
simul_85	30	8	#1	8	46	6.59e-25	2.67e-15
simul_86	30	8	#1	9	44	4.35e-20	1.20e-14
simul_87	30	7	#1	9	47	3.62e-23	1.20e-14
simul_88	30	8	#1	9	53	2.62e-27	1.20e-14
simul_89	30	8	#1	9	57	1.07e-31	1.20e-14
simul_90	30	8	#1	9	51	1.07e-28	1.20e-14
simul_91	30	8	#1	9	56	1.07e-32	1.20e-14
simul_92	30	8	#1	7	58	2.59e-37	3.34e-16
simul_93	30	8	#1	8	63	1.51e-40	2.67e-15
simul_94	30	7	#1	10	59	3.53e-32	3.99e-14
simul_95	30	8	#1	9	66	4.51e-43	1.20e-14
simul_96	30	7	#1	9	58	4.42e-35	1.20e-14
simul_97	30	7	#1	10	66	2.25e-39	3.99e-14
simul_98	30	8	#1	8	64	1.43e-41	2.67e-15
simul_99	30	8	#1	7	67	1.40e-46	3.34e-16
simul_100	30	8	#1	9	72	9.24e-47	1.20e-14

Table S4. Identification of the Hippo-cluster in simulation study II for the detection of simulated disease signals. We performed the NHC method on 100 random cases, in which a random subgroup of 5, 10, 20, or 30 cases was randomly assigned any of the eight simulated mutations of genes from the Hippo pathway. We identified the gene clusters most enriched in the Hippo pathway.

Rank	Enriched Pathway	Pathway *p*-value
1	KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	3.21e-15
2	REACTOME_TOLL_LIKE_RECEPTOR_CASCADES	8.88e-15
3	REACTOME_INNATE_IMMUNE_SYSTEM	1.85e-13
4	REACTOME_TICAM1_DEPENDENT_ACTIVATION_OF_IRF3_IRF7	3.40e-13
5	REACTOME_DDX58_IFIH1_MEDIATED_INDUCTION_OF_INTERFERON_ALPHA_BETA	7.47e-13
6	REACTOME_TOLL_LIKE_RECEPTOR_4_TLR4_CASCADE	2.35e-12
7	REACTOME_ACTIVATION_OF_IRF3_IRF7_MEDIATED_BY_TBK1_IKK_EPSILON	2.89e-12
8	REACTOME_NEGATIVE_REGULATORS_OF_DDX58_IFIH1_SIGNALING	3.11e-12
9	REACTOME_MYD88_INDEPENDENT_TLR4_CASCADE	6.83e-12
10	REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM	8.39e-12
11	REACTOME_TRAF3_DEPENDENT_IRF_ACTIVATION_PATHWAY	1.41e-10
12	KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.20e-09
13	REACTOME_TRAF6_MEDIATED_IRF7_ACTIVATION	9.84e-09
14	REACTOME_DISEASES_OF_IMMUNE_SYSTEM	3.07e-07
15	KEGG_CYTOSOLIC_DNA_SENSING_PATHWAY	5.51e-07
16	KEGG_NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	1.19e-06
17	REACTOME_INTERFERON_SIGNALING	7.78e-06

Table S5. The significantly enriched pathways of the top-ranked gene cluster from the HSE cohort.

Cluster	# Genes	# Variants Hom | Het	# Cases	Cluster *p*-value	Most Enriched Pathway (p-value)
Class-I					
#1	27	0 | 42	37	0.00176	KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY (6.73e-14)
#2	5	0 | 13	12	0.00261	REACTOME_HDR_THROUGH_SINGLE_STRAND_ANNEALING_SSA (9.676e-10)
#3	6	0 | 14	14	0.00445	REACTOME_PEROXISOMAL_PROTEIN_IMPORT(7.262e-11)
#4	15	0 | 28	26	0.00841	REACTOME_RETROGRADE_TRANSPORT_AT_THE_TRANS_GOLGI_NETWORK (1.106e-14)
Class-II					
#5	30	0 | 46	41	0.00041	REACTOME_MITOCHONDRIAL_TRANSLATION (1.121e-57)
#6	9	0 | 22	20	0.00199	REACTOME_PROCESSING_OF_INTRONLESS_PRE_MRNAS (8.875e-15)
This class-I cluster is just above the cluster-level significance cutoff, but functionally interesting					
#7	6	0 | 26	24	0.01023	KEGG_REGULATION_OF_AUTOPHAGY (3.144e-06)

Table S6. The six significant gene clusters (p-value ≤ 0.01) detected by NHC in the HSE cohort of 109 cases of unknown disease etiology, and one gene cluster with a p-value of 0.01023, just above the cutoff, but nevertheless functionally interesting.

Rank	Pathway	# Genes	Gene List	*p*-value
1	REACTOME_TRAFFICKING_AND_PROCESSING_OF_ENDOSOMAL_TLR	13	CNPY3, CTSB, CTSK, CTSL, CTSS, CTSV, HSP90B1, LGMN, TLR3, TLR7, TLR8, TLR9, UNC93B1	0.000247
2	REACTOME_TICAM1_DEPENDENT_ACTIVATION_OF_IRF3_IRF7	13	IKBKE, IRF3, IRF7, RPS27A, TANK, TBK1, TICAM1, TLR3, TRAF3, UBA52, UBB, UBC	0.000635
3	REACTOME_ZBP1_DAI_MEDIATED_INDUCTION_OF_TYPE_I_IFNS	21	CHUK, DHX9, DTX4, IKBKB, IKBKG, IRF3, MYD88, NFKB1, NFKB2, NFKBIA, NFKBIB, NKIRAS1, NKIRAS2, NLRP4, RELA, RIPK1, RIPK3, TBK1, TICAM1, TLR3, ZBP1	0.000691
4	KEGG_RIBOFLAVIN_METABOLISM	16	ACP1, ACP2, ACP3, ACP4, ACP5, ACP6, ENPP1, ENPP3, FLAD1, MTMR1, MTMR2, MTMR6, MTMR7, PHPT1, RFK, TYR	0.000854
5	REACTOME_CROSSLINKING_OF_COLLAGEN_FIBRILS	18	BMP1, COL1A1, COL1A2, COL4A1, COL4A2, COL4A3, COL4A4, COL4A5, COL4A6, LOX, LOXL1, LOXL2, LOXL3, LOXL4, PCOLCE, PXDN, TLL1, TLL2	0.000992

Table S7. The significant pathways (*p*-value < 0.001) identified by running our own scripted pathway-informed SKAT-O test on the HSE cohort. The genes in the list underlined with a solid line are the genes harboring the known HSE-causing mutations, and those underlined with a wavy line are the candidate genes that we selected from the top-ranked NHC gene cluster.

Cluster	#Genes	#Var Hom | Het	#Cases	Cluster *p*-value	#Pathways	Top Pathway	#BP	Top BP	#MF	Top MF
Class-I										
#1	28	4 | 55	49	3.12e-05	16	KEGG_TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY (3.594e-16)	11	GO:0035666:TRIF-dependent toll-like receptor signaling pathway (1.029e-16)	0	.
#2	5	0 | 14	13	0.00274	11	REACTOME_HDR_THROUGH_SINGLE_STRAND_ANNEALING_SSA (9.676e-10)	4	GO:1901796:regulation of signal transduction by p53 class mediator (2.108e-07)	0	.
#3	6	0 | 15	15	0.00564	3	REACTOME_PEROXISOMAL_PROTEIN_IMPORT (7.262e-11)	2	GO:0006625:protein targeting to peroxisome (2.81e-11)	0	.
Class-II										
#4	26	1 | 44	44	6.44e-05	2	REACTOME_MITOCHONDRIAL_TRANSLATION (7.58e-50)	4	GO:0070125:mitochondrial translational elongation (4.128e-55)	2	GO:0003735:structural constituent of ribosome (1.993e-34)
#5	9	0 | 23	22	0.00143	6	REACTOME_PROCESSING_OF_INTRONLESS_PRE MRNAS (5.327e-18)	6	GO:0006378:mRNA polyadenylation (4.458e-17)	0	.

Table S8: NHC-boost detected the same number of significant gene clusters in the HSE cohort of 122 individuals, but with slightly fewer genes in clusters #1, #4 and #5 (see Table 2 for comparison), in a significantly shorter computation time: 5 minutes, versus 40 minutes for the original NHC code.