Supplementary Table 1. Allele frequencies and results from the association analyses for SNPs with association results not reaching nominal significance with consistent effect sizes in the replication.

					Genome-wide analysis				Replication analysis						
					Allele frequencies (Cases/Controls)				Allele frequencies (Cases/Controls)						
Chr	**SNP**	**Position**	**Locus**	**Alleles**	**Scandinavia (332/262)**	**Germany (383/2700)**	***P* - value***	**OR (95% CI)***	**Scandinavia (289/820)**	**Central Europe (561/2063)**	**United States (371/625)**	***P*-value†**	**OR (95% CI)†**	**BD P-value**	**SNP selection strategy**
1	rs2377570	31,071,356	*SDC3/SNORD85*	A/G	0.26/0.19	0.27/0.22	6.3E-05	1.34 (1.55-1.16)	0.23/0.20	0.20/0.21	0.22/0.22	0.82	1.01 (0.90-1.14)	0.29	II
1	rs12144426	164,418,658	*FAM78B*	T/A	0.24/0.31	0.21/0.25	3.5E-05	0.72 (0.62-0.84)	0.25/0.27	0.24/0.24	0.26/0.24	0.94	1.00 (0.89-1.11)	0.48	II
2	rs1990760	162,832,297	*IFIH1*	C/T	0.39/0.38	0.44/0.39	0.0070	1.21 (1.39-1.05)	0.37/0.37	0.40/0.38	0.41/0.41	0.72	1.02 (0.92-1.12)	0.76	I
3	rs7638558	60,020,841	*FHIT*	C/T	0.02/0.00	0.04/0.02	3.6E-05	2.35 (3.52-1.57)	0.01/0.00	0.02/0.01	0.01/0.02	0.83	1.04 (0.69-1.57)	0.24	II
3	rs983513	79,277,847	*ROBO1*	G/A	0.45/0.53	0.39/0.45	2.5E-05	0.75 (0.65-0.86)	0.48/0.49	0.48/0.46	0.49/0.48	0.30	1.05 (0.96-1.16)	0.55	II
4	rs17005387	123,156,648	*KIAA1109/IL2/IL21*	A/G	0.02/0.01	0.03/0.02	2.6E-05	2.41 (3.63-1.60)	0.01/0.02	0.02/0.02	0.03/0.02	0.34	1.19 (0.84-1.67)	0.26	II
4	rs993704	125,931,510	*ANKRD50*	A/G	0.30/0.24	0.36/0.30	2.3E-05	1.35 (1.54-1.17)	0.26/0.26	0.32/0.30	0.27/0.27	0.29	1.06 (0.95-1.18)	0.78	II
4	rs10857102	125,937,433	*ANKRD50*	A/G	0.26/0.20	0.30/0.25	3.2E-05	1.36 (1.57-1.18)	0.25/0.25	0.31/0.29	0.26/0.26	0.23	1.07 (0.96-1.19)	0.78	II
5	rs1000113	150,220,269	*IRGM*	T/C	0.07/0.08	0.05/0.07	0.015	0.72 (0.56-0.94)	0.07/0.07	0.07/0.07	0.08/0.08	0.74	1.03 (0.86-1.24)	0.86	I
5	rs11747270	150,239,060	*IRGM*	G/A	0.07/0.08	0.05/0.08	0.021	0.74 (0.58-0.96)	0.07/0.07	0.07/0.07	0.08/0.08	0.81	1.02 (0.85-1.23)	0.88	I
6	rs394683	5,043,412	*LYRM4*	C/T	0.26/0.31	0.24/0.30	2.2E-05	0.73 (0.85-0.63)	0.26/0.29	0.30/0.28	0.26/0.30	0.37	0.95 (0.86-1.06)	0.036	II
6	rs4713859	35,514,131	*PPARD/FANCE*	C/T	0.02/0.03	0.02/0.05	5.3E-05	0.43 (0.65-0.29)	0.03/0.03	0.05/0.05	0.05/0.06	0.81	1.03 (0.82-1.29)	0.67	II
7	rs590099	18,302,863	*HDAC9*	A/G	0.16/0.13	0.21/0.16	2.6E-05	1.42 (1.67-1.20)	0.13/0.17	0.16/0.17	0.17/0.16	0.13	0.90 (0.79-1.03)	0.16	II
7	rs17806432	76,798,820	*PION/FGL2*	T/C	0.14/0.11	0.15/0.11	1.2E-05	1.59 (1.29-1.95)	0.13/0.10	0.10/0.11	0.11/0.08	0.27	1.09 (0.94-1.27)	0.11	II
7	rs6979188	76,821,451	*PION/FGL2*	T/C	0.10/0.06	0.10/0.06	2.1E-05	1.64 (1.30-2.05)	0.07/0.07	0.07/0.07	0.08/0.06	0.49	1.07 (0.89-1.28)	0.68	II
8	rs10156297	127,621,008	*FAM84B*	C/A	0.16/0.13	0.18/0.12	6.8E-05	1.42 (1.20-1.70)	0.13/0.12	0.12/0.13	0.13/0.12	0.86	0.99 (0.86-1.14)	0.23	II
11	rs11218714	122,000,841	*UBASH3B*	G/A	0.30/0.35	0.27/0.34	3.2E-05	0.75 (0.65-0.86)	0.32/0.31	0.31/0.31	0.33/0.31	0.77	1.02 (0.92-1.12)	0.61	II
11	rs722449	132,656,146	*OPCML*	A/G	0.06/0.03	0.08/0.04	8.8E-06	1.82 (2.36-1.40)	0.04/0.05	0.05/0.04	0.04/0.05	0.74	1.04 (0.83-1.30)	0.45	II
12	rs3764021	9,724,895	*CLEC2D*	T/C	0.41/0.45	0.45/0.47	0.032	0.87 (0.77-0.99)	0.45/0.46	0.47/0.47	0.44/0.48	0.36	0.96 (0.87-1.05)	0.38	I
12	rs608418	10,023,132	*CLEC12A*	T/C	0.57/0.47	0.55/0.49	1.4E-05	1.32 (1.17-1.50)	0.52/0.50	0.49/0.48	0.50/0.50	0.49	1.03 (0.94-1.14)	0.90	II
12	rs2117032	20,965,389	*SLCO1B3*	C/T	0.35/0.31	0.38/0.35	0.045	1.15 (1.31-1.00)	0.35/0.35	0.34/0.36	0.39/0.38	0.73	0.98 (0.89-1.08)	0.54	I
13	rs9576711	38,521,621	*STOML3*	T/G	0.11/0.15	0.06/0.10	2.3E-05	0.60 (0.48-0.76)	0.11/0.13	0.12/0.08	0.09/0.11	0.39	1.07 (0.92-1.24)	2.6E-04	II
13	rs1413040	38,605,939	*STOML3*	A/G	0.26/0.34	0.22/0.26	9.8E-05	0.74 (0.86-0.64)	0.30/0.30	0.33/0.27	0.28/0.30	0.041	1.11 (1.00-1.23)	0.027	II
15	rs289404	83,367,321	*PDE8A*	G/T	0.15/0.21	0.14/0.18	6.5E-05	0.69 (0.83-0.57)	0.19/0.20	0.21/0.23	0.24/0.25	0.078	0.90 (0.81-1.01)	0.70	II
16	rs9888739	31,220,754	*ITGAM*	T/C	0.11/0.13	0.10/0.12	0.026	0.78 (0.63-0.97)	0.09/0.10	0.10/0.11	0.12/0.12	0.61	0.96 (0.82-1.12)	0.88	I
16	rs7190071	71,742,579	*ZFHX3*	T/C	0.39/0.30	0.35/0.30	6.9E-06	1.39 (1.20-1.60)	0.34/0.37	0.33/0.36	0.34/0.35	0.026	0.89 (0.81-0.99)	0.62	II
18	rs4310957	69,756,844	*FBXO15*	G/T	0.32/0.24	0.31/0.27	9.0E-06	1.42 (1.65-1.21)	0.29/0.32	0.28/0.28	0.31/0.30	0.81	0.99 (0.89-1.10)	0.56	II
20	rs6080774	17,600,784	*RRBP1*	A/G	0.04/0.03	0.06/0.04	6.1E-05	2.20 (3.24-1.50)	0.02/0.02	0.03/0.02	0.02/0.02	0.077	1.32 (0.97-1.81)	0.70	II

Supplementary Table 2. Genome-wide association studies on immune-mediated and chronic inflammatory traits taken into consideration in SNP prioritization strategy 1.

Disease/Trait	First Author	Journal	PubMed ID
AIDS	Le Clerc	*J Infect Dis*	19754311
AIDS (progression)	Limou	*J Infect Dis*	19115949
Ankylosing spondylitis	The Australo-Anglo-American Spondyloarthritis Consortium (TASC)	*Nat Genet*	20062062
Anti-cyclic Citrullinated Peptide Antibody	Cui	*Mol Med*	19287509
Arthritis (juvenile idiopathic)	Hinks	*Arthritis Rheum*	19116933
	Behrens	*Arthritis Rheum*	18576341
Asthma	Sleiman	*N Engl J Med*	20032318
	Mathias	*J Allergy Clin Immunol*	19910028
	Himes	*Am J Hum Genet*	19426955
	Moffatt	*Nature*	17611496
	Li	*J Allergy Clin Immunol*	20159242
Asthma (childhood onset)	Hancock	*PLoS Genet*	19714205
Asthma (toluene diisocyanate-induced)	Kim	*Clin Exp Allergy*	19187332
Atopic dermatitis	Esparza-Gordillo	*Nat Genet*	19349984
Atopy	Castro-Giner	*BMC Med Genet*	19961619
Behcet's disease	Fei	*Arthritis Res Ther*	19442274
Bilirubin levels	Sanna	*Hum Mol Genet*	19419973
Biochemical measures	Zemunik	*Croat Med J*	19260141
Celiac disease	Hunt	*Nat Genet*	18311140
	van Heel	*Nat Genet*	17558408
	Dubois	*Nat Genet*	20190752
Chronic Hepatitis C infection	Rauch	*Gastroenterology*	20060832
Chronic Obstructive Pulmonary Disease	Pillai	*PLoS Genet*	19300482
	Cho	*Nat Genet*	20173748
C-reactive protein	Elliott	*JAMA*	19567438
	Reiner	*Am J Hum Genet*	18439552
	Ridker	*Am J Hum Genet*	18439548
Crohn's disease	Barrett	*Nat Genet*	18587394
	Raelson	*Proc Natl Acad Sci USA*	17804789
	Franke	*PLoS ONE*	17684544
	WTCCC	*Nature*	17554300
	Parkes	*Nat Genet*	17554261
	Libioulle	*PLoS Genet*	17447842
Crohn's disease and Sarcoidosis (combined)	Franke	*Gastroenterology*	18723019

Cystic fibrosis severity	Gu	*Nature*	19242412
Diabetic nephropathy	Pezzolesi	*Diabetes*	19252134
Drug-induced liver injury (flucloxacillin)	Daly	*Nat Genet*	19483685
Eosinophilic esophagitis (pediatric)	Rothenberg	*Nat Genet*	20208534
Gallstones	Buch	*Nat Genet*	17632509
Hematological and biochemical traits	Kamatani	*Nat Genet*	20139978
Hepatitis B	Kamatani	*Nat Genet*	19349983
HIV-1 control	Fellay	*PLoS Genet*	20041166
HIV-1 viral setpoint	Fellay	*Science*	17641165
Idiopathic pulmonary fibrosis	Mushiroda	*J Med Genet*	18835860
Inflammatory bowel disease	Kugathasan	*Nat Genet*	18758464
	Duerr	*Science*	17068223
Inflammatory bowel disease (early onset)	Imielinski	*Nat Genet*	19915574
Kawasaki disease	Burgner	*PLoS Genet*	19132087
Knee osteoarthritis	Nakajima	*PLoS One*	20305777
Leprosy	Zhang	*N Engl J Med*	20018961
Lupus	Cervino	*Ann NY Acad Sci*	17911428
Malaria	Jallow	*Nat Genet*	19465909
Multiple Sclerosis	Bahlo	*Nat Genet*	19525955
	De Jager	*Nat Genet*	19525953
	Baranzini	*Hum Mol Genet*	19010793
	Aulchenko	*Nat Genet*	18997785
	Comabella	*PLoS ONE*	18941528
	Hafler	*N Engl J Med*	17660530
	Jakkula	*Am J Hum Genet*	20159113
Multiple Sclerosis (age of onset)	Baranzini	*Hum Mol Genet*	19010793
Multiple Sclerosis (severity)	Baranzini	*Hum Mol Genet*	19010793
Neuromyeltis optica	Kim	*Neurobiol Dis*	19850125
Neutrophil count	Okada	*Hum Mol Genet*	20172861
Osteoarthritis	Zhai	*J Med Genet*	19508968
	Kerkhof	*Arthritis Rheum*	20112360
Periodontitis	Schaefer	*Hum Mol Genet*	19897590
Plasma levels of liver enzymes	Yuan	*Am J Hum Genet*	18940312
Primary biliary chirrosis	Hirschfield	*N Engl J Med*	19458352
Psoriasis	Nair	*Nat Genet*	19169254
	Zhang	*Nat Genet*	19169255
	Liu	*PLoS Genet*	18369459
	Capon	*Hum Mol Genet*	18364390
Rheumatoid arthritis	Gregersen	*Nat Genet*	19503088
	Raychaudhuri	*Nat Genet*	18794853
	Julia	*Arthritis Rheum*	18668548
	Plenge	*Nat Genet*	17982456
	Plenge	*N Engl J Med*	17804836

	WTCCC	*Nature*	17554300
Sarcoidosis	Hofmann	*Nat Genet*	18690218
Serum bilirubin levels	Johnson	*Hum Mol Genet*	19414484
Serum soluble E-selectin	Paterson	*Arterioscler Thromb Vasc Biol*	19729612
Soluble leptin receptor levels	Sun	*Hum Mol Genet*	20167575
Soluble levels of adhesion molecules	Barbalic	*Hum Mol Genet*	20167578
Systemic lupus erythematosus	Han	*Nat Genet*	19838193
	Graham	*Nat Genet*	18677312
	Harley	*Nat Genet*	18204446
	Hom	*N Engl J Med*	18204098
	Kozyrev	*Nat Genet*	18204447
	Yang	*PLoS Genet*	20169177
Systemic sclerosis	Zhou	*Arthritis Rheum*	19950302
Type 1 diabetes	Wallace	*Nat Genet*	19966805
	Barrett	*Nat Genet*	19430480
	Cooper	*Nat Genet*	18978792
	Grant	*Diabetes*	18840781
	Hakonarson	*Diabetes*	18198356
	Hakonarson	*Nature*	17632545
	Todd	*Nat Genet*	17554260
Ulcerative colitis	Asano	*Nat Genet*	19915573
	Barrett	*Nat Genet*	19915572
	Silverberg	*Nat Genet*	19122664
	Franke A	*Nat Genet*	20228798
	McGovern	*Nat Genet*	20228799
	Franke	*Nat Genet*	18836448
Vitiligo	Birlea	*J Invest Dermatol*	19890347

The table lists all the genome-wide association studies on immune-mediated and chronic inflammatory traits that were taken into consideration in SNP prioritization strategy 1. The studies were identified using the Catalog of Genome-Wide Association Studies (http://www.genome.gov/26525384) (accessed 23.04.2010) [36].

Supplementary Table 3. Evaluation of abundance patterns in the major phyla via Kruskal- Wallis and *post hoc* Mann-Whitney U tests.

					post hoc test (MWU)		
Phylum	Factor	*df*	X^2	*P*-Value	Factor*	*W*	*P*-Value
Firmicutes	Secretor status	1	3,725	0.054	NA[†]		
Proteobacteria	Secretor status	1	7,303	0.007	NA		
Proteobacteria	Genotype	2	7,323	0.026	AA - AG	49	0.019
					GG - AA	12	0.015
					GG - AG	85	0.981
Bacteroidetes	Secretor status	1	0,648	0.421	NA		
Actinobacteria	Genotype	2	6,344	0.042	AA - AG	120	0.539
					GG - AA	25	0.197
					GG - AG	30	0.009
Tenericutes	Genotype	2	7,023	0.030	AA - AG	143.5	0.092
					GG - AA	25	0.193
					GG - AG	38.5	0.021

*Genotype at the SNP rs601338 with G being the functional allele.

[†]Test not applied.

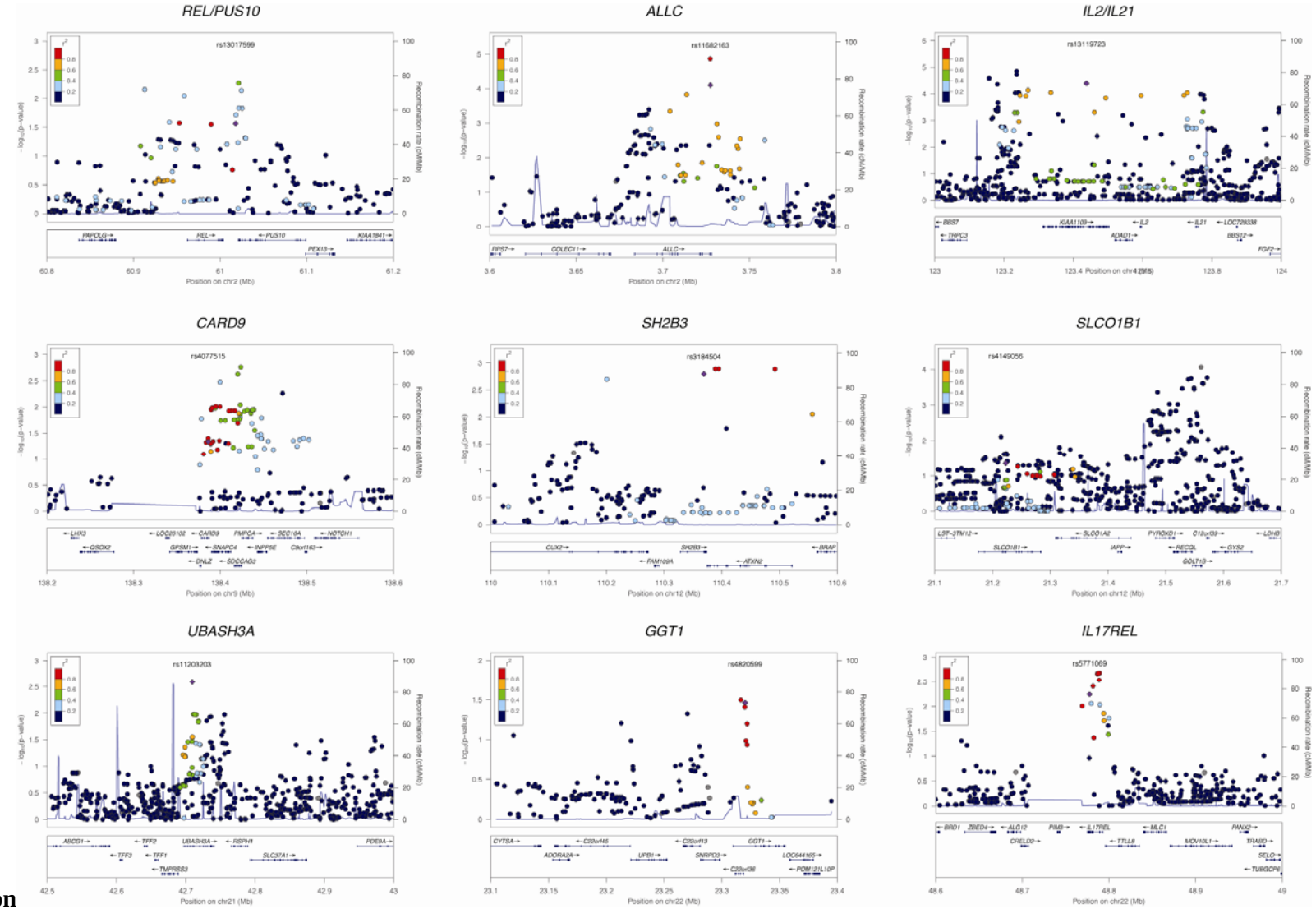
MWU, Mann-Whitney U

Supplementary Table 4. Analysis of alpha diversity via linear modeling.

					post hoc Tukey-HSD	
Alpha diversity	Factor	*df*	*F*-Value	*P*-Value	Factor*	*P*-Value
ACE	Genotype	2	3.015	0.062	AA - AG	0.657
					GG - AA	0.334
					GG - AG	0.049
Phylogenetic Diversity	Genotype	2	3.930	0.029	AA - AG	0.553
					GG - AA	0.259
					GG - AG	0.022

*Genotype at the SNP rs601338 with G being the functional allele.

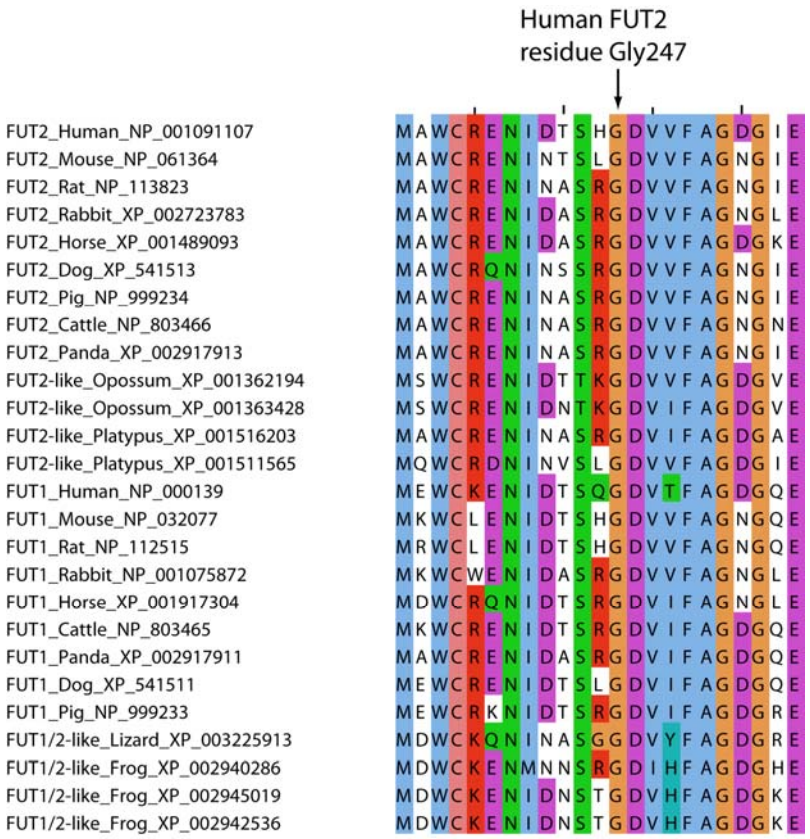
Supplementary Figure 1. Regional association plots for additional nominally replicated region



Association results from the genetic regions with nominally replicated SNPs not shown in Figure 1 in the main manuscript. The plots were generated using the LocusZoom software [8]. The association results for both the genotyped and imputed SNPs are represented by the $-\log_{10}$ *P*-

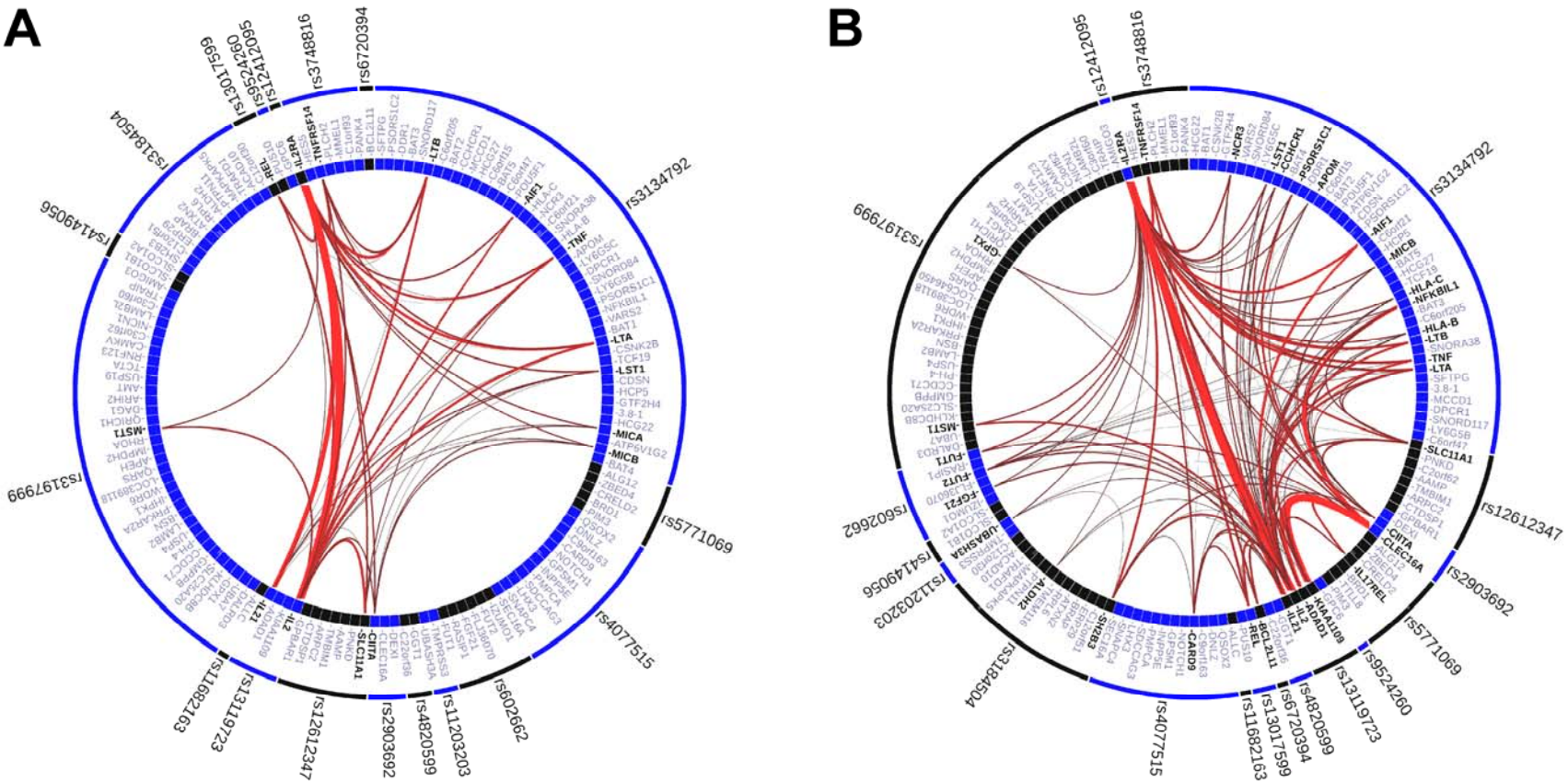
value plotted against the genomic position. The index SNP is marked out with a purple diamond while the colors of the remaining SNPs indicated the linkage disequilibrium (LD) with the index SNP. The recombination rates were derived from the HapMap project and are represented by the thin blue lines.

Supplementary Figure 2. Sequence analysis for a segment of human FUT2 containing Gly247.



A multiple sequence alignment of residues 235-257 of human FUT2 and homologous sequences from tetrapods shows Gly247 to be evolutionary conserved in both FUT2 and FUT1 in mammals, lizard and frogs, strongly suggesting functional importance. The sequences were obtained from the RefSeq protein sequence database [36]. Fold recognition modeling with Phyre [37] indicates that the *Bradyrhizobium* NodZ fucosyltransferase [38] is the closest homolog of FUT2 with a known 3D structure and that FUT2 Gly247 is localized in the active site of the enzyme, in the second loop of the conserved β-α-β glycogen phosphorylase/glycosyltransferase (GPGTF) motif described by Wrabl and Grishin [39].

Supplementary Figure 3. GRAIL pathway analysis

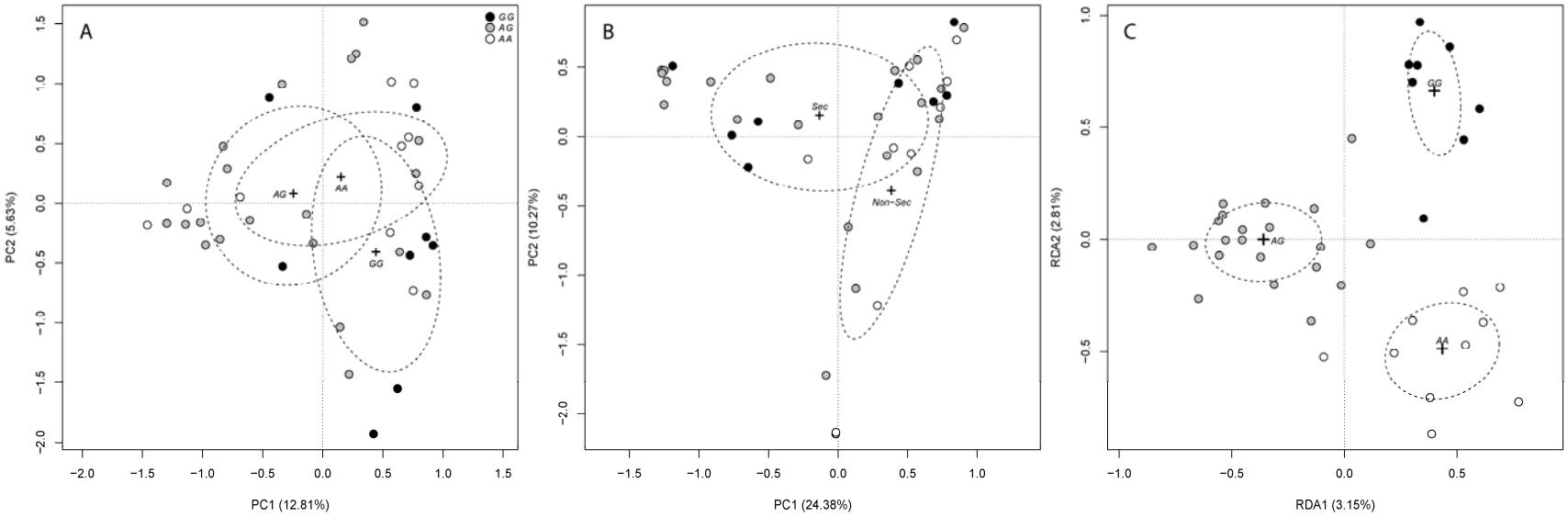


The figures demonstrate GRAIL [31] output results visualized with the VIZ-GRAIL software [33]. Outer circle boxes represent lead SNPs previously or currently identified as suggestive or robust PSC susceptibility loci used as input for the GRAIL analysis (see Supplementary Methods). Inner circle represents genes and genomic regions defined by the outer circle lead SNPs that were identified in the GRAIL analysis based on LD characteristics, genes scored to a $P_{text} < 0.05$ in the GRAIL analysis are indicated in bold The lines between the inner circle genes represent functionally related genes within different loci, the thickness of lines is proportional to the relative similarity of the genes connected by the lines and inversely proportional to the number of genes within the loci that the genes are derived from.

(A) GRAIL pathway analysis based on PubMed abstracts published prior to December 2006.

(B) GRAIL pathway analysis based on PubMed abstracts published prior to April 2011.

Supplementary Figure 4. Influence of *FUT2* genotype on beta diversity



(A) PCoAs of the unweighted UniFrac metric based on presence/absence of phylogenetic branches and (B) normalized weighted UniFrac incorporating the abundances of phylogenetic branches. Centroids were positioned and evaluated using an iterative approach and clusters are denoted by the standard deviation of the weighted averages (dashed ellipses) around the centroids (unweighted UniFrac: r^2=0.110, P=0.074; normalized weighted UniFrac: r^2=0.104, P=0.016). (C) Community relationship (99% OTUs) in an environment spanned only by *FUT2* genotypes explains 5.96% of the total variation in the bacterial species distribution (Redundancy Analysis: F=1.140, P=0.085; RDA1: F=1.206, P=0.093; RDA2: F=1.075, P=0.266).